

HOLLAND BOOKS

IN ECONOMICS

Dani Rodrik
Mark Rosenzweig

Development Economics

VOLUME 5

NORTH-HOLLAND

HANDBOOK OF DEVELOPMENT ECONOMICS

VOLUME

5

Edited by

DANI RODRIK

*John F. Kennedy School of Government,
Harvard University,
Cambridge, MA 02138, USA*

MARK R. ROSENZWEIG

*Department of Economics,
Yale University,
New Haven, CT 06520-8269, USA*



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier



HANDBOOKS IN ECONOMICS

Series Editors

KENNETH J. ARROW
MICHAEL D. INTRILIGATOR



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier



HANDBOOK OF DEVELOPMENT ECONOMICS

VOLUME

5

North-Holland is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

© 2010 Elsevier BV. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary. Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN: 978-0-444-52944-2

For information on all North-Holland publications
visit our website at books.elsevier.com

Printed and bound in the United States of America
10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

INTRODUCTION TO THE SERIES

The aim of the *Handbooks in Economics* series is to produce Handbooks for various branches of economics, each of which is a definitive source, reference, and teaching supplement for use by professional researchers and advanced graduate students. Each Handbook provides self-contained surveys of the current state of a branch of economics in the form of chapters prepared by leading specialists on various aspects of this branch of economics. These surveys summarize not only received results but also newer developments, from recent journal articles and discussion papers. Some original material is also included, but the main goal is to provide comprehensive and accessible surveys. The Handbooks are intended to provide not only useful reference volumes for professional collections but also possible supplementary readings for advanced courses for graduate students in economics.

**Kenneth J. Arrow
Michael D. Intriligator**

This page intentionally left blank

CONTENTS OF THE HANDBOOK

VOLUME 1

PART 1 ECONOMIC DEVELOPMENT—CONCEPTS AND APPROACHES

Introduction

T.N. Srinivasan

CHAPTER 1 The Concept of Development

Amartya Sen

CHAPTER 2 The Roots of Development Theory

W. Arthur Lewis

CHAPTER 3 Alternative Approaches to Development Economics

Pranab Bardhan

CHAPTER 4 Analytics of Development: Dualism

Gustav Ranis

CHAPTER 5 Economic Organization, Information, and Development

Joseph E. Stiglitz

CHAPTER 6 Long-run Income Distribution and Growth

Lance Taylor *and* Persio Arida

PART 2 STRUCTURAL TRANSFORMATION

Introduction

Hollis Chenery

CHAPTER 7 Patterns of Structural Change

Moshe Syrquin

CHAPTER 8 The Agricultural Transformation

C. Peter Timmer

CHAPTER 9 Industrialization and Trade

Howard Pack

CHAPTER 10 Saving and Development

Mark Gersovitz

CHAPTER 11 Migration and Urbanization

Jeffrey G. Williamson

PART 3 HUMAN RESOURCES AND LABOR MARKETS

Introduction

T.N. Srinivasan

CHAPTER 12 Economic Approaches to Population Growth

Nancy Birdsall

CHAPTER 13 Education Investments and Returns

T. Paul Schultz

CHAPTER 14 Health and Nutrition

Jere R. Behrman *and* Anil B. Deolalikar

CHAPTER 15 Labor Markets in Low-Income Countries

Mark R. Rosenzweig

CHAPTER 16 Credit Markets and Interlinked Transactions

Clive Bell

Volume 2

PART 4 PLANNING AND RESOURCE ALLOCATION

Introduction

Hollis Chenery

CHAPTER 17 Short-Run Macroeconomics

Persio Arida *and* Lance Taylor

CHAPTER 18 Multisectoral Models

Sherman Robinson

CHAPTER 19 Income Distribution and Development

Irma Adelman *and* Sherman Robinson

CHAPTER 20 Taxation for Developing Countries

Nicholas Stern *and* Ehtisham Ahmad

CHAPTER 21 Project Evaluation in Theory and Practice

Lyn Squire

PART 5 INTERNATIONAL ASPECTS**Introduction**

T.N. Srinivasan

CHAPTER 22 International Cooperation

Paul Streeten

CHAPTER 23 Trade and Development

Christopher Bliss

CHAPTER 24 Alternative Perspectives on Trade and Development

David Evans

CHAPTER 25 Foreign Public Capital Flows

Jonathan Eaton

CHAPTER 26 Foreign Private Capital Flows

Eliana A. Cardoso *and* Rudiger Dornbusch

CHAPTER 27 Transitional Corporations and Direct Foreign Investment

Gerald Helleiner

CHAPTER 28 Disequilibrium and Structural Adjustment

Sebastian Edwards *and* Sweder van Wijnbergen

PART 6 COUNTRY EXPERIENCE WITH DEVELOPMENT

Introduction

Hollis Chenery

CHAPTER 29 Primary Exporting Countries

Stephen R. Lewis, Jr.

CHAPTER 30 Import Substitution

Henry Bruton

CHAPTER 31 Outward Orientation

Bela Balassa

CHAPTER 32 Large Countries: The Influence of Size

Dwight H. Perkins *and* Moshe Syrquin

VOLUME 3A

PART 7 ANALYTICAL TOOLS

Introduction

Jere Behrman *and* T.N. Srinivasan

CHAPTER 33 Data and Econometric Tools for Development Analysis

Angus Deaton

CHAPTER 34 Human Resources: Empirical Modeling of Household and Family Decisions

John Strauss *and* Duncan Thomas

CHAPTER 35 Applied General Equilibrium Models for Policy Analysis

Jan Willem Gunning *and* Michiel Keyzer

PART 8 RESOURCES, TECHNOLOGY, AND INSTITUTIONS**Introduction**

Jere Behrman *and* T.N. Srinivasan

CHAPTER 36 Savings, Credit and Insurance

Timothy Besley

CHAPTER 37 Technological Change and Technology Strategy

Robert E. Evenson *and* Larry E. Westphal

CHAPTER 38 Institutions and Economic Development

Justin Yifu Lin *and* Jeffrey B. Nugent

CHAPTER 39 Poverty, Institutions, and the Environmental-Resource Base

Partha Dasgupta *and* Karl-Goran Mäler

VOLUME 3B**PART 9 POLICY REFORM, STABILIZATION, STRUCTURAL ADJUSTMENT AND GROWTH****Introduction**

Jere Behrman *and* T.N. Srinivasan

CHAPTER 40 Policy Lessons from Development Experience Since the Second World War

Anne Krueger

CHAPTER 41 Poverty and Policy

Michael Lipton *and* Martin Ravallion

CHAPTER 42 Power, Distortions, Revolt and Reform in Agricultural Land Relations

Hans P. Binswanger, Klaus Deininger *and* Gershon Feder

CHAPTER 43 Human and Physical Infrastructure: Investment and Pricing Policies in Developing Countries

Emmanuel Jimenez

CHAPTER 44 Structural Adjustment, Stabilization and Policy Reform: Domestic and International Finance

Viti'orio Corbo *and* Stanley Fischer

CHAPTER 45 Trade and Industrial Policy Reform

Dani Rodrik

CHAPTER 46 The Contributions of Endogenous Growth Theory to the Analysis of Development Problems: An Assessment

Pranab Bardhan

VOLUME 4**PART 10 NEW INSIGHTS INTO RURAL AND AGRICULTURAL DEVELOPMENT****Introduction**

T. Paul Schultz *and* John Strauss

CHAPTER 47 Economic Development and the Decline of Agricultural Employment

Andrew D. Foster *and* Mark R. Rosenzweig

CHAPTER 48 Information Networks in Dynamic Agrarian Economics

Kaivan Munshi

PART 11 PUBLIC GOODS AND POLITICAL ECONOMY: THEORY AND EVIDENCE**CHAPTER 49 Public Action for Public Goods**

Abhijit Banerjee, Lakshmi Iyer *and* Rohini Somanathan

CHAPTER 50 Understanding Political Corruption in Low Income Countries

Rohini Pande

PART 12 HUMAN RESOURCES AND HOUSEHOLD RESPONSES TO MARKET INCENTIVES AND PUBLIC GOODS**CHAPTER 51 Household Formation and Marriage Markets in Rural Areas**

Marcel Fafchamps *and* Agnes R. Quisumbing

CHAPTER 52 Population Policies, Fertility, Women's Human Capital, and Child Quality

T. Paul Schultz

CHAPTER 53 Health Economics for Low-Income Countries

Germano Mwabu

CHAPTER 54 Health over the Life Course

John Strauss *and* Duncan Thomas

CHAPTER 55 Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy

Peter F. Orazem *and* Elizabeth M. King

CHAPTER 56 The Impact of Child Health and Nutrition on Education in Less Developed Countries

Paul Glewwe *and* Edward A. Miguel

CHAPTER 57 Child Labor

Eric V. Edmonds

CHAPTER 58 Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions

Donald Cox *and* Marcel Fafchamps

PART 13 PROGRAM EVALUATION: METHODS AND APPLICATIONS**CHAPTER 59 Evaluating Anti-Poverty Programs**

Martin Ravallion

CHAPTER 60 Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated

Petra E. Todd

CHAPTER 61 Using Randomization in Development Economics Research: A Toolkit

Esther Duflo, Rachel Glennerster *and* Michael Kremer

CHAPTER 62 Evaluating Conditional Schooling and Health Programs

Susan W. Parker, Luis Rubalcava *and* Graciela Teruel

This page intentionally left blank

PREFACE

Development Policy and Development Economics: An Introduction*

Dani Rodrik

John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138, USA

Mark R. Rosenzweig

Department of Economics, Yale University, P.O. Box 208269, New Haven, CT 06520-8269, USA

Anyone who undertakes to produce a volume of surveys in economic development must confront the question: Does the world really need another one? There have been four volumes in the present series alone, going back to 1988 (Chenery & Srinivasan, 1988), with the latest collection published in 2008 (Schultz & Strauss, 2008). The field changes over time and, one hopes, knowledge accumulates. So, one motive is the desire to cover the more recent advances. And indeed, economic development has been one of the most dynamic and innovative fields within economics in recent years.

But we had another motive as well. We envisaged this Handbook to have a somewhat different focus from earlier ones. In particular, rather than just surveying the “state of the literature” in various subfields, what we sought to accomplish is to present critical and analytical surveys of what we know (and do not know) in different *policy areas*. We asked the authors of each chapter to answer the questions: “What kind of policy guidance does the literature offer in this particular area of development? Where are the gaps? What can we say with certainty that we know? What are the weaknesses of the literature from a policy perspective? What kind of research do we need to undertake to answer burning policy questions of the day? To what extent does actual policy practice correspond to the prescriptions that follow from solid research?” We thus envisioned that the audience for this volume would not only be graduate students and other academics, but well-trained policy makers in developing countries or international financial institutions as well.

While one primary goal is to inform policy makers, we also hope that the volume will assist scholars in designing research agendas that are informed by policy questions, in particular, by the gaps in knowledge that would speak to major policy issues. The

development field has always been one in which the worlds of research and practice are in close relationship with each other and move in tandem. The large number of PhD economists who work in international organizations such as the World Bank and the influence of academia among developing-country officialdom ensure that ideas emanating from the ivory tower often find quick application. But equally important, in principle, is the reverse feedback—the need to tilt researchers’ attention on the questions that are, or should be, on the policy agenda. The organization of the present volume around policy issues is designed to make a contribution toward both of these goals.

1. KEY THEMES

Before briefly summarizing the individual chapters, it is worthwhile to highlight some of the main cross-cutting themes that emerge from the material that follows.

1. *The policies that impact development are wide-ranging, all the way from broad macroeconomic policies such as monetary and exchange-rate policies to interventions in microfinance.* This is perhaps one of the differences between the economics of development and other fields within economics. Poverty reduction, economic growth, and development most broadly are the outcomes of a complex set of interactions across the entire range of economic policies and institutions. From this perspective, “development policies” must have a very broad meaning indeed. As the chapters in this Handbook illustrate, one cannot associate the field of development with policies in just a narrow set of domains—such as education or growth. Nor do these effects and interactions neatly separate along methodological dividing lines such as microeconomics versus macroeconomics. To cite just a few examples: A couple of years’ bout with high inflation can undo the effects on poverty reduction of decades of effective antipoverty programs. In the absence of effective social programs, high growth may not deliver much real development. The returns to schooling may be affected by health interventions that affect life expectancy or cognitive capacity. Improving credit markets may affect schooling. And the analysis of institutions, human capital effects on growth, or environmental policy span across micro- and macroperspectives.
2. *The efficacy of policy is rarely a question of “does it work”; instead it is a question of “when does it work and when not and why?”* Economic theory is rich and flexible enough to justify many different policies, depending on the constraints and opportunities defined by the context. Nowhere is this general point better illustrated than in the area of development policies. The chapters that follow elucidate the point amply. Whether it is in trade, macroeconomics, labor markets, property-rights, education, or microfinance, there is no unique correspondence, as the Washington

Consensus and other general recipes suppose, between policies and outcomes. Earlier attempts to establish universal linkages (such as the impact of tariffs or foreign aid on growth) do not survive the most recent generation of work.

As the newer literature surveyed here makes clear, the trick is to know enough about context and the universe of applicable models to be able to apply the right framework. There are no simplistic rules of thumb and no alternative to the hard work of applying economic reasoning and evidence to the context. Even in such an area as property-rights, around which current orthodoxy has coalesced, there are no clear-cut and immediate policy implications one can draw *in general*. As Besley and Ghatak note, “The creation of effective property rights is heterogeneous in its impact and there are many potential mechanisms that can sustain property rights. This suggests that there should not be a ‘one size fits all’ mantra of extension of private property rights, nor a blind faith that this is a magic bullet that will cure all economic ills.” All of this suggests the need for far greater modesty when generalizing about the role of policy and for studies that evaluate polices or interventions in different contexts.

Relatedly, details of policy implementation matter. The specifics of how a policy or intervention is designed can have a substantial impact on its success. Decentralization of schooling may in principle lead to improved outcomes because educational resource allocations better reflect their suitability to differing local environments, but as studies reviewed by Behrman show, the implicit incentives built in to actual decentralization programs can undermine their achievements. Similarly, Karlan and Morduch emphasize in their chapters that the specific delivery mechanisms for financial products can have substantial effects on outcomes such as repayment rates for similar-sounding financial products (e.g., microloans).

3. *Appropriate development policies typically exhibit high degrees of complementarity.* One reason behind context-specificity is the presence of prerequisites for policies to work appropriately. Policy can work or fail depending on the presence of those prerequisites and the feasibility of implementing the entire package. For example, Kose et al. list a long list of macroeconomic and structural policies that need to be in place for countries to reap the benefits of financial globalization, and emphasize how policies toward the capital account need to be seen as part of a much broader set of policies. Similarly, Harrison and Rodriguez-Clare emphasize the relevance of complementary aspects of the policy regime—such as labor-market policies or the ease of entry and exit for firms—to the success of the trade policy. As Temple points out, we cannot analyze the economics of aid without combining it with the economics of growth, trade, and political economy. As Baland et al. argue, governance is best thought of as a bundle or cluster of institutions and not just a single thing. Hanson points to the interactions between trade in goods, domestic schooling polices, and

international labor flows. Schultz points to the ineffectiveness of schooling interventions where the health environment is such that life expectancy is low. And Dasgupta emphasizes the incompleteness of any account of development that leaves natural capital out of the picture.

4. *Even though developing countries have become significantly more integrated into the world economy in recent decades, their integration remains skewed, has been full of surprises, and has failed to provide many of the expected benefits.* Many of the chapters in this Handbook (those on trade, aid, international labor mobility, monetary and currency policies, financial integration) focus on the linkages between the world economy and development outcomes, and on how domestic policies in both developed and developing countries affect the quality and consequences of those linkages. This is entirely appropriate, as the integration of developing economies in the world economy has been one of the major trends of the last quarter century. Yet, as the chapters demonstrate, this has been a far from smooth process: trade liberalization has not always paid off the anticipated dividends and financial globalization has brought frequent and painful crises. International labor mobility, which could be a great boon, remains mostly restricted, by high-income countries. As the authors document, however, we are now in a far better position to understand these processes and to avoid the blind spots of the past.
5. *Actual experience with development policy has increasingly led economists to broaden and deepen the scope of the analysis into the areas of institutions, governance, and politics.* Scratch any economic issue of consequence, and you are likely to find politics lurking underneath. The chapters that follow demonstrate that this is all the more true of development economics. Economists increasingly acknowledge the importance of institutions—the rules of the game in a society—and the nature of political and power struggles that lie behind them. Encouragingly, they are also increasingly applying their empirical and conceptual tools to analyze those deeper determinants. Issues of governance, politics, and power are no longer a sideshow; they constitute a central element in the field. However, while there has been increased attention to generic issues such as corruption and politician competence, there are few studies that examine the political economy aspects of specific policies, although Behrman's chapter discusses some research that focuses on how elites suppress public schooling in their own interest. As Dasgupta notes in his chapter, the institutions of governance with respect to natural capital—domestic and international—is also beginning to receive some belated attention.
6. *Getting cause-and-effect right is extremely important in designing development policy, and has become a central focus of research in development. But there are dangers.* There is consensus that associations between variables cannot be used as a sound basis for policy. For example, Shultz's chapter on health and population documents the strong positive

historical relationships across countries and over time between income and health measures (e.g., life expectancy). Some have argued that indeed, based on these associations, ill health is a major barrier to development. What is unclear is to what extent these associations reflect the effects of income on health demand and/or jointly reflect differences in both governmental capacities to simultaneously deliver health services and promote income growth. Existing microstudies, which deal with the issue of causation, however, all indicate only modest effects of improvements in nutrition or health on productivity even among the poorest households relative to the aggregate correlations observed in cross-country data.

While to some extent policies should be skewed toward interventions for which evidence is more persuasive, the limitations of empirical methods and of data (see below) mean that only a limited set of interventions will be well-supported by causal evidence. And, indeed, empirical studies will focus on those topics for which obtaining credible, causal estimates is least costly. Thus, the danger is that some interventions that may have large payoffs for generating growth and for which there is little empirical support, because of the absence of evidence, may be neglected. Of course, one aim of this Handbook is to point researchers to areas of potential policy relevance where evidence is weak.

7. *Distinguishing symptoms of underdevelopment from root causes of underdevelopment is key to setting the right policies.* This is distinct from identifying cause and effect. One may find a true causal remedy for a symptom, but remedying the symptom may not lead to economic growth or development. Distributing a certain pill efficiently may cost-effectively and causally reduce the incidence of a particular disease that is prevalent in low-income countries, but even eradicating the disease may not lead to substantial economic growth. Finding which “pills” work for which symptoms thus may not go a long way toward solving problems of underdevelopment. A program redistributing income to the poor alleviates poverty to some extent, but it does not address the root cause of poverty and thus may not be sustainable in the long run. In most low-income countries, school attendance and school inputs are also at low levels. There may be many ways in which schooling can be increased, and there is now increased evidence on the “effectiveness” of various mechanisms (see Chapter 73), through cash transfers that condition on schooling, via policies that relieve credit constraints, or through improvements in school availability or school quality. But if schooling demand is low, principally because payoffs to schooling are low, such interventions will have little effect on poverty reduction and growth. Alternatively, a policy change that accelerated the pace of technical change and raised the demand for skilled workers might both alleviate all of the symptoms of low schooling and enhance economic growth. An understanding of the contexts in which schooling and health contribute to long-term growth is needed along with

evidence on the effectiveness of policies that induce increased schooling demand or increase healthiness. Evidence on the former may be more difficult to obtain compared with evaluating specific interventions that raise the quality of schooling, attract students and teachers to school, or eradicate disease.

8. *We learn from many different types of evidence.* Development policy progresses through the updating of priors on what works, how, and where. These chapters demonstrate that there are multitudes of ways in which our priors can be updated. For example, economists have exploited historical events—for example, colonization and technological inventions; natural events—for example, twinning and disease spikes; and policy variation to assess the impacts of institutions and population and health policies on income growth. Economists are also directly inducing intervention randomness, rather than achieving it through imposing structure on existing data. Randomized controlled trials (RCTs), in which randomly selected subpopulations are selected for an intervention and then outcomes are compared across the treated and untreated populations, have been used to evaluate the causal effects of specific programs (e.g., cash transfers, subsidies to medical inputs), delivery mechanisms (e.g., kinds of financial products), and, less pervasively, to obtain evidence on fundamental behavioral assumptions that underlie models used to justify policy—for example, adverse selection.

As we learn from the chapters on aid, microfinance, health, and education, RCTs are extremely useful in some settings, because, unlike many other empirical techniques, they leave little room for questioning the internal validity of the results. But this method has a number of important limitations. First, in common with other empirical methods, the findings obtained from one population may not be generalizable to others (problems of external validity). This, of course, is just the same principle that applies to development policies—we need to know where interventions work, and why. Second, in common with RCTs applied to medicines but not to other methods, if an intervention is deemed successful in the short run it is almost always then provided to the control population, and if the intervention is deemed ineffective in the short run it is abandoned, thus making it difficult to assess long-term effects of any randomized trial. For example, the original Mexican *Progresa* program, which restricted cash transfers to a randomly chosen subpopulation, was deemed to be a success and then immediately extended to all eligible households. Third, in common with the medical use of RCTs and many other methods, interventions are evaluated one at a time—there are few comparisons of alternative interventions that could achieve the same outcome within a given experiment. Some of these limitations are remediable—multiple interventions can be randomized, treated and nontreated populations can be followed over time, and the same set of interventions can be tried in different populations to assess external validity.

But, at this point in time we have almost no evidence of long-term effects based on RCTs and few systematic comparisons of the outcomes of RCTs across different populations to improve our understanding of the conditions in which particular interventions are most effective.

The limitations of cross-country empirical work have been increasingly recognized in recent years, but as the chapters in this Handbook demonstrate, there is still much that can be learned from these exercises if appropriate judgment is exercised. First, aggregate cross-country associations can be used to establish important regularities (as in Freeman's chapter on labor markets and Hanson's on international labor flows) or to cast doubt on policy presumptions based on incorrect assumptions about cross-country relationships. The chapters by Harrison and Rodriguez-Clare and by Kose provide important illustrations of the latter. As these authors discuss, strongly held professional priors about the impact of trade and financial opening on growth have been considerably weakened by the latest generation of cross-country regressions. These cases represent learning—or perhaps more appropriately, unlearning. History and case studies can also be useful, and many of our chapters make strategic use of them. Schultz, for example, examines the historical demographic transition. He also shows that the existing macroevidence on the relationship between fertility decline, which ages a population, and savings is weak or nonexistent, even though the life-cycle model of savings predicts that savings should rise as a consequence of aging. The impact of fertility decline on savings has been one key motivation for population control policies. A second potential benefit of using relationships among aggregate statistic is that they permit inferences about externalities. For example, the microevidence on the causal effects of, say, health or fertility change on individual wages captures private but ignores social gains from policies that improve health or reduce family size. As Schultz and Behrman point out in their respective chapters, there has in fact been little evidence on the externalities of fertility, schooling, or health that help justify public interventions in these areas. In the field of environment and natural capital, Dasgupta notes that it is possible to learn from about household and village behaviour governing local ecosystems from studies that differ greatly in style, ranging from narratives on individual lives to econometric studies that cover many villages.

9. *The experience of developing countries with diverse policies and institutions provides a rich laboratory for learning about the effect of policies and institutional arrangements—so has made important contribution to economics as an applied science.* Policies and institutions in the advanced nations tend to be stable and relatively uniform. Policies in the developing nations tend to be in a flux, and both policies and institutions vary considerably across time and space. This is both a blessing and a curse. You want to test the idea that nominal variable can have no real effects? What better environment to look at than developing countries with high inflation? Want to scrutinize the diverse ways

in which property rights affect economic incentives or how firms and households cope with absent markets? Why not look across developing countries—where there is huge variance—or indeed within African countries where there can be significant variation in property-right protection even across households? Want to understand the implications of diverse labor-market policies? Again, the developing countries with a range of institutions provide a great laboratory. Of course, it is precisely this heterogeneity in institutional settings and policies that makes it difficult to draw general conclusions about a policy from individual studies confined to particular settings, or from cross-country regressions that assume that policy-response coefficients are identical across countries. Nevertheless, the combination of economic theorizing along with careful and thoughtful empirical work applied to new data has led to an increased understanding of the effectiveness of policies across a wide spectrum of environments, as the chapters in this Handbook demonstrate.

2. AN OVERVIEW OF THE VOLUME

We now provide a brief summary of the chapters that follow.

Harrison and Rodriguez-Clare ask the question: When should the government depart from policy neutrality in trade and foreign investment—employing industrial policies or tilting in favor of (or against) inward foreign investment—and what does the empirical literature say about the record of such policies? The authors draw attention to two important empirical regularities in recent decades: first, there is no significant relationship between average protection levels and growth; and second, there is a positive association between trade volumes and growth. They interpret these regularities to suggest that any successful industrial policy strategy must ultimately raise the share of international trade in GDP. The authors are also skeptical that promotion of FDI is warranted by the existing evidence. They find little support for what they call “hard” interventions (those that distort prices to deal with Marshallian externalities, learning-by-exporting, and knowledge spillovers from FDI). But they believe there is room for “soft” industrial policy, wherein the goal is to develop a process whereby government, industry, and cluster-level private organizations collaborate on interventions to increase productivity. They suggest programs to help particular clusters “by improving the formation of skilled workers, regulation, and infrastructure.”

As Levy Yeyati and Sturzenegger underscore, few things can have as quick and dramatic effect on poverty as mismanaged monetary policy that ends up in a balance-of-payments crisis and high (or hyper-) inflation. Their chapter focuses on the role of “sound monetary policy,” and on the evolving understanding of how that idea is practiced. It emphasizes the role of both economic theory and economic history in shaping governments’ choices with respect to monetary and exchange rate policies. A central argument in the chapter is that these policies not only affect key nominal

variables, but, also as a result, real outcomes such as output volatility, economic growth, and income distribution. In addition, they affect many other variables that are only loosely related to monetary issues. For example, currency stability may foster trade, or augment financial fragility as it undermines the incentives of agents to hedge against currency risk. As Levy Yeyati and Sturzenegger note, the pros and cons of alternative policies depend both on the country context and the changing global context. So the debate on appropriate monetary and exchange rate policies is far from closed.

Kose et al. survey the extent of financial globalization and the nature of the policies that help developing nations put it to better use. They conclude that financial sector development, institutional quality, and trade openness help to derive the benefits of globalization. Low levels of public debt and the avoidance of fixed exchange rate regimes are also important prerequisites. But this chapter also emphasizes that “the relationship between financial integration and economic policies is a complex one and that there are unavoidable tensions inherent in evaluating the risks and benefits associated with financial globalization.” Consequently, the risk-benefit ratio has to be evaluated in a country-specific manner. One of the main contributions of this chapter is the idea that financial integration’s benefits may come not directly in the form of greater investment and better risk sharing, but indirectly in the form of “collateral” benefits: improved macroeconomic discipline, development of the domestic financial sector, competitive discipline, and so on.

Temple reviews the rich and contentious literature on foreign aid and its effectiveness. His chapter’s focus, appropriately, is not on whether aid works, but on when it works. The first half of the chapter takes a general equilibrium perspective, and discusses aggregate models, the Dutch Disease, and the political economy of aid. This part also discusses the cross-country evidence that underpins much of the academic literature on foreign aid. The second half of the chapter is organized more concretely around current debates among donors. Here Temple reviews recent changes in donor policies, against the background of the emergence of a new “partnership” model for donors and aid recipients. He focuses on the shift toward greater targeting of aid, governance and institutions, local ownership of reforms, the streamlining of conditionality, and direct budgeting support for recipient governments. Temple ends on a hopeful note: “Although much remains unknown, arguably enough has been learnt that aid and conditionality can be more effective in the next few decades than in the past. By assigning a central role to governance and institutions, academics and policy-makers may have finally identified the most important binding constraint on growth, development, and aid effectiveness. Donor policies are responding to this, and evolving toward greater flexibility in approach . . .”

It has become trite to say that markets depend on well-defined property rights. But how exactly do these matter, and what is the evidence that they matter in the development process? Besley and Ghatak devote their chapter to these questions. They develop

a unified analytical framework for studying the role of property rights in economic development. The model aims to address two fundamental questions: (i) What are the mechanisms through which property rights affect economic activity? Property rights may work through reduced expropriation, through facilitating gains from trade, or by acting as collateral. (ii) What are the determinants of property rights? Using the analytical framework as an organizing device, Besley and Ghatak survey some of the main empirical and theoretical ideas in the literature. They caution about not fetishizing property rights reform (as it may have diverse effects in different settings), and warn that there may not be a single, uniform manner in which it can be implemented.

Baland, Moene, and Robinson address the broader issue of governance. Their chapter yields four main conclusions. First, governance remains a vague notion which is difficult to unbundled in practice. Nevertheless, the emphasis on governance is quite consistent with the results of the recent academic literature emphasizing the importance of institutions. Second, governance is the outcome of a political process, and as such, the analysis of governance is closely related to the literature on the political economy of development. Third, reformers who want to improve governance need to understand the nature of the underlying political equilibrium which determined prevailing arrangements of governance. Finally, little is understood about the forces that create or impede endogenous improvements in governance, and this remains an important frontier for future research.

Freeman provides a survey of the institutional landscape of labor markets in developing nations, and looks at the effects of labor market practices such as employment-protection laws, minimum wages, and unions on economic outcomes. Developing nations are characterized by labor market dualism. But Freeman's evaluation greatly qualifies the simple Harris-Todaro picture of labor markets where rigidities in urban formal sectors have adverse implications both for efficiency and equity, and in which there are easy benefits to be had from weakening labor regulations. He concludes that "regulations and unions are not the bugbear to development that many believed them to be years ago." Where policies have adverse effects on employment, the magnitudes are generally modest. Freeman emphasizes that practices in formal labor markets have important spillovers on informal labor markets. For example, minimum wages produce spikes in wage distribution in informal labor markets as well. He provides capsule summaries of the Chinese and Argentinean experiences, which suggest that labor market institutions can play a positive role in promoting development and preserving social stability. The chapter calls for more research on the informal sector, about which Freeman says we still know far too little.

One sign of the growing integration between developed and developing countries is the increase in labor flows from low- to high-income countries. The associated phenomenon of remittances has also caught the attention of development economists, as these financial flows from high- to low-income countries exceed in magnitude

foreign aid flows. Hanson's chapter documents the rise in international labor mobility and in remittances. In this area, it is the policies of developed countries that appear to have the greatest impact on these phenomena and thus on the development of poor countries experiencing emigration—high-income countries severely restrict both the quantity and types of immigration, but few countries control labor outflows, although some actively support labor emigration (e.g., the Philippines). Hanson's review of the existing literature suggests that we do not yet know the primary causes of the still relatively low levels of interborder labor movements given the substantial disparities in world wages. We do know that most emigrants are positively selective with respect to human capital and that remittances do not appear to have any special effects on sending-country domestic investment. Hanson asks the policy question of whether increased outflows that would be enabled by less restrictive high-income country immigration policies would aid development in low-income countries, and finds that the evidence is still too weak to provide a reliable answer, but does point to mechanisms to study—effects on wages (generally positive), effects on domestic returns to schooling (unclear), effects on income distribution (depends on the selectivity of receiving-country restrictions). Areas of interaction across topics highlighted in this chapter include domestic schooling policy—schooling abroad is an alternative to domestic investment in education but has implications for the permanent outflow of workers as well—and trade and aid, although the literature on whether trade, aid, and migration are complements or substitutes is not yet clear.

Chapter 72 documents the declines in fertility and improvements in health across the world generally and examines the evidence on the impact of health on productivity and on the effects of fertility reduction on human capital development. He presents evidence that policy-makers in low-income countries wish to lower fertility while those in high-income countries tend to favor policies that are pro-natalist. However, the amount of public domestic or international-donor funds devoted to affecting fertility decisions is relatively modest and has declined over time. Existing microstudies that examine the impact of family size on human capital also point to only relatively small effects. Such microstudies, however, do not capture externalities that may exist from population growth reduction. The aggregate effects of fertility reduction on age-composition also do not appear to have major effects on growth. Emerging evidence does suggest that early interventions in health, including when a child is in the womb, have potentially the largest impacts on adult productivity. Because a child's health is proximately affected by parental decisions, the chapter highlights these and points to how an effective health policy that seeks to promote economic growth must focus on early interventions and thus on the role of parents.

Behrman focuses on schooling policy issues pertaining to the effectiveness of improvements in school inputs such as smaller classes, or textbooks, and the efficacy of decentralization of schooling. There is now a growing literature, using RCTs, that

has examined schooling input effects and incentive schemes to improve both the quality and quantity of schooling resources, but Behrman notes the paucity of evidence of long-term effects of schooling improvements and findings across multiple settings. An emerging theme is that preschool interventions may be especially effective in the majority of low-income countries where primary schooling is now the norm. These findings complement those of the health literature, which, as reviewed in Schultz's chapter, also show stronger effects of early interventions. The set of RCTs is still too limited to draw generalizable conclusions; however, there is a relative absence of studies that look at the general issue of schooling incentives. Behrman also points to a lack of study of the markets in which schooling inputs are supplied, most especially on what determines the quantity and quality of teachers and other education providers.

Many economists believe that poverty itself is a barrier to development, given the limitations of credit and insurance markets—the poor are too poor to save or invest in either human capital or businesses that spur growth. Such a view takes as given financial institutions, however. And it is true that existing formal institutions such as banks find it unprofitable to offer financial services to the poor, and the poor also appear not to be interested in insurance products and to have low savings rates. Karlan and Morduch's chapter takes the point of view that the principal barrier to providing the poor with financial resources is the absence of delivery mechanisms that appropriately take into account market imperfections, informal institutions, and behavior. One well-known institutional innovation in finance is “microcredit,” but evidence of its success is mixed and an understanding of why and whether the specific features of microcredit mechanisms contribute to solving the fundamental problems of credit markets is incomplete. Designing the appropriate financial institutions and delivery mechanisms, of course, requires a deep understanding of behavior and informal sources of finance. The chapter reviews these with respect to how they contribute to designing financial delivery mechanisms. They then review recent RCTs that look at how differing mechanisms of delivery for savings, insurance, and loan products affect both take-up rates and sustainability (e.g., repayment rates in the case of loans), as well as how such products affect savings, contribute to consumption smoothing, and spur business investment.

Of all the issues discussed in this volume, none has the potential to do as much lasting damage to the prospects for development, if mismanaged, as climate change and other sources of degradation in “natural capital.” As Partha Dasgupta notes, policy discussion on these issues vacillates between two perspectives, one pessimistic and held mostly by ecologists and the other optimistic and held mostly by economists. The former points with alarm to the current pattern of apparently unsustainably use of natural resources, while the latter takes comfort in the fact that any potential scarcity is not (yet) reflected in rising prices for marketed resources. The complex interactions

between the environment and economic development is the subject of Dasgupta's chapter.

Dasgupta begins by emphasizing the huge importance of natural capital in the lives of the poor, which he argues is often neglected in economic analyses. He then discusses the externalities involved and the kinds of institutions, particularly community-based ones, most likely to be effective in removing them. Common property resources (CPRs) have both their pros and cons, but have been diminishing in importance for a variety of reasons that Dasgupta identifies. The rest of his paper focuses on ways to value natural capital and proposes a method for evaluating economic programs. He draws a distinction between project evaluation and identifying sustainable development. He shows that GDP is a misleading indicator in the presence of natural capital, as it measures the well-being of only the current generation, and should not be used as a welfare index in either type of exercise. He proposes a comprehensive measure of wealth as an alternative, and discusses the relevance of his preferred index for both identifying sustainable development and undertaking project evaluation. Dasgupta argues that determining the shadow prices of ecological capital assets should be central to the future research agenda in development economics.

3. CONCLUDING NOTE

As these contributions amply demonstrate, development remains a vibrant field, with lively interplay among theory, empirics, and policy. We hope these chapters will give a good account of where the field stands. We also hope their suggestions on important new policy-relevant research will eventually contribute to this Handbook's own obsolescence.

End Notes

- * This is the introduction to *Handbook of Development Economics*, vol. 5, Dani Rodrik and Mark R. Rosenzweig (Eds.), North-Holland, 2009, forthcoming.

References

- Chenery, H., & Srinivasan, T. N. (Eds.). (1988). *Handbook of development economics* (Vol. 1). Amsterdam: Elsevier Science.
 Schultz, T. P., & Strauss, J. (Eds.). (2008). *Handbook of development economics* (Vol. 4). Amsterdam: Elsevier Science.

This page intentionally left blank

CONTENTS OF VOLUME 5

<i>Introduction to the Series</i>	v
<i>Contents of the Handbook</i>	vii
<i>Preface</i>	xv
PART Fourteen The Economics of Development policy	
CHAPTER 63 Trade, Foreign Investment, and Industrial Policy for Developing Countries	4039
Ann Harrison and Andrés Rodríguez-Clare	
Abstract	4040
Keywords	4040
1. Introduction	4040
2. The Theoretical Justification for Industrial Policy	4043
2.1 Multiple equilibria and latent comparative advantage	4045
2.2 Industrial policy without latent comparative advantage	4052
2.3 Industrial policy as sector-specific collective action	4054
2.4 Self-discovery and diversification	4059
2.5 National and global gains from IP	4060
3. Empirical Evidence on Infant-Industry Protection and Other Forms of IP	4062
3.1 Single-industry studies	4063
3.2 Cross-industry studies	4065
3.3 Cross-country studies	4067
3.4 Export subsidies and other forms of IP	4070
4. Trade and Economic Development	4071
4.1 Stylized facts on trade policies, 1980–2004	4072
4.2 Cross-country evidence on trade policies, trade volumes, productivity, and growth	4080
4.3 What can the different effects of trade volumes versus trade policy tell us about the success of IP?	4091
4.4 Identifying the mechanisms for gains from trade	4095
5. Foreign Direct Investment	4099
5.1 FDI and productivity	4103
5.2 FDI and host-country factor markets	4108
6. Concluding Comments	4110
Appendix	4114
End Notes	4196
References	4198

CHAPTER 64 Monetary and Exchange Rate Policies	4215
<i>Eduardo Levy Yeyati and Federico Sturzenegger</i>	
Abstract	4215
Keywords	4216
1. Introduction	4216
1.1 Do nominal variables matter for development?	4217
2. What Do We Talk About When We Talk About MERP?	4220
2.1 Classifying exchange rate policies	4224
2.2 Exchange rate policy trends in the post-Bretton Woods era: A casual glance at the distribution of regime choices	4230
2.3 Inside the nonfloat group	4233
2.4 Monetary policy: From exchange rate anchors to inflation targeting	4236
3. Why Do We Talk About MERP?	4240
3.1 The link between exchange rates and growth revisited	4241
3.2 ERR and output volatility	4246
3.3 MERP on price stability	4247
3.4 MERP on income distribution	4248
3.5 Indirect links: The integration channel	4249
3.6 Indirect links: The financial channel	4251
3.7 Summing up	4254
4. The Making of a Policy	4254
4.1 Exchange rate anchors in the 1980s	4254
4.2 Financial integration and financial crises in the 1990s	4258
4.3 Float cum inflation targeting (FIT)	4260
4.4 The comeback of exchange rate regimes: Leaning against the appreciation wind	4263
5. Where Do We Stand?	4267
End Notes	4269
References	4273
CHAPTER 65 Financial Globalization and Economic Policies	4283
<i>M. Ayhan Kose, Eswar Prasad, Kenneth Rogoff and Shang-Jin Wei</i>	
Abstract	4284
Keywords	4284
1. Introduction	4284
2. Financial Globalization: Measurement and Trends	4288
2.1 How to measure financial integration?	4288
2.2 Evolution of financial globalization: Some basic stylized facts	4289
2.3 Factors driving financial globalization	4292

3.	Macroeconomic Implications of Financial Globalization	4292
3.1	Economic growth	4294
3.2	Macroeconomic volatility and patterns of international risk-sharing	4303
4.	Economic Policies and Growth Outcomes	4306
5.	Structural Characteristics and Economic Policies	4310
5.1	Financial sector development	4310
5.2	Institutional quality	4316
5.3	Trade openness	4323
6.	Macroeconomic Policies	4330
6.1	Theory	4330
6.2	Evidence	4330
7.	Direction of Capital Flows and Economic Growth	4332
7.1	Empirical evidence	4332
7.2	How to interpret these findings?	4334
7.3	Are developing countries savings-constrained?	4336
7.4	Policy implications	4337
8.	Capital Controls as a Policy Tool	4337
8.1	Implementation and measurement issues	4337
8.2	Macroeconomic implications of capital controls	4338
8.3	Microeconomic implications of capital controls	4339
9.	How to Approach Financial Integration?	4341
10.	Conclusion	4343
	End Notes	4345
	References	4350
	Postscript to “Financial Globalization and Economic Policies”	4360
1.	Openness, Structural Features and Policies	4360
2.	Policies Regulating Capital Account	4361
	End Notes	4361
	References	4362
	CHAPTER 66 International Migration and the Developing World	4363
	Gordon H. Hanson	
	Abstract	4363
	Keywords	4364
1.	Introduction	4364
2.	The Dimensions of International Migration	4366
2.1	Data sources	4367
2.2	International migration patterns	4368
2.3	Skilled emigration	4375
2.4	Networks and migration costs	4383
2.5	Discussion	4385

3.	Impact of Emigration on Sending Countries	4386
3.1	Labor markets and fiscal accounts	4386
3.2	Remittances and return migration	4390
3.3	Information and the flow of ideas	4394
3.4	Discussion	4395
4.	Immigration Policy Regimes	4396
4.1	Political economy of immigration policy	4396
4.2	The design of immigration policy regimes	4398
4.3	Discussion	4402
5.	Final Discussion	4403
	End Notes	4405
	References	4408

CHAPTER 67 Aid and Conditionality 4415

Jonathan R. W. Temple

Abstract	4416	
Keywords	4417	
1.	Introduction	4417
2.	Key Concepts and Stylized Facts	4423
2.1	The Millennium Development Goals	4423
2.2	Recent trends in aid	4425
2.3	The forms of aid	4431
2.4	Fungibility	4432
2.5	Debt relief	4433
3.	Aid and Growth in Theory	4435
3.1	Capital shortage and neoclassical growth models	4435
3.2	Poverty-trap models	4437
3.3	Virtuous circles and amplification effects	4439
3.4	Binding constraints	4439
3.5	Complementary inputs	4441
3.6	Growth versus basic needs	4442
3.7	Summary	4442
4.	Has Aid Worked?	4443
4.1	Project evaluations	4443
4.2	Case studies	4444
4.3	Cross-country evidence	4445
4.4	Aid and social indicators	4447
4.5	Summary	4448
5.	The View From Trade Theory	4449
5.1	The transfer problem	4449
5.2	The Dutch Disease	4450
5.3	Evidence on the Dutch Disease	4451

6.	Aid, Fiscal Policy, and Volatility	4452
7.	Aid and Governance	4454
7.1	History and the state	4456
7.2	Aid and political leadership	4457
7.3	Aid and the common-pool problem	4458
7.4	Aid and corruption	4459
7.5	Aid, political instability, and war	4461
7.6	Aid to fragile states	4462
8.	A New Partnership Model?	4463
8.1	Should democracies get the most aid?	4464
8.2	Democracy as a necessary condition	4465
8.3	Democracy as a sufficient condition	4466
8.4	Summary	4467
9.	Conditionality in Theory and Practice	4468
9.1	Policy conditionality	4469
9.2	Commitment problems	4470
9.3	Heroes and villains	4471
9.4	Conditionality versus ownership	4472
9.5	Governance conditionality	4474
10.	Selective Aid Allocation	4475
10.1	Developing rules for aid allocation	4477
10.2	Dynamic aid allocation	4477
11.	The New Conditionality	4478
11.1	The meaning of ownership	4479
11.2	Poverty Reduction Strategy Papers	4479
11.3	Social participation	4480
11.4	New forms of conditionality	4481
11.5	The future of conditionality	4483
12.	Donor Policies	4483
12.1	Proliferation of projects and programs	4484
12.2	Learning and accountability	4485
12.3	Randomized trials and evidence-based aid policies	4486
12.4	Evaluate, evaluate, evaluate	4489
13.	Meeting Basic Needs	4490
13.1	Public service delivery	4491
13.2	Agriculture and rural development	4492
14.	Aid for the Twenty-First Century	4493
14.1	New directions for foreign aid	4495
15.	Conclusions	4499

Appendix	4501
Data	4501
Endogeneity and control variables	4501
Panel data and dynamic responses	4502
Error dependence	4503
Heterogeneity	4503
Publication bias	4505
End Notes	4506
References	4511
CHAPTER 68 Property Rights and Economic Development	4525
<i>Timothy Besley and Maitreesh Ghatak</i>	
Abstract	4525
Keywords	4525
1. Introduction	4526
2. Resource Allocation and Property Rights	4528
2.1 The role of property rights in limiting expropriation	4529
2.2 Insecure property rights as a barrier to trade	4534
2.3 Optimal assignment of property rights	4545
2.4 Evidence	4552
3. Endogenous Property Rights	4559
3.1 Expropriation	4559
3.2 Improving state effectiveness	4583
4. Concluding Comments	4588
End Notes	4589
References	4592
CHAPTER 69 Governance and Development	4597
<i>Jean-Marie Baland, Karl Ove Moene and James A. Robinson</i>	
Abstract	4597
Keywords	4598
1. Introduction	4598
2. What is Governance	4602
3. Measurement	4603
3.1 Political institutions	4603
3.2 State capacity and effectiveness	4603
4. Theoretical Mechanisms	4608
4.1 The Political economy of governance in Sierra Leone	4610

5.	Consequences and Causes of Variation in Governance	4615
5.1	Correlates of governance	4615
5.2	The first stage	4620
5.3	Origins of bad governance in Sierra Leone	4625
6.	Improving Governance	4629
6.1	Persistence of power and incentives—The See-Saw Effect	4631
6.2	Persistence of <i>de facto</i> power	4633
6.3	The Iron Law of Oligarchy	4637
6.4	Successful reform	4640
6.5	Governance and equity—The Scandinavian model	4641
6.6	Promoting reform	4647
7.	Conclusions	4648
	End Notes	4649
	References	4650
CHAPTER 70 Labor Regulations, Unions, and Social Protection in Developing Countries: Market Distortions or Efficient Institutions?		4657
Richard B. Freeman		
	Abstract	4657
	Keywords	4658
1.	The Debate Over Labor Institutions	4661
1.1	Theoretical perspectives	4664
2.	Microevidence on Minimum Wages	4667
3.	Wage Curve	4670
4.	Employment Regulations	4672
5.	Mandated Benefits	4674
6.	Union Effects	4674
7.	Cross-Country Evidence	4680
8.	Country Cases	4682
8.1	China	4683
8.2	Argentina	4684
9.	The Informal Sector	4684
10.	Conclusions	4687
	Appendix A	4688
	Appendix B: Regression	4693
	End Notes	4693
	References	4695

CHAPTER 71 Access to Finance	4703
Dean Karlan <i>and</i> Jonathan Morduch	
Abstract	4704
Keywords	4704
1. Introduction	4704
1.1 Mechanisms matter	4706
1.2 Testing what works	4707
2. Global Financial Access	4710
3. Financial Intermediation and Economic Growth	4712
4. Returns to Capital	4715
4.1 Framing the question	4717
4.2 Evidence from estimating profit functions	4717
4.3 Evidence from field experiments	4719
4.4 What do average returns tell us?	4720
4.5 Sensitivity to interest rates	4721
4.6 A final caveat	4722
5. Credit Market Innovations	4723
5.1 Nature of frictions and policy examples	4724
5.2 Interventions and mechanisms	4729
5.3 Impacts from solving credit market failures	4736
6. The Economics of Saving	4739
6.1 Basic models of saving	4742
6.2 Constraints to saving	4745
6.3 Impacts of saving	4757
7. Risk Management and Insurance	4759
7.1 Why insurance markets fail	4762
7.2 Partnership models and index-based insurance	4766
7.3 Health insurance	4769
8. Governments, Businesses, and Nonprofit Institutions	4770
9. Concluding Comments	4774
End Notes	4775
References	4777
CHAPTER 72 Population and Health Policies	4785
T. Paul Schultz	
Abstract	4785
Keywords	4786
1. Introduction	4786

2.	Historical Transition of Death and Birth Rates and Population Policies	4789
2.1	International population policy assistance: Levels and trends	4792
3.	Stylized Facts Related to Health, Fertility, and Development	4797
4.	Framework for Studying the Determinants and Consequences of Health	4800
4.1	Estimating without bias health production functions with endogenous inputs	4802
4.2	Estimating the marginal product of health human capital	4804
4.3	Parent investment in children in response to their children's initial endowments	4805
4.4	Spillovers from health human capital beyond the individual and family	4810
4.5	Phases of the mortality transition and implications for health inequality	4812
5.	Macroeconomics Evidence on Health Determinants and Economic Growth	4813
5.1	Macroeconomic evidence of the impact of income on health	4816
5.2	Alternative indicators of health status and their economic consequences	4818
6.	Microeconomics of Health and Development: Individuals and Households	4820
6.1	Height, weight, body mass index, and birth weight as indicators of health stocks	4822
6.2	Heterogeneity of height as a measure of the stock of health human capital	4828
6.3	Pathways from fetal and early child development to adult productive life span	4840
7.	Macroeconomics of Fertility and Development: Diminishing Returns to Labor, Human Capital and Savings, and Changing Age Composition	4844
8.	Microevidence of Determinants of Fertility and Consequences of Policy on Fertility	4847
8.1	Income and price effects affecting the quantity and quality of children	4848
8.2	Family planning policies and fertility: Cross-country relationships	4850
8.3	Estimating the cross-effects of fertility variation on family lifetime outcomes	4853
8.4	A social experiment in family planning and reproductive health: Bangladesh	4855
9.	Internal Migration, Demographic Transition, and Evaluation of Policy	4857
10.	Conclusions and Policy Directions	4861
	End Notes	4865
	References	4871
CHAPTER 73	Investment in Education—Inputs and Incentives	4883
Jere R. Behrman		
	Abstract	4883
	Keywords	4884
1.	Introduction	4884

2.	Framework for Analysis and What We Know and Do Not Know	4886
2.1	Demand side—Models, estimates, and policy implications	4886
3.	Supply Side—Models, Estimates, and Policy Implications	4914
3.1	Supply-side modeling	4915
3.2	Estimates and policy implications	4920
4.	Markets for Products and Inputs—Models, Estimates, and Policy Implications	4943
4.1	Sorting in educational markets	4943
4.2	Supplies and sectoral choices of childcare workers, teachers, trainers, and other educational service sector staff	4944
4.3	Educational expenditures and age distributions	4945
5.	Conclusions	4946
	Appendix: Estimation Issues, Possible Methodological Resolutions, and Data	4947
A.1.	Econometric Approaches to Dealing with Estimation Problems	4947
A.2.	Strengths and Limitations of Analysis of Various Types of Data and Related Methods	4952
A.2.1	Some major characteristics pertaining to data quality	4953
A.2.2	Some major types of data for investigating education and development	4955
	End Notes	4959
	References	4965
CHAPTER 74	The Place of Nature in Economic Development	4977
Partha Dasgupta		
	Abstract	4978
	Keywords	4978
1.	Nature in Economics	4978
2.	Natural Capital and Economic Development	4980
3.	Natural Capital and Development Economics	4982
4.	Types of Natural Capital	4985
5.	Contents	4987
5.1	Externalities	4988
6.	Unidirectional Externalities: Exports and Wealth Transfers	4989
6.1	Quantifying externalities	4990
6.2	Internalizing externalities	4990
7.	Reciprocal Externalities: Common Property Resources	4992
7.1	Why CPRs?	4992
7.2	The importance of CPRs	4994
7.3	The good news about CPRs	4995
7.4	The bad news about CPRs	4996
7.5	Valuation, evaluation, and sustainable development	5001

8. Inadequacies in GDP	5001
9. Valuing Goods and Services	5002
9.1 Defining shadow prices	5003
9.2 Estimating shadow prices	5003
10. Social Well-Being and Comprehensive Wealth	5006
10.1 A unifying model	5007
10.2 The welfare significance of net domestic product	5008
11. Comprehensive Investment and Sustainable Development	5009
11.1 The basic theory	5009
11.2 Global public goods	5011
11.3 An application	5012
12. Evaluating Policy Reforms	5016
12.1 Social cost-benefit analysis	5016
12.2 Evaluating projects in practice	5017
13. Discounting Climate Change	5021
13.1 Global activism now or later?	5021
13.2 Negative discount rates	5022
13.3 The welfare economics of climate change	5022
13.4 Social aversions to inequality and risk	5024
13.5 Climate treaties	5025
14. Conclusions	5026
Appendix	5029
A.1. Economic Perturbation as Movement Through Time:	
Sustainable Development	5029
A.1.1 The formal analysis	5029
A.1.2 Accounting for population growth	5031
A.2. Policy Evaluation	5033
End Notes	5034
References	5036
<i>Index</i>	5047

This page intentionally left blank

PART *Fourteen*

The Economics of Development Policy

Trade, Foreign Investment, and Industrial Policy for Developing Countries*

Ann Harrison

University of California at Berkeley and NBER

Andrés Rodríguez-Clare

Pennsylvania State University and NBER

Contents

1. Introduction	4040
2. The Theoretical Justification for Industrial Policy	4043
2.1 Multiple equilibria and latent comparative advantage	4045
2.2 Industrial policy without latent comparative advantage	4052
2.3 Industrial policy as sector-specific collective action	4054
2.4 Self-discovery and diversification	4059
2.5 National and global gains from IP	4060
3. Empirical Evidence on Infant-Industry Protection and Other Forms of IP	4062
3.1 Single-industry studies	4063
3.2 Cross-industry studies	4065
3.3 Cross-country studies	4067
3.4 Export subsidies and other forms of IP	4070
4. Trade and Economic Development	4071
4.1 Stylized facts on trade policies, 1980–2004	4072
4.2 Cross-country evidence on trade policies, trade volumes, productivity, and growth	4080
4.3 What can the different effects of trade volumes versus trade policy tell us about the success of IP?	4091
4.4 Identifying the mechanisms for gains from trade	4095
5. Foreign Direct Investment	4099
5.1 FDI and productivity	4103
5.2 FDI and host-country factor markets	4108
6. Concluding Comments	4110
Appendix	4114
End Notes	4196
References	4198

Abstract

In this chapter we explore the popular but controversial idea that developing countries benefit from abandoning policy neutrality *vis-a-vis* trade, FDI and resource allocation across industries. Are developing countries justified in imposing tariffs, subsidies, and tax breaks that imply distortions beyond the ones associated with optimal taxes or revenue constraints? We refer to this set of government interventions as "industrial policy." We explore the theoretical foundation for industrial policy and then review the related empirical literature. We follow this with a broader look at the empirical work on the relationship between trade and FDI and growth. In this review we find no support for "hard" interventions that distort prices to deal with Marshallian externalities, learning by exporting, and knowledge spillovers from FDI. Nevertheless, we still envision an important role for what we refer to as "soft" industrial policy. The goal is to develop a process whereby government, industry and cluster-level private organizations can collaborate on interventions to increase productivity. We suggest programs and grants to help particular clusters by improving the formation of skilled workers, technology adoption, regulation and infrastructure.

JEL classifications: L50, L51, O12, O13

Keywords

trade
foreign investment
industrial policy
agglomeration

1. INTRODUCTION

...growth was not a passive, trickle-down strategy for helping the poor. It was an active, pull-up strategy instead. It required a government that would energetically take steps to accelerate growth, through a variety of policies including building infrastructure such as roads and ports and attracting foreign funds.

Jagdish Bhagwati, In Defense of Globalization (2004)

During the last three decades, developing countries have made enormous strides in opening up their protected domestic markets to international trade and foreign investment. Yet most countries have not simply opened up their markets. They have also instituted a range of policies to encourage exports, attract foreign direct investment (FDI), promote innovation, and favor some industries over others. This leads to the following question: is openness to trade and FDI alone sufficient to achieve high growth rates in developing countries? If harnessing the gains from globalization requires additional policies, can we identify them? While some types of complementary policies, such as building roads and ports, are not controversial, others are. Bhagwati's suggestion to "attract foreign funds" implies tilting incentives in favor of foreign

investors, which means abandoning policy neutrality. Our goal in this chapter is to explore the popular but controversial idea that developing countries benefit from abandoning policy neutrality *vis-a-vis* trade, FDI and resource allocation across industries.

Policy neutrality does not necessarily mean free trade, or a neutral stance regarding taxation of multinational corporations, or even a common tax structure for all industries. Both optimal tax theory and practical fiscal considerations imply that countries (especially poor ones) will often want to rely on tariffs as a source of revenue or set different tax rates across industries. Are developing countries justified in imposing tariffs, subsidies, and tax breaks that imply distortions beyond the ones associated with optimal taxes or revenue constraints? We refer to this set of government interventions as “industrial policy.”

The presence of externalities is the main theoretical justification for deviating from policy neutrality. Learning externalities from exports could justify export subsidies; knowledge spillovers from foreign companies could justify tax breaks for FDI; production externalities in “advanced” sectors could justify infant-industry protection or other measures to expand those industries. We begin this review in Section 2 with a series of simple models to highlight the role of Marshallian and interindustry externalities, industry-level rents, sector-specific coordination failures and information spillovers as a rationale for industrial policy (“IP”).

The main message that emerges from this review is that the theoretical justification for infant-industry protection requires at a minimum either that the country have a latent comparative advantage in the protected industry or that the international price for this industry is higher than warranted by the true opportunity cost of this good in the rest of the world. Moreover, for protection to deliver large gains, the protected industry must exhibit large Marshallian externalities. In contrast to the temporary trade barriers associated with infant-industry protection, permanent protection of a sector may raise welfare if it generates positive externalities to other sectors. In all these models, however, other policies may be more efficient than protection. Even when protection could improve welfare, a production subsidy would be more efficient since it avoids the temporary consumption losses associated with protection. In addition, protection may not work if the market failure is due to sector-specific coordination problems, since tariff-induced growth does not necessarily help to solve coordination failures. Finally, just as R&D subsidies can address the externalities arising from innovation spillovers, policies to promote entry into new industries can address information spillovers associated with the discovery of new profitable activities.

While government intervention could address a number of market failures in theory, one key question is whether IP has been successful in practice. The theoretical discussion is followed by a review of the empirical literature on industrial policy in Section 3. One challenge that we face in evaluating the empirical literature is the large gap between

the theoretical justification for IP and the quantitative work that has been done to evaluate its effectiveness. Even if we could show that protected sectors grow faster, this is not sufficient evidence to claim that IP increases welfare.

Despite this shortcoming, we proceed in Section 3 to evaluate three different approaches to measuring the effectiveness of infant-industry protection. The first approach focuses on particular industries that have received protection, such as the steel rail industry in the United States and semiconductors in Japan. The few existing studies of this nature suggest that the conditions necessary to generate positive net welfare gains from infant-industry protection are difficult to satisfy in developing countries.

A second empirical strategy exploits the variation in productivity growth and different measures of support (including protection and production subsidies) across industries to see whether supported industries exhibit faster growth. The challenge here is that tariffs and quotas are frequently imposed for reasons other than the pursuit of industrial policy. If we cannot identify the motive for protection, then it is difficult to evaluate the consequences of IP via trade policy instruments. Generating fiscal revenue, improving terms of trade, and political considerations (rent seeking) are likely to be just as important as infant-industry considerations in explaining the pattern of trade policy and other measures of support.

A third approach studies the connection between trade policy and economic growth, both by focusing on particular countries (e.g., South Korea and Taiwan) and by considering the entire cross section of countries. While the literature on trade and growth linkages faces many challenging problems, in Section 4 we suggest two general lessons that may be drawn from the voluminous evidence. First, there was no significant relationship in the second half of the twentieth century between average protection levels and growth. Second, there is a positive association between trade volumes and growth. We interpret the lack of a significant association between average tariffs and growth, combined with the strong relationship between trade shares and growth, to suggest that any successful IP strategy must ultimately increase the share of international trade in GDP. The fact that so many countries have been unsuccessful in offsetting the antitrade bias of their interventions may explain why so many have failed to succeed at IP.

We also review a new set of studies that emphasize the complementarity between trade and FDI reforms and other government policies. These studies suggest that trade liberalization will not generate faster growth unless accompanied by changes in other parts of the economy, such as reducing barriers to new firm entry, encouraging more flexible labor markets, and improving infrastructure. These new studies are consistent with one of the most important new theoretical developments in international trade: the emphasis on heterogeneous firms. The new heterogeneous trade models suggest a new mechanism through which trade affects productivity growth: greater competition forces less productive firms to exit and increases the market share of more productive firms. In this framework, gains from trade cannot be realized when there are

barriers to firm exit and expansion, which suggests a need to take into account complementary policies on entry and exit. We finish Section 4 with an analysis of whether learning by exporting seems to be quantitatively important, and a discussion of the empirical evidence regarding other mechanisms through which trade policy might affect growth.

Aside from intervening in trade, many developing countries deviate from policy neutrality by introducing tax breaks and other policies to attract FDI. While economists are generally skeptical regarding the benefits of intervening in trade, they are much more likely to have interventionist priors when it comes to FDI. In Section 5, we review the microstudies which have identified a number of important effects of inward foreign investment flows. While most of the empirical literature focuses on productivity spillovers, there is also a growing literature that examines export promotion through foreign investment, input linkages, and labor market effects. We argue that FDI is associated with technology transfer and positive labor market outcomes in developing countries, but that the empirical literature on FDI does not show evidence of horizontal spillovers, and that benefits from vertical spillovers may not exceed the cost of FDI subsidies.

At the end of this survey, we provide some broad suggestion for industrial, trade, and foreign-investment policy in developing countries. To preview our conclusions, we find no support for “hard” interventions that distort prices to deal with Marshallian externalities, learning by exporting, and knowledge spillovers from FDI. Nevertheless, we still envision an important role for what we refer to as “soft” industrial policy. The goal is to develop a process whereby government, industry and cluster-level private organizations can collaborate on interventions to increase productivity. We suggest programs and grants to help particular clusters by improving the formation of skilled workers, regulation, and infrastructure.

2. THE THEORETICAL JUSTIFICATION FOR INDUSTRIAL POLICY

The textbook model of IP is based on the idea that some sectors or industries exhibit *Marshallian externalities*, which are local externalities that increase with the size of the industry. These externalities can arise through localized industry-level knowledge spillovers, input-output linkages together with transportation costs to ensure that the externalities remain local, and labor pooling (see Krugman, 1991; Marshall, 1920). Marshallian externalities give rise to geographic agglomeration of industries (e.g., software in Silicon Valley), which have been emphasized in the literature on economic geography.

The simplest model of IP entails a small-open economy with two sectors, 1 and 2. Sector 1 has constant returns to scale (CRS) while Sector 2 has Marshallian externalities. The key result is that under some conditions there are multiple equilibria, with

the equilibrium with complete specialization in Sector 2 being superior to the one with complete specialization in Sector 1.¹ One could say that the economy has a “latent” comparative advantage in Sector 2 but that a coordination failure prevents it from exploiting this advantage.² Since the realization of Marshallian externalities is likely to take some time, it is customary to talk about countries having a “dynamic” comparative advantage (Amsden, 1989; Wade, 1990) in sectors other than the ones in which they are currently specialized, and to think of industrial policy as the way to undertake the necessary transformation to capitalize on that dynamic advantage. We think of this as the main theoretical justification for infant-industry protection.

The following section presents a simple static model that formally captures this idea for an economy facing exogenous international prices. We then move on to endogenize these prices based on production costs in the rest of the world and then show how the results extend to a dynamic framework.

If the South does not have a latent comparative advantage in the advanced sectors likely to have Marshallian externalities, is there still a case to be made for an IP that would promote a structural transformation toward those sectors? In Section 2.2 we show that there are indeed conditions under which this is the case. In particular, IP can increase welfare in two scenarios: first, when there are rents associated with the advanced sector, so that its international price is high relative to its cost; and second, when there are interindustry externalities, so that a large advanced sector increases the economy’s productivity across the board.

Marshallian externalities arise as an automatic consequence of the scale of the sector: a sector necessarily experiences an increase in productivity as it becomes larger. As argued by Baldwin (1969), however, the expansion of a sector does not always bring about positive agglomeration externalities. Such agglomeration effects may instead depend on the way in which production is carried out. In other words, externalities may not be intrinsic to sectors, but to the way in which they are organized (Rodríguez-Clare, 2007). In this case, a sectoral reallocation of resources is not enough; import substitution, for example, may allow an economy to expand its manufacturing sector, but production may take place in unsophisticated ways and no “clustering” benefits may materialize. In Section 2.3 we present a model where, instead of Marshallian externalities, sectors present *opportunities* for collective action that increases their productivity by a certain amount (which may be different across sectors). If there are no rents (i.e., if international prices perfectly reflect production costs in the North), then the best policy is simply to exploit the opportunities for collective action in the sectors where the economy is specialized. Under more general conditions, we show that sectors that would benefit more from IP would be ones that have large opportunities for productivity-enhancing collective action, or that have high world demand relative to the combined size of countries that have achieved such collective action.

In Section 2.4 we turn to IP aimed at “diversification.” This is something that is often stated as a goal by many countries. We present a simple model where diversification is linked to productivity and argue that if there is a market failure reducing the level of diversification below the optimal one (as in Hausmann & Rodrik, 2003), then a policy of encouraging discovery and diversification would indeed be welfare enhancing.

2.1 Multiple equilibria and latent comparative advantage

We first present a static model with exogenous prices, and then discuss the implications of the model when prices are determined by production costs in the rest of the world.

2.1.1 Exogenous international prices

There is a small economy, which we call “South,” two goods and one factor of production, labor, in fixed supply, L .³ Good 1 is produced with CRS and no aggregate externalities: a unit of labor produces λ_1 units of good 1. Good 2 is produced with CRS at the firm level, but there are aggregate externalities, so that labor productivity is:

$$\lambda_2[1 + \alpha \text{Min}(\bar{L}, L_2)]$$

with $\alpha > 0$. The term $1 + \alpha \text{Min}(\bar{L}, L_2)$ captures Marshallian externalities that are increasing with industry-wide employment, L_2 , but that are exhausted once the labor force in a sector reaches the level \bar{L} . The term $\theta \equiv 1 + \alpha \bar{L} > 1$ can be seen as the maximum benefits of *clustering* in sector 2.⁴ We assume that the total labor supply in South is higher than \bar{L} , so that if there is complete specialization in good 2 then productivity is $\theta \lambda_2$.

Let p_i^* be the international price of good i and let $p^* \equiv p_2^*/p_1^*$. Let us derive a condition under which there are multiple equilibria, with one equilibrium characterized by complete specialization in good 1 and the other by complete specialization in good 2. We first check that specialization in good 1 is an equilibrium. Letting w denote the wage in South, then $w = \lambda_1 p_1^*$ if South is specialized in good 1. The unit cost of producing good 2 in South given that all labor is devoted to production of good 1 (and hence no benefits of clustering are realized) is w/λ_2 . Hence complete specialization in good 1 is an equilibrium if and only if $\lambda_1/\lambda_2 \geq p^*$. Similarly, complete specialization in good 2 implies $p_2^* = w/\theta \lambda_2$, and hence this is an equilibrium if and only if $\lambda_1/\lambda_2 \leq \theta p^*$. Thus, there is multiple equilibria if and only if the following condition holds,

$$p^* \leq \lambda_1/\lambda_2 \leq \theta p^* \tag{1}$$

Without loss of generality, in the following discussion we restrict attention to the case in which this condition holds with strict inequalities. If there are multiple equilibria, which

equilibrium yields higher wages? In the equilibrium with complete specialization in good 1 the wage is $w = \lambda_1 p_1^*$, whereas in the other equilibrium we have $w = \theta \lambda_2 p_2^*$. If condition (1) is satisfied with strict inequalities then $\theta \lambda_2 p_2^* > \lambda_1 p_1^*$ so wages are higher in the equilibrium with complete specialization in good 2 is superior.⁵

We will say that South has a *latent comparative advantage* in good i if the opportunity cost of this good given the realization of all Marshallian externalities is lower than the international price.⁶ For good 1 this entails $\theta \lambda_2 / \lambda_1 \leq 1/p^*$, whereas for good 2 this is $\lambda_1 / \theta \lambda_2 \leq p^*$. Thus, condition (1) with strict inequalities implies that South has a latent comparative advantage in good 2. The equilibrium with specialization in good 1 is possible because in this case Marshallian externalities are not realized, and hence the latent comparative advantage of South in good 2 is not what determines the pattern of specialization. Thus, the previous results (i.e., existence of multiple equilibria and the fact that generally the equilibrium with specialization in good 2 yields higher wages than the one with specialization in good 1) can be reinterpreted as saying that a country may be specialized in a sector where it does not have a latent comparative advantage, and that in this case a policy that induces the economy to switch to the equilibrium with specialization in the good where there is a latent comparative advantage *could* be welfare enhancing (see discussion below).

Figure 1 illustrates the previous results. The curve labeled PPF represents the production possibilities frontier for South, which is convex when $L_2 < \bar{L}$ (or $Q_2 < \theta \lambda_2 \bar{L}$) and becomes linear when $L_2 \geq \bar{L}$ (or $Q_2 \geq \theta \lambda_2 \bar{L}$). The curve labeled PPF_{NC} is the hypothetical production possibilities frontier when there are no Marshallian externalities (i.e., $\alpha = 0$), given simply by a line with slope λ_1 / λ_2 , as in the standard Ricardian model. Note that the slope of the PPF is the same as the slope of PPF_{NC} at the corner where

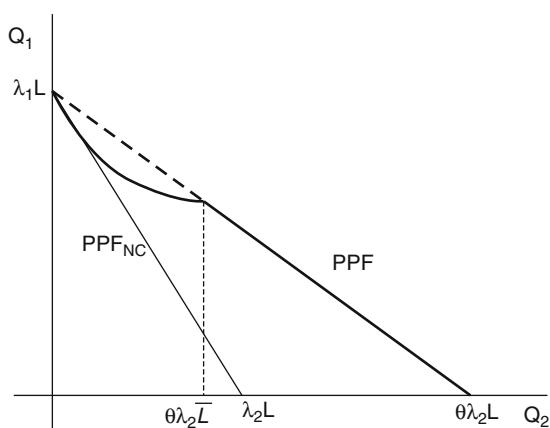


Figure 1 Marshallian externalities and multiple equilibria.

there is complete specialization in good 1. Thus, if the international relative price of good 2 (i.e., p_2^*/p_1^*) is lower than λ_1/λ_2 there is an equilibrium with complete specialization in good 1, whereas if p_2^*/p_1^* is higher than the (negative of the) slope of the PPF along its linear segment—namely, $\lambda_1/\theta\lambda_2$ —then there is an equilibrium with complete specialization in good 2. Clearly, then, if (1) is satisfied, there are multiple equilibria, with the equilibrium with specialization in good 2 delivering a higher consumption possibilities frontier for South.

The standard case for infant-industry protection or IP can now be stated simply by saying that if the South is specialized in good 1, then a high enough tariff would lead this economy to satisfy its own consumption of good 2. This would allow the South to realize the benefits of the Marshallian externalities associated with this sector, and thereby shift the equilibrium toward complete specialization in good 2.⁷ At that point the tariff would no longer be needed, and the economy could maintain free trade.

Sauré (2007) presents an interesting and novel argument for why protection may fail to generate the results predicted by this model. He assumes that goods 1 and 2 are tradable inputs into the production of a final nontradable good via a “modern technology.” The key assumption is that the final good can also be produced directly from labor with a “traditional technology” which exhibits CRS (no Marshallian economies). Under some conditions, protection makes the modern technology unprofitable because of the increase in the price of input 2. Thus, rather than increasing production of the good with Marshallian externalities, protection leads to the contraction of the whole modern sector and a reallocation of resources toward the traditional sector.

2.1.2 International prices determined in north

The previous discussion takes international prices as exogenous. But prices are (at least in part) determined by productivity levels in the rest of the world. One would imagine that rich countries are already enjoying the lower costs associated with clustering in sector 2, so these lower costs would be reflected in p^* . Thus, for a small country to have a latent comparative advantage in sector 2, it must have some deep parameters that confer it such an advantage. To see this, imagine now that there are two countries, North and South, which may differ in the productivity parameters, λ_1 and λ_2 , so that there are exogenous international productivity differences (independent of Marshallian externalities) in the production of both goods 1 and 2. Imagine further that South is small, so that prices are determined in North as if it were a closed economy. Choosing labor in North as the numeraire, international prices are simply given by the North’s unit labor requirements. Assuming that in equilibrium $L_{2N} \geq \bar{L}$ then $p_1^* = 1/\lambda_{1N}$ and $p_2^* = 1/\theta\lambda_{2N}$, so $p^* = \lambda_{1N}/\theta\lambda_{2N}$. Note that the benefits of clustering are reflected in a lower international relative price of good 2. This will be important in the analysis that follows.

Imagine first that there are no Ricardian productivity differences, $\lambda_{ji} = 1$ for all $j = S, N$ and $i = 1, 2$. Using (1) it is easy to confirm that there are multiple equilibria in South (since $p^* = 1/\theta$ then condition (1) is $1/\theta \leq 1 \leq 1$), but since the second part of (1) is satisfied with equality, then the wage is the same in both equilibria. Thus, although there are multiple equilibria, the wage is not higher in the equilibrium with specialization in good 2. This is because even though the economy benefits from clustering in this equilibrium, this is exactly compensated by the lower price of this good, which in turn arises from the higher productivity in North derived from clustering.

The equilibrium with specialization in good 2 may yield higher wages than to the one with specialization in good 1 if we allow for exogenous productivity differences. In particular, the equilibrium with specialization in good 2 will have higher wages if the South has a latent comparative advantage in the good subject to clustering. To see this, drop the assumption that $\lambda_{ji} = 1$ for all j, i , and assume instead that

$$\lambda_{2S}/\lambda_{1S} > \lambda_{2N}/\lambda_{1N} \quad (\text{CA})$$

The condition for multiple equilibria (i.e., condition (1)) is now (using $p^* = \lambda_{1N}/\theta\lambda_{2N}$)

$$\lambda_{1N}/\theta\lambda_{2N} \leq \lambda_{1S}/\lambda_{2S} \leq \lambda_{1N}/\lambda_{2N} \quad (2)$$

The second inequality is satisfied given (CA), so there is always an equilibrium with specialization in good 2. The first inequality (needed for there to be an equilibrium with specialization in good 1) is satisfied if and only if

$$\frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}} \leq \theta \quad (3)$$

That is, the South's comparative advantage in sector 2 must be weaker than the benefits of clustering.

The analysis here is exactly as above, with condition (2) replacing condition (1). Condition (CA) is necessary for South to have a latent comparative advantage in good 2, and this is necessary for the wage with specialization in good 2 to be higher and for IP to raise wages for South.

An important point to note is that for the gains from IP to be large we need South to have a strong latent comparative advantage in good 2. But then condition (3) implies that Marshallian externalities must also be high for there to be multiple equilibria. We can conclude that *IP generates large gains only if the sector that would be promoted exhibits both a strong latent comparative advantage and large externalities*.

One can enrich the model to generate some additional implications. For example, if sector 2 is intensive in physical and human capital relative to sector 1, then if South is

specialized in 1 this goes together with low levels of both types of capital and a lower level of TFP than if it were specialized in sector 2 (Ciccone & Matsuyama, 1996; Rodríguez-Clare, 1996). A policy to shift resources toward sector 2 would then generate endogenous accumulations of capital, as observed in East Asian countries. In other words, some LDCs may have low capital stocks as well as low TFP as a consequence of not exploiting their latent comparative advantage.⁸

2.1.3 Dynamic externalities

Here we extend the model to allow for dynamic Marshallian externalities, as in Bardhan (1970), Krugman (1987), Lucas (1988), Redding (1999), and Melitz (2005). To do so, we assume that productivity in sector 1 in country i is λ_{1i} , just as above, whereas productivity in sector 2 in country i at time t is now $A_{it}\lambda_{2i}[1 + \alpha\min(\bar{L}, L_{2it})]$. Letting $a_{St} \equiv \min\{A_{St}/A_{Nt}, 1\}$ and $a_{Nt} \equiv \min\{A_{Nt}/A_{St}, 1\}$, we assume that A_{it} grows thanks to both learning by doing (which happens if country i has a cluster in sector 2, that is, $L_{2it} > 0$) and international spillovers (which happens if $a_{it} < 1$). Formally, we assume that

$$\dot{A}_{it} = (g/\bar{L})\min(\bar{L}, L_{2it})A_{it} + \varepsilon(1 - a_{it})A_{it}$$

where $\varepsilon > g > 0$. Productivity increases caused by dynamic externalities in one country eventually diffuse to the other country even if there is no cluster there. Thus, in this model clusters are important to generate knowledge but are not critical to benefit from knowledge spillovers.

Note that if the North has a cluster but the South does not, then $\dot{A}_{St} = \varepsilon(1 - a_{St})A_{St}$. There are “benefits of backwardness,” in the sense that a lower relative productivity in South (i.e., lower a_{St}) leads to a faster rate of productivity growth. This implies that given $\varepsilon > g$ there is a steady state productivity gap $A_{St}/A_{Nt} = \hat{a}_S$ given implicitly by $g = \varepsilon(1 - \hat{a}_S)$: if $a_{St} < \hat{a}_S$ ($a_{St} > \hat{a}_S$) then a_{St} increases (decreases) toward \hat{a}_S . For future reference, note that the productivity of North relative to South in this steady state is $(\lambda_{2N}/\lambda_{2S})\theta/\hat{a}_S$: the first-term captures pure Ricardian productivity differences, whereas the second- and third-terms capture the impact of static and dynamic benefits of clustering, respectively. Starting from such a steady state, if South acquires a cluster in sector 2, so that now $L_{2St} \geq \bar{L}$, then there will be full convergence as a_{St} increases from \hat{a}_S toward 1.

As before, assume that prices are wholly determined in North. Then $p_1^* = 1/\lambda_{1N}$ and $p_{2t}^* = 1/A_{Nt}\theta\lambda_{2N}$, and assume that condition (CA) holds. Focusing on South, complete specialization in 1 is a steady state equilibrium if

$$\frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}} \leq \theta/\hat{a}_S \quad (4)$$

On the other hand, complete specialization in good 2 is necessarily a steady state equilibrium under condition CA. Thus, given condition CA, there are multiple steady states if and only if condition (4) is satisfied. It can be readily verified that the steady state with specialization in good 2 entails a higher wage for South than the steady state with specialization in good 1. This is because condition (CA) implies that South has a latent or “dynamic” comparative advantage in good 2.⁹

Imagine that this condition is satisfied and that the system is in steady state with South completely specialized in good 1. Is there an equilibrium in which South moves to the steady state with specialization in good 2? If South becomes specialized in good 2 then it realizes the static externalities, but it will take some time for it to catch up to the North in terms of productivity (i.e., $a_{St} < 1$ for some time). During this period, specialization in good 2 is an equilibrium if

$$\frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}} \geq 1/a_{St}$$

This implies that if

$$1/\hat{a}_S \leq \frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}} \leq \theta/\hat{a}_S$$

then there are multiple steady states, and also an equilibrium in which South switches from specialization in good 1 to specialization in good 2 and eventually reaches the steady state with complete specialization in good 2. The government could achieve this switch with trade protection for an infinitesimally short time, and the gains would necessarily outweigh any associated costs (see below).

On the other hand, if

$$\frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}} < 1/\hat{a}_S$$

then this is no longer the case. In order to guarantee specialization in good 2, there would need to be temporary protection until a_{St} increases from \hat{a}_S to a level a'_S defined implicitly by

$$\frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}} = 1/a'_S$$

After that, complete specialization in good 2 would be an equilibrium for South. Eventually the system would reach the steady state with $a_{St} = 1$.

2.1.4 Discussion

The simple model presented here captures the notion that temporary protection may induce an economy in a bad equilibrium to switch toward the equilibrium where specialization is according to its latent or dynamic comparative advantage. Of course, this is nothing more than the classic case for infant-industry protection, where policy is supposed to turn a latent comparative advantage into an effective one.¹⁰ Such a policy would be welfare enhancing provided it passes both the Mill and Bastable tests: the Mill test is that the protected sector can eventually survive international competition without protection, whereas the Bastable test is more stringent in requiring also that the discounted future benefits compensate the present costs of protection (see Corden, 1997; Kemp, 1960). In the model above, infant-industry protection passes the Mill test if and only if the South has a latent comparative advantage in the protected sector. The Bastable test requires that the discounted gains from IP compensate the temporary consumption losses associated with protection during the period in which the economy is generating the (dynamic) productivity gains associated with clustering. This is the period when a_{St} is increasing from \hat{a} to a' . Bardhan (1971), Redding (1999), and Melitz (2005)—among others—explore the conditions on the learning process under which the benefits of protection justify these initial losses. Bardhan (1970) and Melitz explore the optimal way in which protection should be granted.

Of course, protection does not directly target the presence of Marshallian externalities. A production subsidy would be more efficient, as it would avoid the temporary consumption losses mentioned above. The whole discussion of infant-industry protection is based on the presumption that a production subsidy is simply not feasible, either for fiscal, political, or practical considerations (see Section 3). If this were not the case, then the analysis would be quite simple: provide a (Pigovian) subsidy such that the marginal subsidy is equal to the marginal externality. Naturally, if the source of the externality is not production but some more specific activity (e.g., R&D) then the subsidy could be directed there.

The condition that the South has a nonexploited latent comparative advantage to justify protection is specific to the simple Ricardian model we presented, where the PPF is linear and there is a tendency for complete specialization. If the PPF is strictly concave and international prices are such that under free trade the economy is diversified, then the presence of Marshallian externalities in the import-competing sector implies that a small tariff would necessarily be welfare increasing. But this kind of “marginal intervention” is not what people commonly associate with infant-industry protection.

Another possibility that arises with a strictly concave PPF is multiple equilibria where the bad equilibrium has a low (but positive) production level, whereas the good equilibrium has a higher production level but not enough for any of this production to be exported. Even if the economy does not have a latent comparative advantage,

temporary protection may be welfare enhancing in this case as long as the Mill and Bastable tests are satisfied.

Although it is a result specific to the simple Ricardian model presented here, the result that a latent comparative advantage is necessary for infant-industry protection seems useful to guide policy discussions in practice. First, in all cases in which infant-industry protection has supposedly been successful, the infant sector eventually generates significant exports. Second, if there are resource constraints that prevent the sector from becoming large, then it is likely that the benefits of IP would be small while the associated costs and risk would be relatively large. For this reason, in the rest of this section we restrict attention to the Ricardian model (linear PPF).

2.2 Industrial policy without latent comparative advantage

Can a policy to promote specialization in a good with Marshallian externalities (i.e., good 2 in the model above) be welfare-enhancing in spite of *not* having latent comparative advantage in that sector? For this analysis and the rest of this section we again restrict attention to the simpler static analysis.

2.2.1 Sector-level rents

Imagine that prices are determined not in a single economy, but in a collection of economies. In particular, assume that North is partitioned into two regions, N_1 and N_2 , with labor quantities L_N^1 and L_N^2 , and no labor mobility between N_1 and N_2 . Everything else is as above. Without loss of generality, assume that region N_2 is the one that will produce good 2. If L_N^2 is sufficiently large relative to the world's demand for good 2 then it will not be completely specialized in that good, in which case prices will be determined by technology levels (inclusive of clustering effects), so that $p^* = \lambda_{1N}/\theta\lambda_{2N}$, just as in the previous case. But if L_N^2 is small relative to the world demand for good 2, then in equilibrium one can have

$$\lambda_{1N}/\lambda_{2N} > p^* > \lambda_{1N}/\theta\lambda_{2N} \quad (5)$$

In contrast to what we have with an integrated North, here p^* can be strictly higher than $\lambda_{1N}/\theta\lambda_{2N}$. The difference can be interpreted as the “rents” associated with good 2. To capture this, let

$$R \equiv \frac{p^*}{\lambda_{1N}/\theta\lambda_{2N}}$$

Condition (5) can now be stated as

$$\theta > R > 1 \quad (6)$$

Consider again a “small” South, which takes international prices as given. The condition for multiple equilibria is again given by Eq. (1), with p^* satisfying Eq. (5). Using the definition of R, Equation (1) can be written as

$$\frac{R/\theta}{\lambda_{1N}/\lambda_{2N}} \leq \frac{\lambda_{1S}/\lambda_{2S}}{\lambda_{1N}/\lambda_{2N}} \leq R \quad (7)$$

We are interested in the case where South does not have a latent comparative advantage in good 2, or

$$\lambda_{1S}/\lambda_{2S} > \lambda_{1N}/\lambda_{2N} \quad (8)$$

In this case, the LHS inequality in Eq. (7) is always satisfied (given Eq. 6), while the RHS inequality, which is necessary for specialization in good 2 to be an equilibrium, is satisfied if $R > CA$, where $CA \equiv (\lambda_{1S}/\lambda_{2S})/(\lambda_{1N}/\lambda_{2N})$ is a measure of comparative advantage in South in good 1. In other words, there is multiple equilibria in South if its comparative advantage in good 1 is smaller than the rents associated with international prices. Moreover, just as in the previous cases, the equilibrium with specialization in the good with Marshallian externalities sustains a higher wage.¹¹ This result is reminiscent of the literature on strategic trade policy, where increasing returns and imperfect competition leads to the existence of rents which governments try to capture via trade policy.

An alternative way to get a similar result is by assuming the existence of a wage premium in sector 1.¹² To see this, assume again that North is an integrated region, but with a wage premium in sector 2, so that $w_2 = R w_1$. Then it is easy to verify that $p^* = R \lambda_{1N}/\theta \lambda_{2N}$, which is the same as above.^{13,14}

2.2.2 Interindustry externalities

We have so far focused on intraindustry externalities. Consider instead aggregate externalities, so that all sectors in the country benefit from the externalities realized in a sector (see Greenwald & Stiglitz, 2006; Succar, 1987; Young, 1991).¹⁵ Assume that now productivity in good i in country j is

$$\lambda_{ij}[1 + \alpha_i \text{Min}(\bar{L}, L_{2j})]$$

with $\alpha_1 < \alpha_2$, so that intraindustry externalities are stronger than interindustry externalities. Let $\theta_i \equiv 1 + \alpha_i \bar{L}$ and note that $\theta_1 < \theta_2$. We return to the case in which prices are determined in an integrated North (with no rents), so that $p_i^* = 1/\theta_i \lambda_{iN}$, and assume that South has a latent comparative advantage in good 1. It is readily verified that in this case the only equilibrium in South entails specialization in good 1.¹⁶ But specialization in

good 2 implies $w_S = \theta_2 \lambda_{2S} p_2^* = \lambda_{2S}/\lambda_{2N}$, while specialization in good 1 implies $w_S = \lambda_{1S} p_1^* = \lambda_{1S}/\theta_1 \lambda_{1N}$. The first is higher than the second if and only if

$$1/\theta_1 < \frac{\lambda_{2S}/\lambda_{1S}}{\lambda_{2N}/\lambda_{1N}}$$

If this is satisfied, then forcing the economy toward specialization in good 2, even if this is not an equilibrium, is welfare-enhancing compared to staying in the equilibrium with specialization in good 1. In this case, the losses from going against comparative advantage by specializing in sector 2 are dominated by the gains associated with the economywide externalities generated. This may be one way of interpreting the argument in the late 1980s in favor of protecting the semiconductor industry in the United States (see Borrus, Tyson, & Zysman, 1986).

2.3 Industrial policy as sector-specific collective action

We have so far focused on Marshallian and interindustry externalities as reasons for IP. A more general conceptual framework for thinking about IP is the existence of coordination failures at the industry or sector level. Of course, coordination failures arise in the presence of Marshallian externalities. The difference is that the distortions associated with these externalities (at least as modeled above) disappear when the sector gets sufficiently large, whereas this is not the case with other types of coordination failures. For example, in Rodríguez-Clare (2007) externalities arise only when “modern” technologies are used in a sector. Thus, even sectors that are seen as “advanced” in developed countries can behave as backward sectors when they operate in LDCs, and hence fail to generate any externalities. This captures the idea that what matters for productivity is not “what you produce, but how” (De Ferranti, Perry, Lederman, & Maloney, 2001; Porter, 1998). In these circumstances, a sector can expand and still fail to experience an increase in productivity.¹⁷ Protection or export subsidies would fail, to be welfare-enhancing and other policies would be called for.¹⁸

The existence of coordination failures implies that collective action at the sector level may lead to productivity gains. A concrete example of collective action is the eradication of foot and mouth disease in Uruguay’s cattle industry, which generated significant benefits by allowing the industry to export beef to the United States (see Hausmann, Rodríguez-Clare, & Rodrik, 2005). Simply providing a production or export subsidy to the cattle industry would not have solved the problem. A specific policy to deal with the coordination failure associated with strong externalities was necessary. Another example is the case of flower exports from Ecuador (Hernández, Cely, González, Muñoz, & Prieto, 2007). Several attempts to export flowers in the 1960s and 1970s failed in part because of the lack of reliable air transport to the main destinations. A key difference in the 1980s was an effort by the association of flower exporters,

EXPOFLORES, to convince the government and the national airline to set up the required number of cargo flights for this activity. Thanks in part to this effort, the value of flower exports boomed from less than half a million dollars in 1984 to more than \$400 million in 2006.¹⁹

2.3.1 A simple model

We now want to explore a model where policy can induce higher productivity in a sector through some kind of industry-level collective action, and where prices are determined in a collection of economies (not only in “the North”), so that there may be rents. The goal is not to model the specifics of collective action, but rather to examine the conditions under which this may increase a country’s income level.

There are N countries, indexed by j . Labor is the only factor of production and is available in total quantity L_j in country j . There are M industries indexed by m . There are opportunities for collective action in each industry. Collective action increases productivity in industry m by the factor x_m ; otherwise, productivity is one in all industries in all countries. We refer to x_m as the level of *complexity* in industry m , since it seems reasonable to expect that more complex industries will benefit more from collective action. A country that has achieved high productivity in industry m thanks to collective action will be said to have HP (for high productivity) in that industry. Let k_{jm} be the indicator function for whether country j has HP in industry m , and assume $\sum L_j = 1$ so that L_j is also the share of worldwide labor living in country j . Then $s_m \equiv \sum_j k_{jm} L_j$ is the share of labor in countries with HP in industry m . Also, country j ’s productivity in industry m can be written as $\hat{x}_{jm} \equiv (1 - k_{jm}) + k_{jm}x_m$. Preferences are Cobb-Douglas, with a share v_m devoted to industry m , and $\sum v_m = 1$. Thus, we can think of v_m equivalently as the “size” of industry m , or the extent of the world’s demand for its output.

The model described thus far is a Ricardian model with N countries and M industries, where productivity can be either low or high in each industry. The equilibrium is easy to describe: it consists of a set of wages, w_j , prices, p_m , and an allocation, L_{jm} , for $j = 1, \dots, N$ and $m = 1, \dots, M$ such that for all j and m the following conditions hold: $w_j \geq \hat{x}_{jm} p_m$ and if $L_{jm} > 0$ then $w_j = \hat{x}_{jm} p_m$ (zero profits), and $p_m \sum_j \hat{x}_{jm} L_{jm} = v_m \sum_j w_j L_j$ (i.e., the value of sales of m equals total expenditures on m) for all m .

It is useful to describe an equilibrium without rents. Choosing labor as the numeraire, this entails $w_j = 1$ for all j and $p_m = 1$ if $s_m = 0$ and $p_m = 1/x_m$ if $s_m > 0$; it requires that for each industry either no country has HP or there are enough countries (adjusting for their size) with HP that the large supply drives the price to its marginal cost with unitary wages.²⁰ Note that in this case a country that does not achieve HP in any industry would still enjoy the same wage as other countries. We can think of this as a case in which factor price equalization (FPE) holds.

Rents arise when s_m is small relative to v_m . For example, imagine an equilibrium where $s_m = 0$ for all $m \neq 1$, and only country 1 has HP in industry 1. Then $w_j = 1$ for $j = 2, \dots, N$ while $w_1 > 1$ if and only if $v_1 > s_1$: there are rents in industry 1 (i.e., the price of industry 1 is higher than the marginal cost at unitary wages, $p_1 > 1/x_1$).²¹

Industries differ with respect to three variables: complexity (measured by x_m), size (measured by v_m), and the share of people in the world that live in countries that have HP (measured by s_m). We will refer to the later as “prevalence,” since it measures the extent to which HP is widespread across the world in an industry. The previous result suggests that industries will have rents if they are large relative to their prevalence. Apart from this result, one can learn more from this model only by considering special cases. Instead of doing this, we introduce some additional assumptions to “smooth out” the kinks in the Ricardian model and obtain more general results.

Assume that each industry is composed of a continuum of goods with varying productivity levels. Preferences remain Cobb-Douglas, but now with equal shares across all goods. Thus, assuming that industry m has a measure v_m of goods, then (as above) expenditures on industry m are v_m with $\sum v_m = 1$. (Note that it is natural to think of v_m as the “size” of industry m because it measures both the share of total expenditures devoted to this industry and the measure of goods belonging to that industry.) More importantly, we assume that productivity differs across goods *within an industry* as in Eaton and Kortum (2002). Specifically, productivity for any particular good in sector m in country j is $\hat{x}_{jm}z$, where \hat{x}_{jm} is as above and z is an additional productivity that is independently drawn from the Fréchet distribution with parameters T_j and ϕ , that is, $\Pr_j(z \leq Z) = \exp[-T_j z^{-\phi}]$.²² This distribution has sound microeconomic foundations (see Eaton & Kortum, 2001), but understanding those foundations or its several convenient properties is not important for our purposes here; it is sufficient to know that a higher T_j implies better productivity draws for country j (on average).

Since each good is infinitesimally small and there are no transportation costs, then each good will be supplied to the whole world by the country with the lowest cost. If we consider a particular good in industry m with productivity draws $(z_{1m}, z_{2m}, \dots, z_{Nm})$ in countries $(1, 2, \dots, N)$ then this good will be supplied by country $j = \arg \min_l \{w_l/\hat{x}_{lm}z_{lm}\}$. Eaton and Kortum (2002) show that a country with wage w_j and productivity parameter \hat{x}_{jm} will capture a share

$$D_{jm} = \frac{(w_j/\hat{x}_{jm})^{-\theta} T_j}{\sum_l (w_l/\hat{x}_{lm})^{-\theta} T_l} \quad (9)$$

of total sales in industry m . A country with a lower T_j , a higher w_j , or a lower \hat{x}_{jm} will capture a smaller market share in industry m . Contrary to the standard Ricardian

model, however, a country will have positive production in all industries because it will always have a few goods within any industry where its productivity draws are very high. Letting $Y = \sum w_j L_j$ denote worldwide income, then the trade balance conditions are $\sum_{im} D_{jm} v_m Y = w_j L_j$ (value of sales equal value of purchases for country j). These conditions determine the equilibrium wages w_1, \dots, w_N .

How is the wage in a country affected by acquiring HP in an industry? If a country had a choice, in which industry would achieving collective action have the greatest impact on wages?²³ These are key questions for IP. Note that we have assumed that there are no deep sources of comparative advantage, so this would not be an issue in this choice. Then it seems reasonable that countries would be able to increase wages more by focusing their efforts in industries that have higher complexity, are larger (or have higher demand), and have a lower prevalence: higher complexity means that there is more to gain from collective action, while larger demand combined with low prevalence implies higher rents. Under some conditions, one can in fact prove this result. In particular, assume that countries 1 and 2 are identical except that they have HP in industries 1 and 2, respectively, with no HP in the rest of industries (i.e., country 1 has HP only in industry 1 and country 2 has HP only in industry 2 and $\hat{x}_{j1} = \hat{x}_{j2}$ for all $j = 3, \dots, N$). Then one can show that the wage in country 1 is higher than in country 2 if $x_1 > x_2$ or if $v_1 > v_2$. Also, assuming that $x_1 = x_2$ and $v_1 = v_2$, then $w_1 > w_2$ if $s_1 < s_2$ (see Appendix).²⁴

We have assumed thus far that there are no differences in industry-level productivity across countries. If this were not the case, then it is clear that a country may want to go against its latent comparative advantage and specialize in goods with combinations of high complexity, high demand, and low prevalence. Thus, this model may be suitable to think about IP, although it is important to note that trade protection is not an effective policy in this framework.

2.3.2 Relation of the model to some recent contributions

We now show how to place some recent arguments for IP in the context of this simple analytical framework. Hausmann, Hwang, and Rodrik (2006) argue that the varied economic performance of different countries is partly explained by the goods that they produce. Other things equal (including physical and human capital stocks), countries that specialize in what they call “rich country goods” are richer. Their explanation is that such goods provide more opportunities for learning by doing (similar to Young, 1991) and for technological and institutional upgrading that ultimately benefits the whole economy.

The model presented here can capture this notion in a slightly different way. In this model we may talk about “rich country industries” as ones that are more complex have higher worldwide demand or exhibit lower prevalence. Countries that achieve HP in these industries would be able to sustain higher income levels, and getting there would entail higher growth rates.

Hausmann and Klinger (2006) argue that goods are connected, so that productivity in one good would be higher if the country has already achieved HP in a related good. This has similar implications as interindustry externalities (think of “industries” as “goods”). For example, coming back to the role of specialized inputs, if such inputs exist for one good, this would also help in the production of a similar or related good. Hausmann and Klinger show that some goods are connected to many other goods, while other goods are relatively isolated. They think of the space of goods as a forest, with each good being represented by a tree, and talk about how this forest is more dense in some areas. Hausmann and Klinger, and Hausmann and Rodrik (2006) suggest that if a country could choose, it would want to locate in the denser parts of the forest.

The idea that there are regions in the forest that are more dense is captured in the model above by having industries differ in their size or worldwide demand. The suggestion that countries in dense part of the forests are better off corresponds to the result above that countries are better off if they manage to achieve HP in an industry with a high measure of goods, v_m , since this corresponds to high demand. But the model reveals a weakness of this notion: industries with high demand would not be attractive if they also have high prevalence. Returning to the forest metaphor, being in a dense part of the forest would not be better if this is also more crowded. For example, although the electronics industry may be a “highly connected area” of the forest, there may also be many countries (and large ones, e.g., China) participating in this area. In principle, it could be better for a country to remain in an isolated but relatively uncongested part of the forest.

The measure of the “income level” of an industry developed by Hausmann et al. (2006) takes this into account: an industry with high prevalence would exhibit low prices and would thus be classified as one with a lower income level. Thus, in principle, this measure may be seen as a reasonable way to guide countries in choosing industries that are ideal for productivity-enhancing collective action. Unfortunately, however, this may be a noisy measure of the relevant notion needed for IP because it may be capturing capital (physical, human, or knowledge capital) intensity, which leads countries with good conditions for capital accumulation to produce and export these goods. In other words, if there is an exogenous variation in conditions for capital accumulation across countries, and if goods differ in their capital intensity, then rich countries will tend to produce and export capital intensive goods, and this will have nothing to do with industry-specific collective action and IP. The same association between rich countries and certain sectors may arise because of an exogenous variation in the quality of institutions. For example, if rich countries have institutions that are conducive to capital market development, then they would tend to specialize in goods that rely heavily on outside capital (Manova, 2006). If one could somehow adjust Hausmann, Hwang, and Rodrik’s measure to isolate the income correlation of goods that is *not* explained by capital intensity or exogenous variation in institutions, then this

could be a useful measure for IP. In particular, countries could consider inducing collective action in sectors with a high “adjusted income level.”

2.4 Self-discovery and diversification

The notion of IP that we have emphasized so far is that there are “special industries,” and that countries can increase welfare by reallocating resources to those industries. An alternative idea that we now explore is to think of a policy to increase diversification. This has been a particular concern in countries that specialize in natural-resource intensive industries (see CAF, 2007; De Ferranti, 2001, for recent treatments focusing on Latin America). Diversification could be desirable as a way to reduce volatility, or as a way to increase productivity. Here we focus on the later.

To think about the connection between diversification and productivity, consider the Eaton and Kortum (2002) model of trade. As explained above, Eaton and Kortum model productivities as being drawn from a distribution that is common across countries except for a technology parameter T . This technology parameter determines the location of the productivity distribution: countries with a higher T have “better” distributions in the sense that, on average, productivity draws will be higher (formally, this entails first-order stochastic dominance). Apart from T , countries also differ in size, L . Assuming away trading costs for simplicity, wages are determined by the ratio of technology to size, T/L . A high T/L means that the country would have many sectors in which it has absolute advantage relative to its size, leading to a high equilibrium wage. Moreover, given L , a higher T implies the production (and export) of more goods, or more diversification.

Of course, higher productivity and higher wages may not go together with diversification. For example, high productivity in a sector that is not “diversified” or “differentiated” would draw resources away from the diversified sector and reduce overall diversification even as it increases overall productivity and wages. It is also important to recall that the data reveal that after a certain level of development, higher income goes together with less, not more diversification (see Imbs & Wacziarg, 2003).

An interesting way to think about diversification is via what Hausmann and Rodrik (2003) call “self-discovery.” They argue that countries do not really know their cost structure, and hence they do not know the goods in which they have comparative advantage. This must be discovered through costly experimentation, which is plagued by information spillovers that render its private benefits low relative to its social benefits. The following model shows a simple way to capture the connection between diversification and productivity, and then between self-discovery and diversification.

Consider an economy with labor as the only factor of production and two sectors: agriculture and manufactures. Agriculture is a homogenous good, produced with decreasing marginal returns to labor, with $Q_A = \lambda F(L_A)$. There are a continuum of manufacturing goods indexed with $j \in [0, 1]$ that are produced from an input $H = \varepsilon L$

(ε converts raw labor units L into efficiency units H) with productivities as in Eaton and Kortum (2002): one unit of H produces z units of manufactures, with z for each good $j \in [0, 1]$ drawn from the Fréchet distribution with parameters T and ϕ , as above. Assuming no transportation costs for either agriculture or manufacturing, and letting agriculture serve as numeraire, then the wage is $w = \lambda F'(L_A)$. The cost of producing a manufacturing good with productivity draw z is $w/\varepsilon z$, and, analogous to the expression in Eq. (9), the share of manufactures that will be exported by a country with wage w , labor efficiency ε , and technology parameter T , is

$$\frac{(w/\varepsilon)^{-\phi T}}{\sum (w_l/\varepsilon_l)^{-\phi T_l}}$$

Consider now two countries that differ only in λ , ε , or T . The country with higher λ will exhibit a higher wage, but will also devote more resources to agriculture and less to manufacturing, with a smaller share of manufacturing goods exported and a lower level of diversification (i.e., the number of goods exported is lower). This is a case of “good concentration.” On the other hand, a country with a lower ε or a lower T would also have a higher share of labor in agriculture, with higher concentration, but in this case the equilibrium wage would be lower. This is a case of “bad concentration.”

Now let us think about the determinants of T . In a dynamic setting, T can be seen as the stock of ideas (see Eaton & Kortum, 2001). Imagine that there is a worldwide stock of ideas, T^* , that grows at rate n . Discovery can be seen as the process by which a country adopts these worldwide ideas to the national environment. This is not exactly as in Hausmann and Rodrik (2003), where productivity is determined *ex ante*, and experimentation simply reveals what that productivity is. But the basic implications are the same. If x is the rate of discovery, then $\dot{T} = xT^*$, and in steady state $T/T^* = x/n$. Thus, a higher rate of discovery leads to diversification and a higher equilibrium wage.²⁵

The rate of discovery depends on its cost and associated private returns. If there are knowledge spillovers, as in Hausmann and Rodrik (2003), so that once an entrepreneur adopts a foreign idea then it diffuses rapidly to others in the economy, then the market by itself would lead to a suboptimal level of discovery and a level of diversification that would be too low. Policies to encourage discovery would lead to more diversification and higher welfare.

2.5 National and global gains from IP

The gains from IP for a country could come at the expense of other countries, or they could lead to global efficiency gains. We now explore the nature of the gains from IP for the different models discussed above.

Consider first the case in which the existence of Marshallian externalities leads to multiple equilibria. Above we considered a small country, but to explore whether gains

are national or international we have to allow for the possibility that a country's IP hurts other countries. Consider a model with two countries (A and B) and two goods (1 and 2). As above, good 2 exhibits Marshallian externalities. In this context, IP entails a shift from an equilibrium in which countries are *not* specialized according to their comparative advantage to one in which they are. Clearly such a shift increases worldwide efficiency. What happens to welfare in each of the two countries? First note that the country that implements IP (country A) experiences an increase in productivity in industry 2. If the terms of trade did not change, country A would be better off (as in Section 2.1.2). But if A is not small, then the relative price of good 2 will fall, and this will erode some of A's productivity gains. Even if A was so large that the terms of trade were now equal to its domestic prices, however, it would still gain, so country A necessarily gains from IP.²⁶ This is a simple application of Helpman and Krugman's (1985) proposition that a country necessarily gains from a move to an equilibrium in which it allocates more resources to a sector with positive externalities. For the other country (country B) welfare can decrease. To see this, note that for this country the effect of IP is exactly the same as the effect of an increase in productivity in industry 2 in country A. This could be either positive or negative for country B. For example, if the opportunity cost of good 2 in country A is just barely lower than the opportunity cost of good 2 in country B when both countries fully exploit the Marshallian externalities, then it is clear that country B loses from the IP implemented in country A.

Different results arise when IP is geared toward the appropriation of rents or broad (interindustry) externalities associated with certain goods, as in the literature on strategic trade policy (see Brander, 1995). In this case, the use of IP by different countries entails essentially a zero-sum game, or worse, as it could lead to an allocation of resources in which countries are not specialized according to comparative advantage.

Consider next the model where IP entails industry-specific collective action. If collective action is costless, then the answer is very simple: IP necessarily increases global welfare, although it may hurt another country by worsening its terms of trade. Things become more interesting when collective action entails costs. If countries bargain efficiently among themselves regarding trade barriers, then (by definition) only IP that increases global welfare will be implemented. One could optimistically argue that global negotiations that include trade and other policies actually take into account the ability of countries to implement IP, and that the resulting agreements allow only policies that increase global welfare. In this case, countries that can follow IP that increases their own welfare but lowers global welfare would be paid (perhaps through better trade access to other countries) not to do it, and countries that can implement IP that lowers national welfare but increases global welfare would be paid to do it. In fact, this is the way that trade agreements work in Bagwell and Staiger's (1998) model of GATT.

Bagwell and Staiger show that as long as countries choose their policies while disregarding their effects on international prices, then they will lead to a globally efficient

outcome. If international agreements are not efficient in this sense, however, then countries would evaluate the impact of IP on welfare while taking into account its effect on international prices. Clearly, this could lead to a decline in global welfare. For example, imagine that countries have committed to free trade. If IP for some good X implies a reduction in the supply of some other good Y then the price of this good would increase. If IP just barely makes the Home country implementing it better off, then for the world as a whole (which necessarily sees a decline in TOT) this entails a loss. Note, however, that this is a general feature of domestic policies implemented by countries that affect international prices, and not something particular to IP.

Finally, we turn to the model where IP aims at promoting diversification. As in the cases discussed above, the problem for global welfare may arise from the influence of IP on international prices. Imagine that the cost of a policy to encourage “self-discovery” just barely justifies the associated benefits. If such benefits include an improvement in a country’s terms of trade (which would arise from the decline in the supply of the non-diversified good), then the global efficiency would decline. Again, global efficiency would necessarily increase from IP only if countries evaluate it while disregarding its impact on international prices.

3. EMPIRICAL EVIDENCE ON INFANT-INDUSTRY PROTECTION AND OTHER FORMS OF IP

Since the best known form of IP is infant-industry protection, we devote most of this section to a critical review of the empirical literature on this type of IP. As discussed in Section 2, the theoretical justification for infant-industry protection relies on the existence of Marshallian externalities. There is extensive evidence that these externalities exist and are significant (Rosenthal & Strange, 2004). When such externalities exist, temporary protection may help the country switch toward a better equilibrium, and this may be welfare improving if the short-run costs are not too high. Protection may also be welfare enhancing if the protected industry generates positive (interindustry) externalities to the rest of the economy.²⁷

As we outlined in Section 2, such a policy would be welfare enhancing only if it passes both the Mill and Bastable tests. Recall that the Mill test requires that the protected sector can eventually survive international competition without protection, whereas the Bastable test requires that the discounted future benefits compensate the present costs of protection. We emphasize in our review that very few studies of IP have examined whether industries pass either the Mill or the Bastable test. Consequently, even if we could show that protected sectors grow faster, this is not sufficient evidence to claim that IP is welfare-enhancing.

Even if the conditions necessary for protection to increase welfare are satisfied, protection is not the most efficient policy. The theory of targeting (see Bhagwati &

Ramaswami, 1963) suggests that a government's objectives can be met more efficiently using instruments other than tariffs, and that when these instruments target the distortion at the source, a tariff of zero maximizes welfare for a small economy. Unless there are specific distortions or externalities associated with trade, the most efficient policy instrument to conduct industrial policy is not likely to be tariffs. In a recent review on the fiscal implications of trade reform, Pelzman and Shoham (2006) point out that "in the theoretical public finance literature it is well established that an optimal policy for a small open economy is to reduce tariffs to zero and raise consumption taxes (see Diamond-Mirrlees, 1971; Dixit, 1985), thus maintaining production efficiency" (p. 9).

In practice, however, there are several reasons why countries have continued to use tariffs to promote domestic industry. The most direct policy instruments are frequently not available in countries with limited abilities to collect income, consumption, or production taxes. Anderson (1996) shows that in a budget constrained economy, where it is difficult to compensate tariff cuts with increases in consumption or other taxes, tariff reductions lead to a curtailment of government spending and a resulting under-provision of public goods, which lowers welfare. Irwin (2007) compares the deadweight losses per dollar of revenue raised by tariffs with other forms of taxation in the United States. He argues that although the deadweight losses (per dollar) for tariffs were high in the nineteenth century, "import duties were probably much easier to collect and enforce . . . than other modes of taxation" (p. 19). Irwin (2002) draws similar conclusions for Argentina and Canada at that time. For the same reasons, it could be argued that in some developing countries trade protection could be an effective policy tool to implement IP. An important implication is that if fiscal considerations are the reason to use trade protection rather than production subsidies, then clearly tariffs would be the efficient policy and not quantitative restrictions.

In the rest of this section we review the empirical research on the effectiveness of infant-industry protection. We begin by discussing case studies where protection was clearly motivated by infant-industry considerations and then move on to other approaches that exploit the cross-industry and cross-country variation in trade barriers to see whether protection had the consequences predicted by infant-industry models.

3.1 Single-industry studies

There are very few detailed evaluations of infant-industry protection. Some important papers that explicitly take into account learning effects include Baldwin and Krugman (1986, 1989), Hansen, Jensen, and Madsen (2003), Head (1994), Irwin (2000), and Luzio and Greenstein (1995). Baldwin and Krugman (1986) study protection to the semiconductor industry in Japan. They use a simulation model to show that the Japanese semiconductor industry in Japan could not have emerged as a global player without the protected domestic market. Protection was needed to achieve the kinds of economies of scale and learning effects that would allow the industry to move down

its marginal cost curve and be competitive on world markets. However, Baldwin and Krugman (1986) also find that the costs to Japanese consumers outweighed the benefits, leading to net welfare losses for both Japan and the United States. Thus, although the semiconductor sector in Japan satisfied the Mill test, it did not satisfy the Bastable test.

Baldwin and Krugman (1989) estimate the impact on US and European welfare of Airbus's entry into the imperfectly competitive aircraft model. They show that subsidies to Airbus may have resulted in net welfare gains for Europe, primarily due to the high degree of imperfect competition (and monopoly rents) that characterized the industry. However, their simulation also makes clear that these results depend heavily on the assumed parameters, including the elasticity of demand. In any case, it is possible to evaluate that sector in such a way that the European subsidy to Airbus passes both the Mill and the Bastable test.

Head (1994) studies the effect of tariff protection on the emergence of the steel rail industry in the United States. This case fits the infant-industry protection view almost perfectly: the local industry was initially uncompetitive (1860s), but a few decades after the imposition of an import tariff the United States was the world leader in this market and the duty was repealed. Head concludes that "the domestic industry did 'grow up' and the duty was eventually removed. Hence, protection certainly did not cause stagnation and gross inefficiencies. Furthermore, the duty led to long-run reductions in domestic prices. While the savings to railroad builders were too small and came too late to yield a net gain to consumers, the overall effect on welfare appears to have been positive" (p. 163).

Hansen et al. (2003) examine the effect of production subsidies in Denmark for the production of electricity from wind power. They conclude that the subsidies elicited strong learning by doing in the industry, which achieved a dominant position in the world market. Moreover, according to their calculations, the direct and indirect (environmental) benefits outweighed the overall costs of the policy.

Irwin (2000) evaluates the effects of protection in the tinplate industry in the United States. The industry flourished after receiving tariff protection in 1890. Whereas there were no US producers at all prior to the imposition of the tariff, after the imposition of the tariff (at rates exceeding 70%) the industry became entirely self-sufficient. According to his counterfactual simulations, the tariff accelerated the industry's development by about 10 years, which would have developed anyway due to falling costs of iron ore. However, the costs to consumer surplus were so large that welfare declined as a consequence of protection. Irwin concludes that a lower tariff of around 50% could have improved welfare, but that the actual tariff imposed exceeded the optimal level and actually decreased welfare.

All the previous studies are for cases of protection or subsidies in developed countries. One single-industry study of infant-industry protection in a developing country is that of Luzio and Greenstein (1995), who study the effects of protection

of the microcomputer industry in the 1980s in Brazil. They show that although there was rapid productivity growth in the protected industry, it never caught up with the also rapidly growing technological frontier. As a result, welfare declined by a significant amount (around 20% of domestic spending on microcomputers) and the policy was abandoned in the early 1990s.

More studies like these analyzing the welfare implications of infant-industry protection would be very useful. Yet even this brief review makes it clear that protection may lead to higher growth but result in net welfare losses. For tinplate, steel rail, wind power, semiconductors, and aircraft, protection allowed domestic producers to grow and eventually become world class producers. Yet for tinplate, semiconductors, microcomputers, and possibly aircraft, protection led to net welfare losses. These case studies suggest that designing policies that increase welfare is very difficult.

3.2 Cross-industry studies

The theoretical framework makes a number of predictions for cross-country empirical studies. If the conditions necessary for infant-industry protection to improve welfare are satisfied (see Section 2), then protected sectors should experience faster productivity growth than nonprotected sectors. With time, protected sectors should increase their importance in the economy, and eventually become exporters. Under these conditions, one would expect to find a positive correlation between trade protection and productivity growth.

One of the first studies to look for a correlation between trade protection and productivity growth is Krueger and Tuncer (1982). Using cross-industry data on protection and productivity growth for Turkey in the period 1963–1976, these authors conclude that the empirical evidence does not provide support for the infant-industry argument. However, Harrison (1994) uses their data to show that more protected sectors did in fact exhibit higher productivity growth. As pointed out by Harrison, “Krueger and Tuncer (1982) applied *no* statistical tests to support their conclusion . . . If one runs simple correlations . . . one obtains striking results: Krueger and Tuncer’s data show a statistically significant *positive* relationship between increased protection and higher productivity growth. In contrast to their stated conclusions, it is possible to show that, in Turkey, protected industries did in fact achieve greater productivity gains during the sample period.”

While Krueger and Tuncer’s results appear to show some support for infant-industry protection, most studies find little support. Cross-industry studies usually show that the removal of protection generates both intrafirm and intraindustry productivity gains (possibly through market share reallocations, just as in Melitz, 2003). This includes work by Pavcnik (2002) for Mexico, Tybout and Westbrook (1995) for Mexico, Harrison (1994) for Cote d’Ivoire, Nishimizu and Page (1982) for Yugoslavia, Kim (2000) for South Korea, Topalova (2004) for India, Muendler (2004) for Brazil, Beason and Weinstein (1996) for Japan, and others.

Rodrik (2007) has criticized the use of cross-industry studies to test for the success of IP on the grounds that if IP is designed to deal with market failures that impede sector growth, then one should not be surprised to find a negative correlation between protection and growth. He assumes that $g_i = (1-\theta_i)A$, where g_i is the growth rate of industry i , θ_i is an index of market failures, and A is a parameter that captures productivity growth that is common across industries. In this framework, industries with stronger market failures would exhibit slower productivity growth. If there are political or fiscal costs associated with the promotion of an industry then in equilibrium one would find industries with a higher θ_i exhibiting stronger promotion and lower growth. Rodrik's point is that lower growth could be perfectly consistent with a successful IP, just as it is consistent with a politically motivated policy of promoting sunset sectors. Rodrik's argument may be correct for certain types of IP, but not for IP *associated with the infant-industry argument* (as defined in Section 2), since this would generate a positive correlation between protection and productivity growth. Even if Rodrik were correct in the short run, we would expect that in the longer term successful examples of infant-industry protection would lead eventually lead to growth.

In any case, there is a more fundamental problem with existing tests of infant-industry protection. There is no evidence to suggest that intervention for IP reasons in trade even exists. If intervention were motivated by IP reasons, we would expect the pattern of protection to be skewed toward activities where positive externalities or market failures are largest. Instead, existing evidence suggests that protection is motivated by optimal tariff considerations (Broda, Limao, & Weinstein, 2006), for revenue generation, or to protect special interests (Gawande, Krishna, & Olarreaga, 2005; Goldberg & Maggi, 1999). Tariff protection is frequently granted to less successful firms or declining industries. Beason and Weinstein (1996) study the pattern of industrial targeting in Japan and specifically ask whether the government targeted industries with increasing returns or emerging sectors such as electrical machinery. They find the opposite result: protection and other forms of targeting such as capital subsidies were highest for declining industries and industries without increasing returns. Most tariff protection was heavily concentrated in processed food and textiles, while most subsidized loans and tax relief went into mining. Beason and Weinstein conclude that "industrial policy considerations were dominated by the desire to aid declining sectors or protect the interests of large unproductive industries." Lee (1996) reaches similar conclusions for South Korea.

Harrison and Hanson (1999) find that in Mexico in the 1980s the pattern of trade protection was skewed toward food processing and garments, presumably for political economy reasons since these were sectors where Mexico already had a comparative advantage. More recently, Mobarak and Purbasari (2006) use a database on firms granted import licenses for raw materials and commodities in Indonesia to show that politically connected firms are more likely to be granted protection. However, firms that export are significantly less likely to be granted support. This suggests that firms

most likely to succeed on world markets in Indonesia were in fact penalized by restricting their access to import licenses.

There are a number of other areas where additional research is needed. To properly test the type of model presented in Section 2, we would expect initial output and productivity gains from the imposition of tariffs; later removal of these tariffs would generate no effects. We are not aware of any studies that test for these asymmetric effects of imposing protection and later removing tariffs through trade reform. Nor do existing empirical studies address the many other reasons why infant-industry protection may not work. Domestic demand in LDCs may lack the level of sophistication that would induce firms to meet the quality standards necessary for penetrating richer markets (Porter, 1990). Porter also argues that weak competition in small protected markets may not provide the incentives that firms need to upgrade their technologies and increase their efficiency. Small markets may also fail to reach a critical mass where Marshallian externalities are fully exploited. Moreover, as mentioned above, industries that use simple production methods may expand without the generation of any externalities, even if those same industries do exhibit strong Marshallian externalities in rich countries (Baldwin, 1969; Rodríguez-Clare, 2007). Firms may believe that protection will fail to create a cluster, and so they may decide not adopt the production methods that lead to externalities. In other cases protection may actually favor the use of backward technologies that do not contribute to the generation of higher industry-level productivity (Sauré, 2007).

The realization of Marshallian externalities is a much more complex process than what the model in Section 2 suggests. As argued by Rosenthal and Strange (2005), “there are many different aspects of a location that may matter to firms. A well-intentioned policy could easily fail because it failed to attend to one or two of these . . . it may not be possible to duplicate elsewhere the circumstances that led to a successful agglomeration in another place. . . This is not to say that government policy has never contributed to the formation of clusters. It certainly has, but the formation of clusters has been a side-effect rather than the primary goal of the policy . . .” (p. 19). In other words, the econometric evidence regarding Marshallian externalities may in fact be telling us that agglomeration may be *necessary but not sufficient* for increased productivity. If a certain factor, policy or institution is necessary for geographic concentration of an industry to lead to higher productivity, then we may observe Marshallian externalities taking place in advanced countries, but not in LDCs. To put it crudely, subsidizing the software sector may not generate a Silicon Valley in a developing country.

3.3 Cross-country studies

There are a number of cross-country studies evaluating the success of IP. These can be grouped into (1) testing for an association between protection and country performance and (2) qualitative case study evidence. Within the first group, recent studies emphasize the importance of the *pattern of protection* in understanding possible linkages between IP

and economic growth. These include Chang (2002), Clemens and Williamson (2001), Irwin (2002), Lehmann and O'Rourke (2008), O'Rourke (2000), Estevadeordal and Taylor (2008), and Nunn and Trefler (2006).

Clemens and Williamson (2001) and O'Rourke (2000) both find a positive correlation between import tariffs and economic growth across countries during the late nineteenth century. These two studies hypothesize that protection was associated with growth because it allowed countries to accelerate the growth of what were then “emerging” sectors (industry) out of “declining sectors” (agriculture). These emerging sectors were characterized by learning effects and the kinds of Marshallian externalities modeled in Section 2. These new explanations for the observed positive relationship between growth and protection at the turn of the twentieth century could also be used to explain how first Britain, and then the United States, were able to emerge as economic leaders in conjunction with tariffs that were very high in the eighteenth and nineteenth centuries. Chang (2002) claimed that protection was essential in the transformation of the United States into an industrialized economy in the nineteenth century. Others (for one such view, see Sir Arthur Lewis, 1955) have argued that Latin America's relatively high growth rates during the 1960s and early 1970s in a time of high effective rates of protection was also no coincidence.

Irwin (2002) casts doubt on the meaning and robustness of such correlations, arguing that rather than a causal relationship from protection for IP reasons to growth, the positive correlation comes from the fact that a few fast growing countries (e.g., the United States, Canada, Argentina) imposed high tariffs as a means of raising fiscal revenues. Lehmann and O'Rourke (2008) are able to address this criticism by purging their results of tariffs imposed for revenue purposes. They find no relationship between tariffs imposed to raise revenue and economic growth.

As pointed out by Lehmann and O'Rourke (2008), *what you protect matters*. If the pattern of protection is skewed toward increasing returns sectors where there are important externalities, then IP would be much more likely to work than if protection is given to declining sectors or sectors without externalities. In their book analyzing whether openness to trade is likely to be associated with higher growth, Grossman and Helpman (1991) make a similar point. Lehmann and O'Rourke examine the pattern of protection and growth for a sample of developed countries during the period between 1875 and 1913. They find that while agricultural tariffs were negatively correlated with growth, industrial tariffs were positively correlated with growth. They do not, however, estimate the impact of protection on productivity growth, nor do they calculate the net welfare effects of these policies.

Two other recent studies that emphasize that it is the pattern of protection which matters and not the average level of protection include Estevadeordal and Taylor (2008) and Nunn and Trefler (2006). Nunn and Trefler find that countries which protect skill-intensive sectors grow more rapidly than countries which protect unskilled-labor-intensive industries. Estevadeordal and Taylor disaggregate tariffs into consumption,

capital goods, and intermediate goods tariffs. For the 1970s through the current decade, Estevadeordal and Taylor show that tariff protection affects growth more negatively if tariffs are on capital or intermediate goods. We continue and expand this discussion to include a more general evaluation of the literature on trade and growth in Section 4.

Another approach relies on qualitative case study evidence to contrast the apparent success of East Asia countries relative to Latin America and elsewhere in the use of industrial policy. There is significant debate over whether the use of a range of industrial policy instruments, including infant-industry protection, helped or hurt development in East Asia. A common view is that East Asian countries used export subsidies whereas Latin American countries used import tariffs, and that this explains part of the difference in performance in these two regions. In fact, East Asian countries used both import tariffs and export subsidies, and this created a setting in which the incentives were neutral regarding import substitution versus exports, although manufacturing as a whole enjoyed some net promotion (World Bank, 1993). China's policies over the last 25 years could similarly be described as using both import tariffs and export subsidies.

South Korea and Taiwan had tremendous rates of physical and human capital accumulation over the 1960s, 1970s, and 1980s, and this went together with rapidly changing comparative advantage toward capital intensive goods. Of course, the standard explanation of this experience is that capital accumulation caused by some exogenous factors led to a changing comparative advantage. But an alternative interpretation, consistent with the model presented in Section 2.1.2, is that such a structural transformation was not inevitable because of multiple equilibria. In particular, without protection or promotion of the capital intensive sectors, countries would have remained specialized in the sectors where they enjoyed a static comparative advantage; since these sectors were not capital intensive, then capital accumulation would not have taken place. Amsden (1989) and Wade (1990) have argued that IP was crucial for some East Asian countries to capitalize on their latent comparative advantage in advanced manufacturing.

It is clear that East Asian countries indeed pursued several policies to encourage particular sectors, such as production subsidies, subsidized credit, fiscal incentives, and trade protection. But what was the actual effect of these policies relative to what would have happened in their absence? Can IP be credited with bringing about the successful industrialization experienced in East Asia? One approach in the literature has been to check whether the industries that received most support are the ones that have grown most rapidly.²⁸ As we discussed above, protection is typically motivated by political or terms of trade reasons rather than prospects for higher growth through IP. In the late twentieth century, in contrast to the last century when industrial countries protected emerging industries, it appears that trade barriers were often designed to protect “sunset” industries rather than to encourage “sunrise” industries (see Noland & Pack, 2003).

3.4 Export subsidies and other forms of IP

We have focused our discussion of the empirical evidence regarding the effectiveness of IP on infant-industry protection. But clearly there are many other forms of IP: countries could subsidize exports across the board or in particular industries, they could impose differential taxes, as well as differentiated production, credit and R&D subsidies. Since a comprehensive review of all forms of IP is not possible, in the rest of this section we focus our discussion on export subsidies.

Consider again the case in which some sectors exhibit Marshallian externalities. An overall export subsidy would simply preserve the allocation associated with the current pattern of comparative advantage. If export subsidies are targeted to the sectors that exhibit Marshallian externalities, then they could also be effective in switching the economy to the equilibrium with higher welfare. Again, production subsidies are less distortionary than export subsidies but they impose stronger fiscal demands. Thus, for governments with great fiscal needs or where taxation is very distortionary at the margin, export subsidies could be a reasonable option, although import tariffs would be less costly.

In any case, the advantages of export subsidies relative to import tariffs in improving productivity are threefold: (1) that by promoting exports, a country makes sure that firms are subject to the “discipline of the international market,” which forces firms to become more productive; (2) that by subsidizing only exporting firms, a country effectively limits the subsidy to firms with high productivity; and (3) that domestic markets may be too small to allow the protected industry to reap the full benefits of Marshallian externalities. All of these arguments are relevant but require some qualification. First, the discipline of the international market applies both to firms that export and to those that sell in domestic markets as long as there are no quantitative restrictions. Second, there is in principle no reason to subsidize highly productive firms over low productivity firms (see Demidova & Rodríguez-Clare, 2008), unless there are barriers that prevent resources from flowing from the latter to the former, in which case the most efficient policy would be to remove those barriers. Finally, if the economy is small in relation to the industry size needed to fully exploit the Marshallian externalities, this is not going to be fixed by an export subsidy. A different and more reasonable argument is that domestic demand is not sufficiently sophisticated, hence firms selling to domestic consumers will not develop the necessary level of sophistication needed for success in international markets. Export orientation for infant industries would avoid this problem.²⁹

One could redefine the case for industrial policy as trying to change incentives to produce (or export) some goods and not others. What evidence is there that what a country exports (or imports) matters? There are a number of studies listed in Appendix Table 1 which suggest that the growth effects of openness hinge on the composition of trade. These include An and Iyigun (2004), Choudhri and Hakura (2000), Dodaro (1991), Giles, Giles and McCann (1992), Ghatak, Milner, and Utkulu (1997), Hansen

(1994), Khalafalla and Webb (2001), Pineres and Ferrantino (1997), and Xu and Wang (1999). With one exception, all these studies find that exports are more likely to lead to growth if they are in nontraditional sectors such as manufacturing or skill-intensive goods rather than primary products or raw materials; studies also find that greater export diversification is more likely to be associated with growth.

Hausmann, Hwang, and Rodrik (2005) develop a measure of the sophistication of a country's exports based on the level of GDP *per capita* associated with the export of different goods worldwide. Once they have attached a level of GDP *per capita* to each detailed export category based on which countries export the good, they can then derive the implied GDP level of a country's exports. Countries whose export baskets contain items typically produced by higher income countries have a "high" level of export sophistication relative to what would be expected given their level of development. In principle, one could construct a measure of the distance between the implied income level of a country's exports and the country's actual level of GDP *per capita*. In a related paper, Rodrik (2006) shows that China's exports in 1992 were associated with an income level more than six times higher than China's *per capita* GDP at that time. Rodrik suggests that the gap between the implied level of income of China's exports and China's actual GDP *per capita* is too large to have occurred naturally and is one outcome of China's activist industrial policy.

These studies suggesting that what a country exports is more important than how much the country exports for long-run growth are suggestive that IP has played an important role in country growth experiences, but are not conclusive. No researchers have identified an association between the existence of Marshallian externalities in particular sectors, industrial promotion of those sectors, consequent changes in the commodity composition of a country's export basket, and growth. Consequently, we can only loosely infer that rapid changes in the composition of a country's exports toward more sophisticated products could be indicative of IP.

4. TRADE AND ECONOMIC DEVELOPMENT

Countries intervene in trade for many reasons, including the desire to shift production toward sectors with positive externalities (industrial policy), to raise revenue, to affect terms of trade, and to satisfy special interests. We reviewed in Section 3 some of the studies that explicitly evaluate the success of IP. There is almost no research that has been able to isolate the effects of IP on welfare, taking into account all the other motives for protection. However, there is a large literature which estimates the reduced reform relationship between openness to trade and economic growth. In this section, we review nearly 200 prominent studies that use some kind of reduced form approach. We discuss the different measures of openness used in this literature, the datasets, the identification strategies, and the results.

Can we use this vast literature to cast light on the IP debate? In defining openness to trade, we make a distinction between trade volumes, such as exports or imports, and trade policy, such as tariff reductions or changes in quotas. Most of the studies we review below find a positive relationship between trade volumes and growth. While there are far fewer studies that evaluate the linkages between tariffs and growth, they tend to get insignificant or weak effects. This should not be surprising given the mixed motives for imposing tariff protection. We hypothesize that these different results imply that industrial policy is more likely to induce growth if it is “pro-trade.” Yet from a practical standpoint, it is difficult to envision a successful set of policies that are both pro-trade and protectionist; at a minimum, this would require policies that fully offset any antitrade biases. The positive correlation between trade shares and growth is consistent with our hypothesis that IP implemented via export promotion and encouragement of FDI has generally led to higher productivity and growth outcome relative to IP implemented through protection from trade. Before we turn to the literature on trade volumes, trade policies, and country outcomes, we begin with some stylized facts.

4.1 Stylized facts on trade policies, 1980–2004

If we use a conventional measure of trade volumes such as export shares in GDP, developing countries are now more integrated into the world economy than the industrial countries. Figure 2 shows that export shares for developing countries overtook industrial countries in the early 1990s.

Table 2 shows the evolution of country trade policies over the last 25 years. Countries are ranked according to the net change in statutory tariff levels between 1985 and 2004. We have included both developing and developed countries to provide an indication of how protection changes at different stages of development. The first

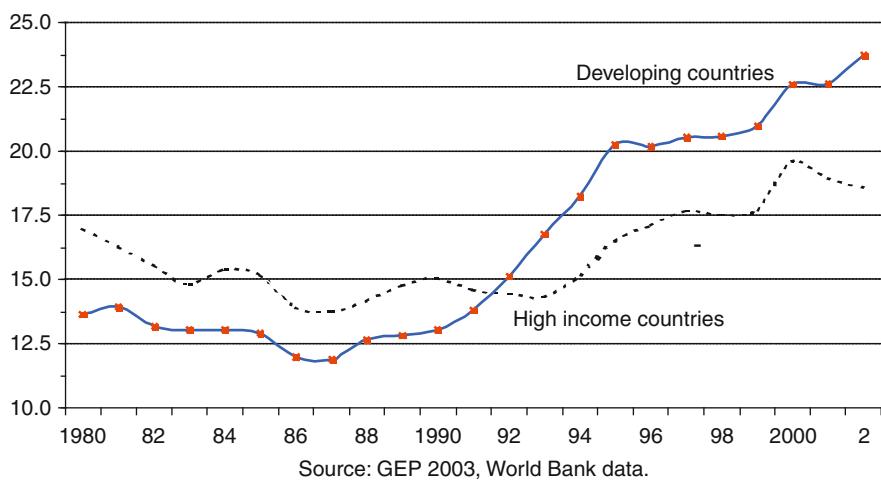


Figure 2 Export shares in developing and high-income countries.

Table 2 The evolution of country trade policies for 1980 to 2004

Countries	Revenue Tariffs (RT)					Statutory Tariffs (ST)					Changes 1985-2004		
	1980	1985	1990	2000	2004	1980	1985	1990	2000	2004	Changes in RT	Changes in ST	Change in Standard Deviation of ST
India	15.5	24.2	21.1	8.4	5.7	74.3	98.8	79.2	32.5	28.1	-18.5	-70.7	-32.8
Bangladesh	13.4	17.9	12.1	7	8.2	99.9	86	102.2	21.6	16.4	-9.7	-69.6	-36.8
Paraguay	6	2.2	3.3	3.4	2.2	71	71.7	16	10.9	8.5	0	-63.2	-6.5
Pakistan	15.3	14.7	15.2	5.6	3.9	77.6	78	58.8	46.6	16.1	-10.8	-61.9	-14.8
Peru	10.6	8.3	3.9	4.5	3.4	57	64	36	13.4	9.6	-4.9	-54.4	-22.1
Nicaragua	8.7	7.4	4.3	1.8	1.3	54	54	24.8	3.2	4.9	-6.1	-49.1	-
Costa Rica	5.3	6.9	7	1.1	1.1	55	53	16.4	5.4	5.7	-5.8	-47.3	-5.4
Guatemala	4.8	7.5	3.6	2.5	2.7	50	50	23	7.2	5.1	-4.8	-44.9	-19
Brazil	10	3.2	3.7	4.2	4.2	-	55.6	30	14.4	13.1	1.0	-42.5	-10.3
Madagascar	8.5	-	14	10.2	10.3	31.5	46	36	-	5.2	-	-40.8	-
China	5.7	10	3.7	1.5	1.8	49.5	39.5	40.3	16.3	9.8	-8.2	-29.7	-22.1
Ecuador	7.2	6.2	4	5.2	5.2	-	37.7	37.1	12.9	11.5	-1	-26.2	-20.7
Senegal	11.4	8.3	11.8	8.3	9.3	-	39	35	12	13.4	1	-25.6	-
Chile	2.8	5.7	3.7	2.3	0.7	10	30	15	9	4.9	-5	-25.1	0
Turkey	6.3	3	2.8	0.7	0.4	44	26.6	22.7	7.1	2.6	-2.6	-24	-21
Egypt	13.1	12.1	5.9	7	6	47.4	42.8	33.5	20.5	18.9	-6.1	-23.9	-386.3

Continued

Table 2 The evolution of country trade policies for 1980 to 2004—Cont'd

Countries	Revenue Tariffs (RT)					Statutory Tariffs (ST)					Changes 1985-2004		
	1980	1985	1990	2000	2004	1980	1985	1990	2000	2004	Changes in RT	Changes in ST	Change in Standard Deviation of ST
Benin	10.6	—	—	—	—	37.4	—	12	14	—	-23.4	—	—
Uganda	3.1	11.6	10.8	11.4	9.8	—	30	30.1	8.3	6.7	-1.8	-23.3	—
Philippines	6.8	6.2	6.6	2.7	2.5	38	27.6	24.3	7.6	4.4	-3.7	-23.2	-1.9
Kenya	6.1	7.4	6.3	7.5	7.5	40.3	39.2	43.7	19.3	16.2	0.09	-23	-8
Colombia	7.8	7.5	6	2.5	2.3	33.5	33.6	30.4	11.7	11.3	-5.2	-22.3	-2.1
Mali	3.8	5	4.6	—	—	35	35	25	12	12.8	—	-22.2	—
Zambia	2.4	6.4	4.8	7.3	7.3	—	34.8	25.6	14.7	13.2	0.9	-21.6	—
Venezuela	3	9.1	2.2	3.1	1.8	—	32.9	30.6	13.5	12.2	-7.3	-20.7	-18.5
Uruguay	8.9	5.8	5.6	2	1.9	—	31	30.5	11.1	10.8	-3.9	-20.2	2.1
Sri Lanka	11.7	10.6	8.8	2.1	2.5	41.3	29	26.9	9.9	9.8	-8.1	-19.2	-16.2
Thailand	6.9	6.5	5.4	1.4	1.3	32.3	31.2	40.8	17	13.3	-5.2	-17.9	-7.2
Burundi	18.1	17	22.9	13.3	13.3	—	37.9	37	—	20	-3.7	-17.9	—
Tanzania	7.7	6.3	7.5	11.9	11.9	—	32	29.7	19.1	14.1	5.5	-17.9	—
Ghana	17.3	21.7	11.6	4.6	4.6	—	30	17.5	15.2	13.1	-17.1	-16.9	—
Argentina	9.5	12.7	9.9	3.1	7.4	27.8	27	20.5	12.6	11.4	-5.3	-15.6	-17.1
Mexico	17.6	2.6	2	1	1	30.9	30	11.1	16.2	14.6	-1.6	-15.4	2.3
South Korea	4.1	3.6	3.4	1.3	1.3	20.4	23	13.3	8.7	9.3	-2.3	-13.7	-7

Central African Rep.	10.6	—	11.5	8.6	8.6	—	32	32	18.6	18.9	—	—13.1	—
Congo, Rep.	3.8	—	8.2	1.1	1.6	—	32	—	17.6	19.6	—	—12.4	—
Syria	7.1	5.6	2.9	3.7	3.7	37.3	27	20.4	—	14.6	—1.9	—12.4	—
New Zealand	2.5	2	1.7	0.8	1.6	13	16	14.5	3.4	3.7	—.4	—12.3	—11.3
Bolivia	7.8	7	2.3	2.4	2	—	19.8	16.7	9.5	7.5	—5	—12.3	—
Nigeria	8.5	5.1	4	—	—	32.6	37	34.3	—	24.8	—	—12.2	—15.2
Indonesia	2.9	1.6	2.4	0.5	1	29	18.1	20.3	8.4	6.4	—.6	—11.7	—5.9
Mauritius	9.6	9.6	7.6	4.6	3.6	34.9	34.9	27.6	31	23.5	—6	—11.4	—63.7
Gabon	7.3	6.4	4.8	—	—	43.3	29.6	—	20.4	18.6	—	—11	—
Cote d'Ivoire	12.8	11.8	10.9	9.3	9.9	27.5	23.3	33	12	12.7	—1.9	—10.6	—
Congo, Dem. Rep.	10.3	8.4	9.1	6.4	5.5	23.6	22.4	20.7	—	13	—2.9	—9.4	—
Poland	—	8.6	6	1.2	1	—	13.6	11.7	10	4.3	—7.6	—9.3	0.90
Australia	3.6	3.2	3.1	1.3	1.6	14.5	13.3	14.2	5.8	5.1	—1.6	—8.2	—7.8
Austria	0.7	0.6	0.7	0.3	0.2	11.6	10	8.7	2.4	1.8	—0.4	—8.2	—5.2
Nepal	8.6	7.7	8.8	5.1	6.9	22.1	22.6	22.6	14.7	14.8	—0.79	—7.8	—
Jamaica	0.9	1.7	—	2.2	3	16.5	17	17	10.6	9.4	1.3	—7.6	—
Trinidad & Tobago	3.2	5.7	2.7	10.2	10.2	—	17	18.6	18.4	9.7	4.5	—7.3	4.7
Canada	2.4	1.7	1.2	0.3	0.3	11.2	10.5	9.4	3.9	3.7	—1.4	—6.8	—2.1

Continued

Table 2 The evolution of country trade policies for 1980 to 2004—Cont'd

Countries	Revenue Tariffs (RT)					Statutory Tariffs (ST)					Changes 1985-2004		
	1980	1985	1990	2000	2004	1980	1985	1990	2000	2004	Changes in RT	Changes in ST	Change in Standard Deviation of ST
Malaysia	7.7	5.7	3.2	1.6	0.6	10.6	13.6	13	9.2	7.4	-5.1	-6.2	20.8
Italy	0.4	0.3	0.3	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.1	-5.7	-2.3
Denmark	0.5	0.4	0.3	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.2	-5.7	-2.3
Belgium	0.3	0.3	0.4	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.1	-5.7	-2.3
Luxembourg	0.1	0.1	0.1	0.3	0.2	8.8	7.5	7.4	2.4	1.8	0.1	-5.7	-2.3
Netherlands	0.4	0.4	0.5	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.2	-5.7	-2.3
Germany	0.5	0.5	0.5	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.3	-5.7	-2.3
Ireland	3.5	3	3	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-2.8	-5.7	-2.3
Spain	2.7	3	1.5	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-2.8	-5.7	-2.3
France	0.5	0.4	0.4	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.2	-5.7	-2.3
Greece	3.2	0.7	0.6	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.5	-5.7	-2.3
United Kingdom	0.7	0.6	0.6	0.3	0.2	8.8	7.5	7.4	2.4	1.8	-0.4	-5.7	-2.3
Finland	0.8	0.4	0.6	0.3	0.2	10.4	7	7.7	2.4	1.8	-0.2	-5.2	-5.2
Norway	0.3	0.3	0.3	0.3	0.2	6.2	5.7	5.7	2.9	0.5	-0.1	-5.2	7.1
Guyana	1.8	3.9	—	—	—	—	17	—	17	12	—	-5	—
Belize	8.7	10.9	10.8	6.6	6.6	—	17	—	—	12.9	-4.3	-4.1	—
Malawi	6.6	8.8	5.7	8.1	8.1	—	16.7	15.2	19.6	12.9	-0.6	-3.8	—

Japan	0.9	0.8	0.9	1.5	1.5	9.5	6.7	6.9	4.5	2.9	0.7	-3.8	-0.8
United States	1.1	1.7	1.5	0.8	0.9	7.3	6.6	6.2	4	3.2	-0.8	-3.4	3.4
Algeria	-	-	-	5.5	7.6	11.7	21.7	24.6	24.6	18.4	-	-3.3	-
Iran	17	14.2	5.4	3.1	4.9	20.7	20.7	20.7	4.9	17.5	-9.2	-3.2	-
Sweden	0.7	0.3	0.4	0.3	0.2	7.5	4.1	4.7	2.4	1.8	-0.1	-2.3	0.09
Haiti	9.9	8	6.7	-	-	27.7	11.6	-	10	10	-	-1.6	-
Iceland	6.5	4.6	4	0.5	0.4	9.6	6.7	3.8	1.9	5.1	-4.2	-1.6	0.6
Namibia	-	-	7.3	12.5	8.5	9.4	6	11	7.2	4.5	-	-1.5	-1.3
Cyprus	4	4.4	4.2	1.3	1.3	17.1	17.6	16.5	16.4	16.4	-3.1	-1.2	-
Switzerland	2.4	2	1.9	0.3	0.3	4.6	4.4	4.4	0	3.2	-1.7	-1.2	0
Botswana	12.8	7.1	6.6	6.1	6.1	9.4	6	11	7.2	5	-1	-1	-1.3
Papua New Guinea	3.3	5.9	7.1	7.3	6.1	-	7	7	21.1	6.4	0.1	-0.5	-
Kuwait	-	-	1.2	1.1	1.3	-	4	-	-	3.5	-	-0.5	-
Singapore	0.4	0.3	0.1	0.1	0	0.3	0.3	0.4	0.4	0	-0.3	-0.3	-1.8
Hong Kong	0.5	0.6	0.4	0.3	0	0	0	0	0	0	-0.6	0	0
United Arab Emirates	0	0	0	0	0	-	4	-	4	4	0	0	-
Jordan	7.1	6.1	5	4.6	3.1	-	13.8	-	24	14.4	-3	0.5	-2.9
Tunisia	9	13.3	9.5	3.6	2.4	26.4	24	27.5	30.6	25.3	-10.9	1.2	2.5

Continued

Table 2 The evolution of country trade policies for 1980 to 2004—Cont'd

Countries	Revenue Tariffs (RT)					Statutory Tariffs (ST)					Changes 1985-2004		
	1980	1985	1990	2000	2004	1980	1985	1990	2000	2004	Changes in RT	Changes in ST	Change in Standard Deviation of ST
Barbados	3.7	3.4	3.6	—	3.1	—	17	—	19	19	-0.3	2	—
South Africa	1.2	1.4	2.2	1.5	1.6	9.4	6	11	8.5	8.5	0.2	2.5	0.3
Bahamas	8.1	9.3	9.6	11	11	29.8	32.3	—	35	35	1.7	2.7	—
Morocco	10.7	6.4	8.6	7.4	7.4	54	23.5	23.5	33.6	28.3	1	4.7	—
Oman	—	1.4	1	0.9	0.8	—	3	—	4.7	8	-0.6	5	—
Zimbabwe	1.7	8	9.2	7.3	7.3	10	8.7	10.1	21.4	15.9	-0.6	7.2	6.8
Romania	—	—	0.4	1.6	0.9	—	—	—	15.2	13.3	—	—	—
Guinea-Bissau	—	7.7	—	—	—	—	—	—	12	13.9	—	—	—
Ukraine	—	—	—	1	1.7	—	—	—	10.5	7.9	—	—	—
Sierra Leone	13.3	11.9	4	10	10	25.8	25.8	—	—	—	-1.9	—	—
Myanmar	14.2	17.6	19.4	19.8	15.2	—	—	—	—	4.4	-2.4	—	—
Rwanda	13.3	—	14.2	—	0.	—	—	—	—	8.3	—	—	—
Russia	—	—	—	4.6	9.8	—	—	—	12.6	10.3	—	—	—
Czech Rep.	—	—	—	0.5	0.1	—	—	—	6.5	5.1	—	—	—
Niger	8.4	9.5	9.8	—	0.	—	—	—	12	12.8	—	—	—
Albania	—	—	—	7.3	3.4	—	—	—	17	8.3	—	—	—
Togo	12.4	8.6	9.2	—	0.	—	—	—	12	14.4	—	—	—

Fiji	5.8	7.9	6.4	4.5	4.5	—	—	—	12.4	12.4	—3.4	—	—
El Salvador	6.2	7.1	4.1	1.5	1.6	48	—	21.1	7.4	5.4	—5.5	—	—
Croatia	—	—	—	2.6	0.7	—	—	—	6.7	4.1	—	—	—
Portugal	2.1	1.2	1.3	0.3	0.2	—	—	7.4	2.4	1.8	—1	—	—2.3
Vietnam	—	—	—	—	2.8	—	—	—	—	13.6	—	—	—
Honduras	6.7	—	—	2.9	2.9	41	—	20	7.9	5.2	—	—	—
Lithuania	—	—	—	0.3	0.1	—	—	—	5.4	1.2	—	—	—
Dominican Republic	9.8	6.5	6.3	7.2	6.5	—	—	28	19.5	10.3	0	—	—
Hungary	5	3.7	5.1	0.8	0.6	—	—	—	8.2	8.9	—3.1	—	3.7
Bahrain	0.9	—7	1.1	1.3	0.9	—	6	—	—	—	0.2	—	—
Panama	3.1	4.1	1.8	4	1.6	—	—	40	9.5	7.9	—2.5	—	—
Israel	5.1	2.9	—9	0.3	0.4	—	—	—	2	4.5	—2.5	—	—
Latvia	—	—	—	0.3	0.2	—	—	—	5.6	3.2	—	—	—
Taiwan	3.6	2.8	2.1	1.5	0.8	—	—	9.7	8.8	6.9	—2	—	—
Slovenia	—	—	—	0.8	0.3	—	—	—	11.8	3.9	—	—	—
Slovak Rep.	—	—	—	1	0.2	—	—	—	6	22.1	—	—	—
Cameroon	11	6.1	5.4	9.2	9.2	—	—	—	18.4	18.1	3.1	—	—
Malta	4.9	4.5	4.9	0.6	0.7	—	—	—	8.8	8.8	—3.8	—	—
Bulgaria	—	5.4	1.3	0.7	0.7	—	—	11.4	13.7	10.2	—4.7	—	—

four columns indicate the ratio of tariff revenues to trade flows, taken from the World Bank. The next four columns report the actual statutory tariffs, averaged across import categories between 1980 and 2004. The next two columns report changes in these tariffs between 1985 and 2004. The last column indicates the change in the standard deviation of statutory tariffs over time; this measure is an indication of the change in the dispersion of tariffs.

It is evident from Table 2 that there has been a dramatic decline in tariff protection among developing countries. Statutory tariffs in India declined from 98.8% in 1985 to 28.1% in 2004; in Bangladesh from 86% in 1985 to 16.4%; in Costa Rica from 53% to 5.7%; in China from 49.5% in 1980 to 9.8% in 2004; in Turkey from 44% to 2.6%, and in Chile from 30% to 4.9%. Not all countries had such enormous declines, however: in Algeria, average tariffs only declined slightly, from 21% to 18%.

There are several other features worth noting in Table 2. The last column of Table 2 shows that there has been a dramatic decline in tariff dispersion, as indicated by the fall in the standard deviation of tariffs. There is also a large discrepancy between tariff revenues as a percentage of trade flows (“revenue tariffs”) and actual statutory tariffs. To illustrate the difference, Table 2 shows that average statutory tariffs in India in 1985 were 98.8% but tariff collections as a percentage of trade were only 24.4%. In Paraguay, average tariffs were 71% in 1980, while tariff collections were 6%; in Costa Rica, tariffs were 55% and collections were 5.3%; in Chile, average tariffs in 1985 were 30% but tariff collections were 5.7%. The difference between these two measures reflects in part the role of tariff exemptions—typically state enterprises were exempt from paying tariffs, as were many exporters and foreign enterprises—as well as selective imposition of duties by customs officers and the negative impact of high tariffs on imports.

We also highlight the historically high levels of protection in China. China has arguably had the most spectacular success in integrating into the world economy in the last two decades. Yet in 1990, it was still one of the most protected economies in the world, with an average tariff rate of 40%. According to Table 2, in 1990 China was tied for fifth place in average levels of tariff protection, behind Bangladesh, India, Pakistan, and Kenya. The dispersion of tariff levels was also high, and the maximum tariff exceeded 200%. Figure 3, taken from Rodrik (2006) documents that much of China’s export surge occurred simultaneously with the imposition of high tariffs. Nevertheless, countries such as China and India—because they had such high tariffs to begin with—also exhibited the highest tariff reductions between 1985 and 2004.

4.2 Cross-country evidence on trade policies, trade volumes, productivity, and growth

A standard approach in the cross-country literature is to regress an outcome of interest for country i at time t (GDP growth, real GDP per worker, or total factor productivity (TFP) growth) on a preferred measure of openness and a set of controls Z ,

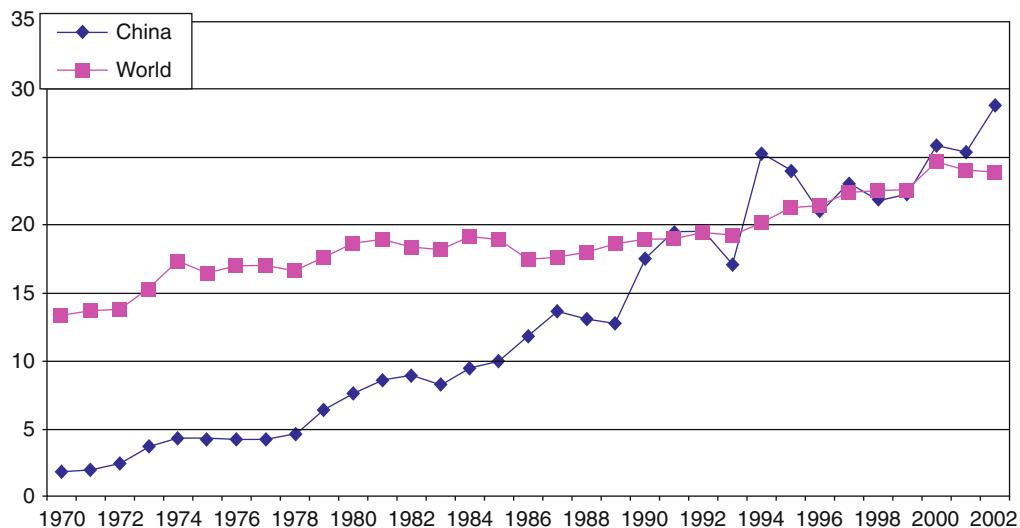


Figure 3 Exports as a share of GDP. Source: World development indicators database.

$$Y_{it} = \text{Constant} + \beta \text{OPENNESS}_{it} + \phi Z_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (10)$$

Most controversies have arisen over the following three issues:

1. How to measure OPENNESS.
2. How to account for the endogeneity between Y and OPENNESS.
3. Which variables to include in the set of controls Z .

(1) How to measure openness.

There is a large debate over how to measure openness. The ideal measures for understanding the linkages between trade *policies* and outcomes are measures of policies themselves—such as tariffs and quotas, but until recently, these measures were hardly ever used (see Harrison, 1996, for a discussion). How much of a problem is the lack of information on statutory tariffs (in contrast to revenue tariffs) in practice? If the difference between actual tariffs and revenues are due to exemptions in the tariff schedule, then the tariff schedule is misleading, and it would actually be better to use revenues as a share of import value. But if the differential between revenue tariffs and statutory tariffs highlighted in Table 2 reflects the restrictive impact of high barriers on trade volumes, barriers, or corrupt practices which impose rent-seeking costs not reflected in revenues, then using trade revenues to proxy for tariffs is not ideal.

A more fundamental problem which has plagued the literature on the relationship between trade policies and growth is the continued use of trade volumes as a proxy for policy. Trade volumes are affected by many different factors, including policies,

Table 3 All countries

	Correlation between different openness measures					
	Exchange rate	Revenue tariffs (Trade taxes/ Trade volumes)	Statutory Tariffs	Openness ($X + M/GDP$)	Real openness	DFI/GDP
Revenue tariffs (Trade taxes/ Trade volumes)	0.0452 283	1.0000 919				
Statutory Tariffs	-0.0314 274	0.6271* 666	1.0000 716			
Openness ($X + M/GDP$)	-0.0172 291	-0.2459* 561	-0.2470* 464	1.0000 607		
Real openness	-0.0662 293	-0.2200* 579	-0.2970* 461	0.8921* 566	1.0000 630	
DFI/GDP	-0.0066 281	-0.2501* 559	-0.3016* 445	0.3708* 567	0.4129* 576	1.0000 607

Notes: Data from World Bank. Time period includes 1970–2004. The asterisk indicates significant at the 5% level. Number of observations are underneath correlation coefficient.

distance to neighbors and trading partners, country size, exchange rate movements, terms of trade changes, and barriers to entry. Consequently, simply using trade volumes to proxy for changes in trade policies may be misleading. We evaluate the relationship between these different openness measures in Table 3. We present correlations between statutory tariffs, trade taxes as a percentage of trade, two measures of trade volumes, the nominal exchange rate, and the ratio of foreign investment inflows to GDP using annual data from 1980 to 2004. For trade shares, we include both nominal and real trade shares, where real trade shares are defined as the ratio of trade to GDP in constant prices from the Penn World Tables (version 6.1). Table 4 repeats the same exercise, but restricts the sample to developing countries. The correlations reported in Tables 3 and 4 highlight the following:

- Although Table 2 indicates a big difference in magnitude between the ratio of tariff revenues to trade and statutory tariffs, the correlation coefficient reported in Table 3 between the two measures is actually quite high at 0.63 and statistically significant.
- There is a significant negative correlation between trade shares and tariffs. The correlation with nominal openness is -0.25. The correlation with real openness about the same, between -0.2 and -0.3, depending on which measure of tariffs is used.

Table 4 Developing countries only

Correlation between Different Openness Measures					
	Exchange rate	Revenue tariffs Trade taxes/ Trade volumes	Statutory Tariffs	Openness ($X + M/GDP$)	Real openness DFI/GDP
Revenue tariffs (Trade taxes/ Trade volumes)	-0.0194 188	1.0000 435			
Tariffs	-0.1051 180	0.5863* 301	1.0000 323		
Openness ($X + M/GDP$)	0.0153 196	-0.3594* 392	0.3556* 313	1.0000 432	
Real openness	-0.0625 200	-0.2370* 423	-0.3473* 315	0.8379* 419	1.0000 470
DFI/GDP	0.0377 196	-0.2465* 406	-0.3291* 311	0.4568* 415	0.3685* 439
					1.0000 450

Notes: Data from World Bank. Time period includes 1970–2004. The asterisk indicates significant at the 5% level. Number of observations are underneath correlation coefficient.

- The negative correlation coefficient between trade policies (tariffs) and outcomes (trade shares) does not depend on how tariffs or trade shares are measured.
- Tariff levels are highly (negatively) correlated with the ratio of foreign investment inflows to GDP, and trade volumes are highly (positively) correlated with foreign investment inflows. In fact, trade flows are more highly correlated with foreign investment inflows than they are with tariffs. These correlations suggest that measures of openness may also be capturing the gains from foreign investment inflows.
- The correlations are the same or stronger if we restrict the sample to developing countries (see Table 4). Trade taxes as a share of trade flows continue to be highly correlated with actual tariffs. The correlation coefficient between trade shares and both tariff measures increases to (negative) 0.36.

These stylized facts suggest that trade taxes as a share of trade (revenue tariffs) are a much better proxy for statutory tariffs than trade shares. The correlation coefficient of statutory tariffs with revenue tariffs is significantly higher than the correlation of statutory tariffs with trade shares (0.70 versus -0.35). Yet researchers continue to rely on trade shares as a measure of trade policy, despite the easily available (World Bank or IMF) tariff

revenue measure. The other broad conclusion we can draw from these data is that there is a significant positive relationship between less restrictive trade policies and higher trade shares. Pritchett (1996) suggested that:

alternative objective measures of trade policy are completely uncorrelated across countries. This result has serious implications for empirical research that attempts to assess the effects of liberalization on economic performance using comparisons across countries; it also highlights the difficulties of interpretation in these types of empirical studies.

We would argue that this is not the case: statutory tariffs are highly correlated with revenue tariffs, indicating that they are excellent measures of trade policy. There is also a significant and negative correlation between tariff measures and outcome measures such as real or nominal trade shares in GDP. Nevertheless, the magnitude of the (inverse) correlation between trade shares and trade policies is not nearly large enough to allow proponents of free trade to argue that high trade shares always reflect a free trade stance.

Most studies listed in Appendix Table 1 use trade volumes as a measure of openness, but trade volumes are outcomes of trade policies as well as a host of factors including geography, shifts in terms of trade, exchange rate shocks, and changes in transport and communication costs. Much of the criticism in the important and widely cited Rodriguez and Rodrik (1999) NBER Macro Annual paper is directed at the inadequacy of typical proxies for openness. Rodriguez and Rodrik find fault with Dollar (1992), Edwards (1998), and Sachs and Warner (1995) for using exchange rate distortions as measures of trade policy, since exchange distortions reflect macroeconomic distortions, not trade policies *per se*. They also critique Edwards for using a World Bank classification of trade regimes which is subjective. While Dollar's openness measure seems ideal because it directly measures the deviation of domestic from international prices, Rodriguez and Rodrik argue the measure is primarily correlated with swings in the exchange rate. Dollar uses the following definition of openness:

$$\text{OPENNESS}_i = 100 \left(\frac{P_i}{E_i P_{\text{USA}}} \right)$$

OPENNESS is the relative price level compared to the United States, with all price levels converted to US dollars, using Summers-Heston country-specific consumption price indices. A higher price level should indicate a higher degree of distortions. Rodriguez and Rodrik (1999) argue that the law of one price does not hold in general, and that domestic prices could be high for reasons other than trade policy. These could include high transport costs or monopolies in distribution channels. In practice, Rodriguez and Rodrik show that there is no relationship between the openness measure

calculated by Dollar and actual tariffs or nontariff barriers. In fact, tariffs and nontariff barriers enter with the wrong sign if this measure of openness is regressed on tariffs and nontariff barriers. They then show that this measure is primarily capturing exchange rate movements.

Both Harrison and Hanson (1999) and Rodriguez and Rodrik (1999) critique a heavily used measure of openness created by Sachs and Warner (1995). The Sachs and Warner measure has been updated by Wacziarg and Welch (2003), but the updated measure may suffer from the same shortcomings as the Sachs and Warner measure. Harrison and Hanson show that the Sachs and Warner (1995) measure of OPENNESS does not pick up differences in trade policy but instead reflects differences across countries in exchange rate policies and political regimes. One way to reinterpret the evidence presented in Dollar (1992) and Sachs and Warner is that real exchange rate overvaluation is bad for growth, a theme recently emphasized by Rodrik (2007). One obvious implication for researchers is that any study which measures the impact of real price distortions on growth due to protection should also control for exchange rate movements.

(2) *Endogeneity problems.*

Endogeneity problems could arise for many reasons. Policy makers may prefer not to open up to trade until firms are capable of competing on world markets, suggesting that the causality runs from income to openness. Even statutory measures of trade policy (tariffs, quotas) are endogenously determined. The pattern of protection is likely to be skewed toward protecting weak sectors, promising infant industries, special interests, or vocal minorities.

While empirical work in the 1970s and 1980s largely ignored endogeneity problems, newer studies give much greater weight to constructing plausible identification strategies. This progress is evident in Appendix Table 1, which lists prominent studies on the linkages between openness and growth from the 1980s onward. Most of the early studies had no identification strategy at all, as indicated in column (4). More recent work addresses this omission, using one of two general approaches.

The first approach is to use granger-causality tests that exploit lags in studies that use time-series datasets. As indicated in Appendix Table 1, these studies often find that causality runs in the reverse direction, from Y to OPENNESS: more successful economies (or sectors) are more likely to open up to global competition. Related to this approach is the use of lags as instruments, which depends on some strong assumptions about the lack of correlation between the instruments and the error term.

The second general approach to identification has been to seek additional instruments for OPENNESS. One path-breaking study along these lines is Frankel and Romer (1999). Frankel and Romer use the insights from gravity models to derive an instrument based on geographic proximity. Gravity models predict that countries closer to each other trade more with each other. This means that distance can be used as an

instrument for bilateral trade. In the first-stage regressions, Frankel and Romer regress the log of country i 's trade with country j as a share of country i 's GDP on distance and other variables:

$$\ln(\tau_{ij}/GDP_i) = d'X_{ij} + \zeta_{ij}$$

The vector X includes the log of distance between country i and j , the log of population and area in both countries, and dummy variables indicating whether the two countries share a common border and whether they are landlocked. There are no subscripts for time in this specification: this is a pure cross-section using data for 1985. Frankel and Romer show that greater distance from a trading partner reduces bilateral trade, and they are able to explain 36% of bilateral trade in the first stage. Using the first stage estimates, Frankel and Romer then generate an OPENNESS variable by aggregating predicted bilateral trade with all of country i 's trading partners. In the second stage, Frankel and Romer regress log of income *per capita* in 1985 on the predicted trade share, log of population and log of area. They show that OPENNESS positively affects income *per capita*.

The beauty of this approach is that geographic proximity is without question exogenous with respect to income. There are several problems, however. Since distance does not change over time, the authors cannot allow for country-specific fixed effects α_i in Eq. (1) and are restricted to pure cross section estimation. While one solution in principle would be to control for factors that vary across countries but remain fixed over time—such as cultural or institutional differences—it may be difficult to control for all these omitted determinants of income. Another concern is that Frankel and Romer's original results are not very robust: the statistical significance on predicted openness disappears once we add continent dummies, which is not surprising since all the identification is from the cross section. Frankel and Romer also omit observations with zero bilateral trade in the first stage, which probably contributes to the poor first-stage R -square and the resulting weak instrument problem. Rodriguez and Rodrik (1999) are also critical of Frankel and Romer because they argue that greater openness to trade generated through geographic proximity may have different effects from trade generated through trade policy interventions. One further concern is that the instrumental variable estimates magnify the impact of trade on incomes, in contrast to what one would expect if trade is a positive function of income. The explanation given by Frankel and Romer is that the bias goes in the opposite direction because of measurement error, but one is still left wondering whether or not the authors have successfully addressed the endogeneity of trade to income.

Alcalá and Ciccone (2004) use the insights of Frankel and Romer to improve on their initial specification. They use all bilateral trade data available in the first stage, including those bilateral trade pairs with zero trade, which improves the first stage

F-statistic from 3.06 using Frankel and Romer's bilateral trade pairs to 11.66. This gives them two and a half times the number of observations in the first stage relative to Frankel and Romer. Consequently, the second stage relationship between their chosen measure of openness and their dependent variable *Y* (the log of PPP GDP *per capita* in 1985) is more robust. They also add a measure of institutional quality to the *Z* vector, which addresses the concern that trade is positively correlated with income or growth because greater openness is correlated with better institutions. They instrument institutional quality with language and settler mortality data, drawn from Hall and Jones, and Acemoglu, Johnson, and Robinson. Nevertheless, any analysis which uses geography as an instrument is still restricted to a pure cross section analysis, which requires the researcher to find all possible covariates which could induce a spurious correlation between OPENNESS and *Y*.

There are other aspects to Alcalá and Ciccone (2004) which suggest that the relationship between openness and income in a pure cross section is not very robust. Trade openness is only significantly correlated with *Y* if the authors use a "real" measure of openness, defined as the ratio of PPP trade to GDP. Nominal trade shares are not significantly associated with GDP *per capita*, which leads Rodrik, Subramanian, and Trebbi (2004) to suggest that Alcalá and Ciccone's results are driven by movements in the price level, not by trade. It is difficult to be sure, however, since Rodrik, Subramanian, and Trebbi do not use exactly the same specification as Alcalá and Ciccone.

Romalis (2007) suggests another clever instrument for a country's OPENNESS: tariffs imposed by a country's trading partners. In particular, Romalis uses US most-favored nation (MFN) tariffs as an instrument for developing country trade shares. Using this instrument, he shows that the change in real *per capita* GDP is positively and significantly affected by trade, and that the magnitude is economically important. Using MFN tariffs is particularly clever, since these are unlikely to be influenced by developing country behavior and are consequently exogenous. This is at the same time a limitation of the approach: the instrument only varies over time, not across countries since the United States must apply the same MFN tariffs to all its trading partners. The results also could be interpreted to suggest that other country policies matter for developing country growth. What Romalis shows is that access for developing country exports is beneficial for growth, but his research does not indicate whether opening up import-competing sectors to competition through reductions in protection is also beneficial for growth.

Two very recent studies which make an important contribution to resolving the endogeneity of trade and growth are Feyrer (2009) and Donaldson (2009). Feyrer (2009) suggests a solution to the problem faced by Frankel and Romer in using geography as an instrument. To overcome the fact that their instrument did not vary over time, which made it impossible to control for country fixed effects, Feyrer allows the

importance of air and sea distance across countries to vary over time. In the first stage of his estimation, Feyrer allows the coefficient on air and sea distance to be time-varying in explaining bilateral trade. He then uses those time varying coefficients to construct time and country-specific instruments for the first stage. Feyrer shows that trade has important, robust effects on growth using a panel of countries for 1960 through 1995. Donaldson (2009) uses historical data for 1861 through 1930 across Indian districts to estimate the impact of the introduction of railroads on trade costs, income volatility, price convergence, and economic growth. He traces out the mechanisms through which railroads affected income and shows that the reduction of trade costs through the introduction of railroads had large and important effects on economic growth.

(3) Which variables to include in the set of controls Z .

The third major area of controversy in this literature is which variables to include in the set of controls Z . There is a growing literature which claims that two key omitted variables from the Z vector are institutions and geography. Indeed a recent literature has sought to distinguish between institutions, economic geography, and trade as sources of economic growth, including Alcalá and Ciccone (2004), Easterly and Levine (2003), and Rodrik et al. (2004). Only Alcalá and Ciccone find that openness matters; the other two studies find that “institutions rule.” There are several reasons why Alcalá and Ciccone get very different results from Rodrik et al. First, Alcalá and Ciccone use real trade shares while Rodrik et al. use nominal trade shares as their measure of openness. Second, Alcalá and Ciccone improve upon the Frankel and Romer measure by expanding the first stage and using more countries, improving the first-stage F and reducing the fragility of the instrument.

None of these three studies, which have been extensively cited in the empirical literature on the determinants of growth, uses trade policy as a measure of openness. Easterly and Levine (2003) use the Sachs and Warner (1995) and Dollar (1992) measures to proxy for openness; the flaws of these two measures are discussed above. Rodrik et al. use the average of nominal trade shares for 1950–1998 as their openness measure. All three papers focus on a pure cross section of countries. As pointed out by Harrison (1996), trade policies and trade shares have changed too much over the last 40 years to make long-run averages very meaningful.

Given the problems inherent in the openness measures, and the reliance on pure cross-sectional estimation, it is not surprising that openness is trumped by institutions in two of these three studies. This research also highlights the tremendous problems associated with measuring institutions in a way which is distinct from trade policy. The correlation between the openness and institutions measures in Easterly and Levine (2003) is 0.68, which suggests that multicollinearity is likely to be a significant problem. Both Rodrik et al. and Alcalá and Ciccone use the Kaufmann, Kraay, and

Zoido-Lobaton measure of institutions, which is constructed from World Bank surveys based on responses for 1997–1998. Yet the dependent variables in these two studies are PPP GDP *per capita* prior to that period: 1985 PPP GDP per worker for Alcalá and Ciccone and PPP GDP *per capita* in 1995 for Rodrik et al. It seems odd to try to understand growth in 1985 or 1995 using measures of institutions based on data from the end of the 1990s, unless institutions change very little. Yet if institutions are not time-varying, then they may simply be capturing the country fixed effect α_i in Eq. (1).

While this survey has highlighted some of the shortcomings of cross-country work on openness and growth, there are several promising new areas of research which deserve mention. Most of the work surveyed so far uses real GDP *per capita* or *per capita* growth as a measure of Y . Yet a number of studies have suggested that openness is important because it allows countries to invest more. Levine and Renelt (1992) show that there is no robust relationship between different measures of openness and average *per capita* GDP growth in their cross-country sample. Replacing Y with investment shares in GDP, however, they find that openness is robustly correlated with investment rates. They conclude that “the relationship between trade and growth may be based on enhanced resource accumulation and not necessarily on the improved allocation of resources.”

Levine and Renelt show that trade matters for growth because it increases investment. One mechanism could be by reducing the price of investment goods. Delong and Summers show that countries with lower investment prices grow faster, and Lee (1995) shows that a higher share of imported capital goods in total investment is associated with higher growth. More recently, Hsieh and Klenow (2007) argue that on the contrary there is no link between lower relative prices of investment goods and trade policy. They cite as evidence the fact that the actual level of prices for investment goods in poor countries is not higher than in the rest of the world. Instead, they argue that investment rates are low in poor countries because the relative price of investment is high relative to nontraded consumption goods, such as services.

The importance of barriers to investment in understanding linkages between trade and growth is taken up once more by Estevadeordal and Taylor (2008). They estimate a version of Eq. (1) in differences, but separate their measure of openness into tariffs on consumption goods, intermediates, and capital goods. They also allow for a country fixed effects in differences, leading to a difference-in-difference specification for (1). They show that this approach successfully addresses the problem of whether institutions or trade policy is responsible for higher incomes, since in first differences there is no clear correlation between the two. In addition, they address the potential endogeneity of changes in openness by using as instruments for the change the level of tariffs in 1985 interacted with two measures of what they refer to as “GATT potential”: GATT membership in 1975 and a measure of diplomatic pressure constructed from number of diplomats. Estevadeordal and Taylor show that tariff protection affects growth more

negatively if tariffs are on capital goods or intermediates, which is consistent with Levine and Renelt's 1992 hypothesis that openness matters because it affects resource accumulation.

A second promising area of research is related to an emerging consensus on the need for openness to trade to be accompanied by key complementary policies. Recent research emphasizing the importance of complementarities between trade and other policies includes Chang, Kaltani, and Loayza (2005), Bolaky and Freund (2004), and DeJong and Ripoll (2006). One reason why the relationship in Eq. (1) may be fragile could be because openness to trade is most successful if implemented in conjunction with other policies which make it possible for firms to effectively compete on world markets. If such a policy can be characterized as X (there could be overlap between X and Z), then it would lead to a slightly different specification:

$$Y_{it} = \text{Constant} + \beta \text{OPENNESS}_{it} + \phi Z_{it} + \delta X_{it} \\ + \lambda (\text{OPENNESS} \times X)_{it} + \alpha_i + \tau_t + \varepsilon_{it}$$

Figure 4, taken from Bolaky and Freund (2004), makes this point graphically. Bolaky and Freund use three different measures of openness, including real and nominal trade

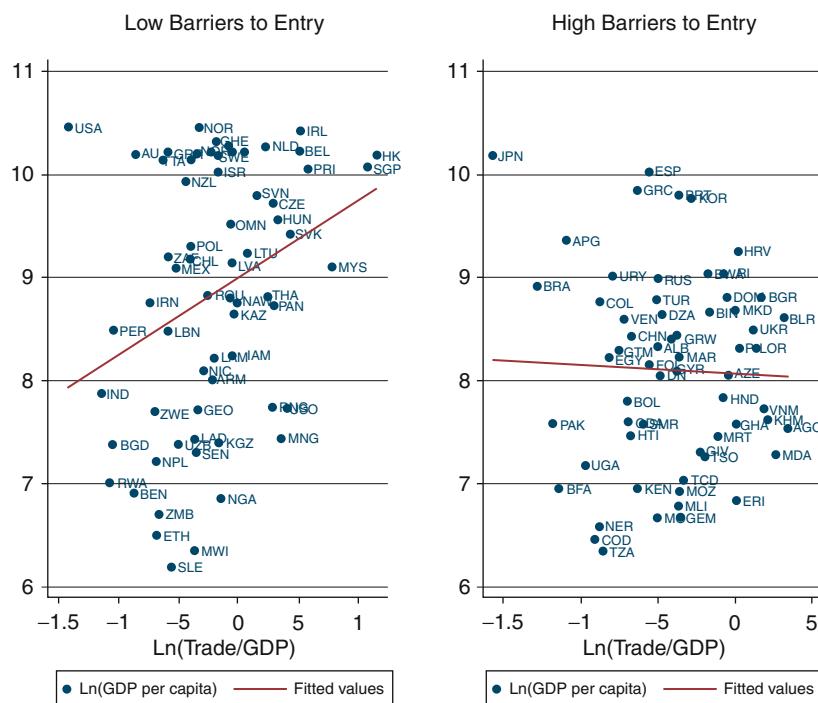


Figure 4 Trade, growth, and regulation of entry.

shares, and tariffs. In countries with low barriers to entry, there is a positive relationship between openness to trade and growth; in regulated economies, there is a negative relationship. The importance of other policies—in this case, low entry barriers—provides one explanation for why it is so difficult to find a robust relationship between openness to trade and good performance. There is simply too much heterogeneity in outcomes, in large part because other types of policies are so different. The necessity for openness to trade to be accompanied by low barriers to entry and exit can be understood in light of new trade theories that emphasize firm heterogeneity, as illustrated by Melitz (2003). In his model, gains from trade occur through reallocation of market share from less productive to more productive firms. If firms cannot easily expand or exit, this important source of productivity gains through trade reforms is lost.

Two other studies also find that gains from trade are contingent on other policies. Chang et al. (2005) use panel data instead of a cross section and trade shares corrected for country size are their measure of openness. In their work, the key complementary policies for ensuring that openness to trade is associated with growth include infrastructure development, labor market flexibility, and low barriers to entry. Given the current levels of those variables, they conclude that “there are many countries that currently stand to lose from opening their markets.” Chang et al. argue that other types of reforms are not so critical for ensuring growth gains from openness, including educational attainment, financial depth, and good governance. DeJong and Ripoll (2006), using tariffs as a measure of openness, also find that the effect of openness on growth is conditional on the level of income. In particular, using cross-country data for 1975–2000, they find a positive relationship between tariffs and growth rates for the world’s poorest countries, but a negative relationship for rich countries.

These three papers have several implications for trade policy. When eliminating both trade and other distortions simultaneously is not possible, these papers suggest that the benefits of trade reform depend on the existence and the degree of nontrade distortions and the feasibility of removing them. Packages tailored to the specific circumstances of each country may be more appropriate than a “one size fits all” policy. Another implication is that opening up to trade is not enough to ensure growth; in particular, key complementary reforms include lowering barriers to new firm entry, encouraging more flexible labor markets, and improving infrastructure.

4.3 What can the different effects of trade volumes versus trade policy tell us about the success of IP?

Our review of recent evidence on trade policies, trade volumes, and growth suggests that researchers face several challenges. In particular, measuring openness to trade, identifying the direction of causality between openness and growth, and adding

additional controls to include in cross-country estimation are ongoing concerns. Nevertheless, our review of the studies in Appendix Table 1 suggests that in the post-World War II period, tariffs on average have not been successful in generating higher growth. We confirm these general conclusions using a panel dataset on non-OECD countries, for the 1960-2000 period. In particular, we contrast the results using trade shares ($X + M/GDP$) as our openness measure relative to using the World Bank's revenue tariffs (tariff revenue divided by imports). The dependent variable is log income *per capita*. The use of a panel of annual observations allows us to control for country-specific fixed effects and also for time effects. Country fixed effects are one approach to controlling for country characteristics that vary across countries (such as institutions) that are not perfectly measured and do not vary systematically from one year to the next. Time effects allow us to control for worldwide shocks, such as an oil price shock or a worldwide currency crisis.

The top two rows in Table 5 show that openness measured using trade shares is positively and significantly associated with growth. These results are consistent with the majority of the studies listed in Appendix Table 1 that use trade shares as a measure of openness and generally find that changes in trade shares are associated with higher growth. However, the bottom two rows show that revenue tariffs are not significantly associated with growth. Even skeptics, such as Rodriguez (2007), generally conclude that there is a strong correlation between trade volumes and growth, while the association between trade policy—as measured by the World Bank's revenue tariff measure—and growth is weak.

The positive correlation between trade shares and growth is very strong and remains after we add other controls. The negative correlation between tariffs and growth is significant in some specifications but is not a robust finding. What should we make of this? As we discussed in Section 3, tariffs are imposed for many reasons. Foremost among these reasons are the need to raise revenue, political economy considerations, and infant-industry concerns. Either IP via protection has not increased growth, or protection on average has been imposed for other reasons, leading to no net gains in output.

A promising area for new research is to move beyond reduced form evidence on the linkages between openness and growth so that we can identify *how* openness to trade affects growth. This is particularly important from a policy perspective. If openness yields benefits because it allows firms to import new technology embodied in capital goods, the policy implications are quite different than if openness is beneficial because it forces firms to compete internationally or leads to market share reallocation toward more efficient firms. Identifying the mechanisms leading from openness to growth is precisely the focus of new micro-based studies of firms (see, e.g., Acemoglu, Antras, & Helpman, 2007 or Melitz, 2003). These theoretical advances have been accompanied by a growing empirical literature, to which we now turn.

Table 5 Incomes, trade shares, and import tariffs in a panel of non-OECD countries

Using trade shares as a measure of openness, 1960-2000								
	Dependent variable: ln income <i>per capita</i> , in PPP \$1993							
	(1) OLS No controls	(2) OLS No controls	(3) OLS Includes controls	(4) OLS Includes controls	(5) IV No controls	(6) IV No controls	(7) IV Includes controls	(8) IV Includes controls
3-Year lag trade share	0.907 (0.036)***	0.514 (0.037)***	0.214 (0.038)***	0.203 (0.035)***				
Trade share					0.978 (0.037)***	0.857 (0.057)***	0.426 (0.067)***	0.402 (0.064)***
Country fixed effects?	NO	YES	YES	YES	NO	YES	YES	YES
Time fixed effects	NO	NO	YES	YES	NO	NO	YES	YES
Observations	3294	3294	1996	2657	3288	3288	1996	2657

Continued

Table 5 Incomes, trade shares, and import tariffs in a panel of non-OECD countries—Cont'd

Using import tariffs as a measure of openness, 1960-2000								
Dependent variable: ln income <i>per capita</i> , in PPP \$1993								
	(1) OLS No controls	(2) OLS No controls	(3) OLS Includes controls	(4) OLS Includes controls	(5) IV No controls	(6) IV No controls	(7) IV Includes controls	(8) IV Includes controls
3-Year lag import tariff	-3.586 (0.377)***	-0.721 (0.142)***	-0.298 (0.117)**	-0.137 (0.119)				
Average import tariff					-4.830 (0.441)***	-4.830 (0.441)***	-0.635 (0.328)*	-0.338 (0.379)
Country fixed effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Time fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Observations	1617	1617	1261	1485	1415	1415	1125	1306

Notes: Source is Aisbett, Harrison, and Zwane (2005), as reported in Harrison (2007). Huber robust standard errors in parenthesis. Data are annual data, with one observation for each country and year. The dependent variable is the log of income *per capita*, in PPP \$1993. Trade share is the share of exports plus imports in GDP. Import tariffs are import revenues divided by the value of imports. Both are taken from the World Bank's indicators. OLS indicates ordinary least squares and IV indicates instrumental variables. All regressions exclude OECD high-income countries. Columns (3) and (7) include controls for inflation, government expenditure in GDP, currency crises, investment in GDP, and the fraction of the population that is literate. Columns (4) and (8) include controls for inflation, government expenditure, and currency crises. In instrumental variable regressions, trade share is instrumented using 3-year lagged value, and import tariff is instrumented using 3-year lagged value.

*Huber robust standard errors in parenthesis with *** indicating significance at 1%, ** indicating significance at 5% and * indicating significance at 10%.

4.4 Identifying the mechanisms for gains from trade

Much of the new research evaluates the importance of international trade for growth using micromodels of consumer and firm behavior. This new research focuses on the following mechanisms for understanding the linkages between openness to trade and growth: (1) gains from consumption of increased variety (2) gains from importing goods that embody new technology (3) gains from increasing competition (4) gains from reaping economies of scale (5) gains through reallocation of market shares to the most productive firms, and (6) learning by doing through exporting. We describe these in more detail below.

Gains from consumption of increased variety. Quantitative models are useful for measuring gains from trade coming from increased variety, as in Romer (1994). Feenstra (1994) and Romer showed that these gains could be large, while Arkolakis, Demidova, Klenow, and Rodríguez-Clare (2008) show that Romer's results are sensitive to modeling assumptions. Under heterogeneity, gains could be small, as the new varieties that are imported after liberalization are "marginal varieties," in the sense that total consumption of these goods is small. Moreover, the benefits from the increase in foreign varieties could be compensated by the losses associated with the displacement of domestic varieties. Another important paper in this literature is Broda and Weinstein (2006). Broda and Weinstein show that an important part of growth comes from the increase in imported variety over the last several decades. While this may in part result from a country's own trade reforms, it is also driven by diversification on the part of exporting countries like China.

Gains from importing goods that embody new technology. Eaton and Kortum (1999) argue that tariffs affect the price of capital, and through this they affect the capital-labor ratio in steady state. Coe and Helpman (1995), Keller (1998), and others reviewed in Keller (2004) study the role of trade as a vehicle for "international R&D spillovers." The idea is that by importing intermediate and capital goods, a country benefits from the R&D done in the exporting countries. This is a key feature of the model of R&D and trade in Eaton and Kortum (2001). A different notion is that trade accelerates the international flow of technical know-how (see Grossman and Helpman, p. 165). Several papers have explored this empirically with mixed results (see Aitken et. al., 1997; Bernard and Jensen, 1999; Clerides et. al., 1998; Rhee & Belot, 1990).

A number of theoretical papers have explored the role of intermediate inputs in raising productivity growth (Ethier, 1982; Grossman & Helpman, 1991; Markusen, 1989; Romer, 1986, 1990). Some recent studies find that increasing intermediate goods inputs or lowering input tariffs are associated with large productivity gains. This includes Amiti and Konings (2005), Broda, Greenfield, and Weinstein (2006),

Goldberg, Khandelwal, Pavcnik, and Topalova (2008), Halpern, Koren, and Szeidl (2005), and Kasahara and Rodrigue (2008). Goldberg et al. use Indian data to show that accounting for new imported varieties lowered the price index for intermediate goods by 4.7% per year relative to conventional gains from lower import prices. They also find that lower input tariffs account for a third of the new products introduced by domestic Indian firms, suggesting potentially large gains from trade.

Gains from increasing competition. Harrison (1994), Levinsohn (1993), and Muendler (2004) show that increasing competition due to lowering of trade barriers reduced price-cost margins, using microlevel data for Turkey, Cote d'Ivoire, and Brazil. Enhanced competition may encourage firms to engage more in innovative activity or may simply lead to lower prices for consumers. One important implication is that studies which measure the productivity gains from trade reforms without allowing markups to fall postreform are likely to mismeasure the productivity gains from trade, as pointed out by Harrison. Harrison shows that the direction of the bias in productivity measurement cannot be identified *a priori*. She also shows that a simple way to address this problem is to allow measured factor shares to vary with changes in trade policy when doing production function estimation.

Gains through reallocation of market shares to the most productive firms. New heterogeneous trade models suggest that trade enhances productivity by reallocating output toward more efficient firms. In the original framework developed by Melitz (2003), firms are endowed with different productivity draws, which are predetermined and unchanging over time. When a country opens up to international trade, only the more productive firms remain as the less productive firms are forced to exit.

There are a number of implications of this framework which could be tested using firm-level data. As we discussed earlier, these models which emphasize reallocation of firm shares imply a critical role for other complementary policies that make it easy for firms to expand, contract, enter, or exit the market. Another implication is that productivity gains from trade reforms should operate through market share reallocation. An early study for the United States by Baily, Hulten, and Campbell (1992) did not focus on trade policy *per se* but provided a framework for decomposing productivity growth into components due to within-firm changes versus reallocation of output. They concluded that the bulk of growth in aggregate TFP was accounted for by the reallocation of output shares. For developing countries, this decomposition has been performed by Pavcnik (2002) for Chile and by Van Biesebroeck (2003) for Colombia.

Since Pavcnik's sample begins at the end of the Chilean trade reforms, she cannot use changes in trade policy as her openness measure. Instead, she separates enterprises into import-competing, export-competing, and nontraded. She then examines

whether import- or export-competing firms had higher productivity relative to other types of firms. Using the Olley-Pakes (1993) approach to estimating productivity, Pavcnik finds that import-competing firms improved their productivity over time, while export-competing firms did not. This suggests that in Chile there was no learning among exporters, but import-competing firms did exhibit productivity growth. Pavcnik calculates that two thirds of productivity growth in Chile was due to reallocation of market shares toward more efficient producers, and the remaining one third was due to improved productivity among surviving firms.

Van Biesebroeck (2003) uses a variety of methods to compute productivity growth and finds that for Colombia the majority of changes in productivity occur within the same plant. Van Biesebroeck finds that between two thirds and three quarters of the total change comes from within-plant changes. The second most important effect is the entry of more productive plants into the economy. While Pavcnik (2002) finds that two thirds of the increases in productivity growth are due to reallocation of market shares, Van Biesebroeck finds the opposite: in Colombia, the bulk of aggregate increases in productivity growth are driven by within-plant changes in productivity.

One possible explanation for these different results is that entry and exit barriers (due to restrictions on hiring and firing) in Colombia were higher than in Chile during this period. Barriers to entry or exit make it difficult for a reallocation of market share toward more efficient producers to occur. More case studies are needed before we can reach any definitive conclusions. However, the limited evidence suggests that productivity growth stems from *both* (1) a reshuffling of production toward more efficient producers and (2) increasing productivity within the firm. Consequently, Melitz's (2003) simplifying assumption that firms receive an exogenous productivity draw which is unchanging over time is not consistent with actual firm behavior. The latest heterogeneous firm research modifies Melitz to allow firm-level productivity to evolve over time, instead of being fixed at the initial distribution. Atkeson and Burstein (2007a, 2007b, 2007c), Costantini and Melitz (2007), Lileeva and Trefler (2007), and Edington and McCalman (2008) all develop models with heterogeneous firms where productivity is allowed to evolve within the firm. Lileeva and Trefler allow reductions in foreign tariffs to induce lower productivity firms to invest in raising labor productivity, engage in more product innovation, and increase their adoption of advanced manufacturing technology, which in turn leads to within-firm productivity gains. Bernard, Redding, and Schott (2006) have also developed a model where firms produce multiple goods, and trade liberalization may lead them to focus on the goods in which they are most productive. Empirically this would be seen as an increase in productivity within firms, even though the mechanism is essentially the same as in Melitz.

Learning by doing through exporting. One likely channel through which international trade leads to productivity gains is through learning by doing for exporters. Yet a first round of empirical studies suggested that there was very little, if any, learning from exporting. Instead, these first generation studies found that the best firms select into exporting, leading to a strong positive correlation between productivity levels and export status. This research included case studies of Colombia (Clerides, Lach, & Tybout, 1998), Spain (Delgado, Farinas, & Ruano, 2002), Germany (Bernard & Wagner, 1997), and the United States (Bernard & Jensen, 1999). Pavcnik (2002) also found a similar result: firms operating in export-competing sectors are the most efficient in manufacturing, but these firms do not show productivity improvements over time.

This first-generation evidence is drawn from primarily high and upper middle income countries. A second wave of empirical studies casts doubt on the hypothesis that there is no learning by doing for exporters. For developing countries, this includes Aw, Chung, and Roberts (2000) for Korea, Van Bieseboeck (2005) for nine African countries, Lileeva (2004), Blalock and Gertler (2004) for Indonesia, Hallward-Driemeir, Iarossi, and Sokoloff (2002) for East Asian countries, Fernandes and Isigt (2007), Park, Yang, Shi, and Jiang (2006) for China, Aw, Roberts, and Winston (2007), and De Loecker (2007) for Slovenia.

All of these second generation studies find evidence of learning through exporting. Blalock and Gertler (2004), for example, test for whether exporting confers productivity gains using a panel of Indonesian manufacturing enterprises. They find strong evidence that firms experience a jump in productivity of 3–5% following the initiation of exporting. Van Bieseboeck (2005) examines the evidence in favor of learning by exporting in Africa and also finds a causal link from exporting to productivity. Aw, Roberts and Xu (2009) also find evidence of learning by exporting in the Taiwanese electronics industry.

Blalock and Gertler argue that previous tests of learning by exporting were done on industrialized or middle income developing countries. They point out that “while firms in developed and middle income countries are likely to be as efficient as those in their trading partners’ countries, firms in the poorest countries may have much more to gain from exposure to international export markets.” Lileeva and Trefler (2007) propose a similar explanation for these disparate findings using Canadian data. They show that lower productivity firms are more likely to invest and learn to access foreign markets, in contrast to higher productivity firms that are able to export without additional investment. De Loecker (2007) finds that for Slovenian firms, there was evidence of a different kind of selection into export markets: only the most productive firms began exporting. He then uses nonlinear matching estimation to test whether there was learning by exporting during the 1994–2000 period. He finds that while there was evidence of selection into export markets, these same exporters increased productivity as a result of exporting. The identification strategy uses both matching techniques and the

difference in productivity growth for those who began exporting this period relative to a control group that did not. De Loecker also finds that learning by exporting was higher for firms exporting to high-income destinations.

All of the identification problems present in Eq. (10) for the cross-country literature are challenges for these microstudies as well. How can we distinguish between selection into exporting (i.e., the most productive firms choose to become exporters) and the impact of exporting on learning and productivity? A number of studies exploit the panel nature of the data, testing whether firms were more productive prior to becoming exporters. This involves constructing an indicator for the period prior to when the firm becomes an exporter and testing whether it is statistically significant in a regression of productivity on export status. Another approach has been to find an instrument for export status. Van Biesebroeck (2005) uses as instruments the location of the firm, ethnicity of the owner, foreign ownership, and state ownership, although the validity of these instruments is not completely clear. Lileeva and Trefler (2007) use US tariff cuts mandated by the Canada-US Free Trade Agreement as an instrument for Canadian export status.

Despite ongoing controversies regarding the importance of learning by doing through exporting, it is probably safe to conclude the following:

1. The most productive firms are likely to become exporters.
2. While there is selection into exporting, there is also learning through exporting.
3. Learning from exporting is most likely in technologically backward countries and among less productive firms.

5. FOREIGN DIRECT INVESTMENT

Many countries encourage inward FDI because they expect that foreign firms will enable domestic enterprises to become technologically more advanced. This is nothing other than industrial policy, although it is rarely identified as such. Hanson (2001) reviews the many incentives offered to foreign firms, which include income tax holidays, tariff exemptions, and subsidies for infrastructure. In 1998, 103 countries offered tax concessions to foreign companies that set up production or other facilities within their borders. Some of these concessions were enormous: the Government of Alabama paid the equivalent of 150,000 dollars per employee to Mercedes for locating its new plant there in 1994; the British government provided between 30,000 and 50,000 dollars per employee to attract Samsung and Siemens in the late 1990s; Ireland offered until recently a corporate tax rate of 10% to all foreign manufacturers who chose to locate there (Gorg & Greenaway, 2004). Until 2008, China offered significantly lower corporate tax rates to foreign companies locating there, and continues to subsidize infrastructure investments for multinationals locating in foreign enterprise zones.

Most countries encourage FDI inflows to specific sectors; typically these are sectors which are technologically more advanced. Chandra and Kolavalli (2006) document that “attracting FDI has been an important strategy in technological adaptation” in a number of cases (p. 33) Table 6 reproduced from Alfaro and Charlton (2008) identifies the specific sectors targeted by 29 countries for the 1985–2000 period. The most targeted sectors include machinery, computers, telecommunications, and transport equipment.

Alfaro and Charlton find that FDI inflows are likely to be higher in targeted sectors, and that FDI in these sectors generates higher growth. Their study suggests some support for the argument that IP with respect to FDI has been successful at generating growth, possibly because governments appear to target sectors with positive externalities. Nevertheless, it is difficult to econometrically identify whether targeting has worked, since FDI may be attracted to sectors with high potential for growth.

Aggregate cross-country studies on FDI and growth are reviewed in greater detail in another chapter of this volume. Given the strong correlation between FDI inflows as a share of GDP and openness to trade reported in Tables 3 and 4, identifying a separate impact of FDI from trade on country outcomes is likely to be challenging. All of the problems associated with the cross-country evidence on trade and growth—measuring policies, identifying the direction of causality, and omitted variable bias—are present in this literature as well. Researchers typically use FDI flows, rather than policies targeted at FDI, to measure its effects, since panel data on FDI policies are even more difficult to obtain than data on tariffs. Reverse causality is also a problem, since extensive evidence suggests that FDI is attracted to large, less risky, and growing local markets. Nevertheless, there is generally mixed evidence on the relationship between FDI and a country’s growth (Borensztein, De Gregorio, & Lee, 1998; Bosworth & Collins, 1999; Carkovic & Levine, 2002). Recent work on FDI and growth suggests that other complementary policies need to be in place to maximize the gains from inward foreign investment, reinforcing the theme introduced in Section 4 on trade. Important policies that need to be in place for a country to benefit from FDI include a minimum level of human capital (Borensztein et al.), developed financial markets (Alfaro, Chanda, Kalemli-Ozcan, & Sayek, 2004) and openness to trade (Balasubramanyam, Salisu, & Dapsoford, 1996).

There is significant research interest in FDI as a vehicle through which developing country firms learn about new technology. While most of the empirical literature focuses on productivity effects of FDI, there is also a growing literature that examines the impact of FDI on factor markets. We review these two research areas below.

Table 6 The specific sectors targeted by 29 countries for the 1985-2000 period (from Alfaro & Charlton, 2008)

Industries	Code	Targeted by number of countries	Growth in value added	FDI/ value added	Share of value added	Dep. external finance	White high skill	Blue and white high skill
Agriculture and fishing	1	4	2.617	0.001	0.055	-0.297	0.040	0.707
Mining of metals	2	2	3.309	0.000	0.015	0.455	0.168	0.687
Extraction of petrochemicals	3	5	1.473	0.009	0.012	0.318	0.341	0.674
Food products	4	5	5.314	0.013	0.041	-1.717	0.074	0.427
Textile and wood activities	5	5	3.362	0.015	0.026	-0.946	0.145	0.594
Petroleum, chemical, rubber, plastic products	6	9	5.755	0.033	0.043	0.285	0.397	0.546
Metal and mechanical products	7	3	5.104	0.024	0.037	-0.796	0.181	0.615
Machinery, computers, RTV, communication	8	10	5.821	0.013	0.051	-0.692	0.352	0.619
Vehicles and other transport equipments	9	11	5.999	0.032	0.022	-0.664	0.357	0.644
Electricity, gas and water	10	3	4.471	0.004	0.038	-0.359	0.257	0.599
Construction	11	0	6.563	0.001	0.087	-0.919	0.071	0.711
Trade and repairs	12	1	6.004	0.010	0.163	-0.416	0.151	0.262
Hotels and restaurants	13	5	7.518	0.003	0.031	-0.100	0.062	0.349

Continued

Table 6 The specific sectors targeted by 29 countries for the 1985-2000 period (from Alfaro & Charlton, 2008)—Cont'd

Industries	Code	Targeted by number of countries	Growth in value added	FDI/ value added	Share of value added	Dep. external finance	White high skill	Blue and white high skill
Land, sea, and air transport	14	1	5.733	0.002	0.071	-0.150	0.055	0.140
Telecommunications	15	12	6.762	0.007	0.031	-0.119	0.033	0.190
Monetary intermediation	16	1	5.071	0.030	0.053	-2.445	0.820	0.827
Other financial intermediation	17	1	9.184	0.050	0.006	-3.613	0.714	0.737
Insurance	18	1	5.974	0.050	0.013	-3.586	0.761	0.773
Real estate and business activities	19	11	8.741	0.008	0.210	-0.173	0.564	0.612

Notes: Number of industries corresponds to the 3-digit ISIC Rev. 3. FDI corresponds to Foreign Direct Investment Flows from OECD International Direct Investment Database. Dependence on external finance is the difference between investment and cash generated from operations in the US following Rajan and Zingales (1998). Skill data are the ratio of high-skilled workers to other workers in German Industries. Skilled workers include: White-collar high-skill (WCHS): Legislators, senior officials and managers (Group 1), Professionals (Group 2), Technicians and associate professionals (Group 3). White-collar low-skill (WCLS): Clerks, service workers (Group 4), shop and sales workers (Group 5). Blue-collar high-skill (BCHS): Skilled agricultural and fishery workers (Group 6), Craft and related trade workers (Group 7). Blue-collar low-skill (BCLS): Plant and machine operators and assemblers (Group 8), Elementary occupations (Group 9). See Appendix for detailed explanation of all variables and sources.

5.1 FDI and productivity

To justify special treatment, foreign enterprises would need to confer some type of positive externality which is not internalized by firms. A typical test of whether foreign firms transfer technology to domestic enterprises would be to estimate firm-level productivity, and measure whether a more extensive foreign presence increases domestic firm productivity. As an illustration, Aitken and Harrison (1999) estimate the following production function for firm i in sector j at time t :

$$Y_{it} = \text{Constant} + \beta_1 \text{DFI_Plant}_{ijt} + \beta_2 \text{DFI}_{\text{Sector}_{jt}} \\ + \beta_3 (\text{DFI_Plant}_{ijt} \times \beta_2 \text{DFI_Sector}_{jt}) + \beta_4 Z_{ijt} + \alpha_i + \tau_t + \varepsilon_{it}$$

where

$$\text{DFI}_{\text{Sector}_{jt}} = \frac{\sum_{i \in j} \text{DFI_Plant}_{ijt} \times \text{Employment}_{ijt}}{\sum_{i \in j} \text{Employment}_{ijt}}$$

The dependent variable could be log of output, in which case the Z vector would control for input use, leaving foreign investment to affect the residual determinants of output, which is typically interpreted to be TFP levels. Alternatively, the dependent variable Y could be output per worker or TFP, and the vector Z could include a number of other controls, such as size, openness, or other determinants of productivity. DFI_Plant is typically measured as the share of the plant which is foreign-owned. The coefficient β_1 then measures whether firms with foreign investment are more productive than other plants. Most researchers find that the own-plant effect is large and significant. In other words, firms with foreign equity participation typically have higher output, higher output per worker, or higher levels of TFP.

This is an important point which has not been emphasized enough in the literature on FDI. While many researchers are focused on identifying some sort of externality or technology spillover, probably the most important contribution that foreign firms make is the direct effect captured by the coefficient β_1 on DFI_Plant . Ramondo (2008) develops a model where countries gain access to foreign technologies through the activities of multinationals. In counterfactual exercises with her estimated model she finds that for small countries the associated benefits can be quite large.

Policies that promote joint venture activity are consistent with the empirical results showing productivity gains from FDI. As an illustration, promoting joint ventures has been the core of China's policy to benefit from inward foreign investment. Foreign investors in key sectors (mobile phones, computers) were required to

enter into joint ventures with domestic firms (Rodrik, 2006). In electronics, Huchet (1997) writes that China's technology strategy has been to "allow foreign firms access to the domestic market in exchange for technology transfer through joint production or joint ventures." Wholly owned foreign firms are a rarity in China; most firms are joint ventures between local (frequently state-owned) and foreign enterprises.

China is a particularly interesting case to study in seeking to understand the role played by technology transfer via FDI. Since opening its economy to the outside world in late 1978, China has absorbed an increasing amount of FDI and is now among the world's largest hosts of FDI inflows. A number of recent papers use Chinese firm-level data to evaluate the own-plant effect (the coefficient β_1 that measures whether firms with foreign investment are more productive than other plants). All these studies, including Du, Harrison, and Jefferson (2008), Hu and Jefferson (2002), and Liu, Liu, and Yifan (2008) find that joint ventures in China exhibit not only higher productivity levels than other enterprises but also higher productivity growth. Du et al. find that these effects are largest for state-owned enterprises that form partnerships with foreign firms. The impact on their productivity is 10 times that on privately owned enterprises. Du et al. suggest that forming joint venture partnerships for Chinese state-owned enterprises has provided an alternative approach to privatization as a way of increasing public sector performance. Bartel and Harrison (2005) also find that public sector enterprises in Indonesia improved performance after forming joint ventures with foreign enterprises.

Although there is a strong relationship between foreign ownership and productivity levels, that effect often cannot be separately identified from a firm fixed effect. So if the equation above is estimated either in first differences or with firm-specific effects, then the coefficient on DFI_Plant may not be significant. What this implies is that firms with foreign equity participation are at a higher level of technology, but technological change is not always higher for these enterprises. This should not be surprising, since many of these firms are already at the technology frontier.

Evidence suggests that encouraging less productive state enterprises to make alliances with foreign firms may raise productivity. However, most research has not used any sort of identification strategy to tease out the direction of causality. One possibility is that foreign firms acquire the most productive domestic enterprises, or form alliances with the most productive firms. Arnold and Javorcik (2006) attack this problem. Combining differences-in-differences estimation with nonlinear matching techniques, they show that the acquired firms outperformed the control group in every observable dimension, including exhibiting higher productivity growth, higher investment, and higher sales growth. They conclude that foreign equity infusions do confer benefits to domestic enterprises, and that the effect is not simply one of picking winners. However, more studies in this regard would be useful.

The coefficient β_2 on DFI_Sector measures the extent to which foreign ownership in the sector positively affects the productivity of domestic enterprises. This effect is sometimes referred to as a horizontal spillover. Early studies, such as Blomstrom and Wolff (1994) for Mexico, typically estimated a pure cross section or failed to include firm or industry-specific effects. Consequently, the coefficient β_2 was always positive and frequently significant, indicating the presence of positive horizontal spillovers. Aitken and Harrison (1999) showed that this positive coefficient for Venezuelan manufacturing enterprises was spurious, indicating that foreign firms were attracted to highly productive sectors. When they included firm or sector effects, the coefficient on DFI_Sector switched from positive to negative. Aitken and Harrison interpreted the negative coefficient as indicative of market-stealing: foreign firms grab market share from domestic firms, driving them up their cost curves in sectors with economies of scale.

Other recent studies measuring the extent of horizontal spillovers are listed in Appendix Table 7. In contrast to earlier studies which generally found support for positive (horizontal) spillovers, more recent studies that have attempted to reproduce Aitken and Harrison (1999) for other countries have typically found insignificant or negative horizontal externalities. This includes Djankov and Hoekman (2000) for the Czech Republic, Smarzynska (2002) for Lithuania, Lopez-Cordova (2002) for Mexico, Damijan, Majcen, Knell, and Rojec (2001) for eight transition economies, Kathuria (2000) for India, Hu and Jefferson (2002), and others.³⁰

One explanation for the lack of positive horizontal spillovers is that foreign firms have no incentives to transfer knowledge or technology to competitors within the same industry. They should, however, have an incentive to help the productivity of their suppliers, by transferring knowledge to them (see Kugler, 2001). More recent work has sought to identify what is now referred to as vertical spillovers, which are positive externalities stemming from the relationships of foreign enterprises with domestic suppliers or customers. Forward spillovers could occur if foreign firms that locate domestically supply inputs that embody new technologies or processes. Backward spillovers could occur if domestic suppliers to downstream foreign firms benefit from contacts with the firms to increase productivity. Smarzynska (2004) defines horizontal, forward, and backward FDI spillovers as follows:

$$\text{Horizontal FDI}_{jt} = \frac{\sum_{i \in j} \text{DFI_Plant}_{ijt} \times \text{OUTPUT}_{ijt}}{\sum_{i \in j} \text{OUTPUT}_{ijt}}$$

and

$$\text{Forward FDI}_{jt} = \sum_{m \neq j} \sigma_{jm} \frac{\sum_{i \in m} \text{DFI_Plant}_{ijt} \times (Y_{ijt} - X_{ijt})}{\sum_{i \in m} (Y_{ijt} - X_{ijt})}$$

where σ_{jm} is the share of inputs purchased by industry j from industry m in total inputs sourced by sector j . Finally, if α_{jkt} is the proportion of sector j 's output that is supplied to sector k (taken from the input-output matrix), then backward FDI spillovers are defined as:

$$\text{Backward FDI}_{jt} = \sum_{k \neq j} \alpha_{jk} \times \text{Horizontal FDI}_{kt}$$

Smarzynska finds a zero or negative impact from forward linkages, and a positive impact of backward linkages, indicating that technological gains from FDI are primarily concentrated among domestic suppliers interacting with downstream foreign firms. Recent studies on other countries (Appendix Table 7) also find these positive backward FDI spillovers, including Blalock and Gertler (2003) for Indonesia, Liu (2008) for China, Gorodnichenko, Svejnar, and Terrell (2008) for 15 transition economies, and Lopez-Cordova (2002) for Mexico. The evidence in favor of backward spillovers may be one factor that could explain China's emphasis on the use of domestic content requirements. Sutton (2004) describes the use of domestic content requirements in both India and China as follows:

From the early '90s onwards, a wave of multinational firms entered both markets. In both countries, these entrants were required to achieve a high level of domestic content within a specified period (typically, 70 percent within 3 years). For at least some of the new entrants, this was seen as an unreasonable target, as domestic suppliers could not meet the price and quality requirements of the car makers. Achieving the 70 percent target required the car makers to switch rapidly from a reliance on imported components to sourcing from local vendors; and this in turn gave the car makers a strong incentive to work closely with (first-tier) suppliers, to ensure that quality standards were met, within an acceptable price.

These requirements illustrate how IP has been actively used to shape China and India's auto sector through the use of domestic content requirements. Even if there are backward spillovers, however, these spillovers could be captured by subsidizing domestic input purchases by foreign firms instead of imposing them through domestic content requirements. Domestic content requirements could deter some investments that would have created backward spillovers and may lead to inefficient outcomes by protecting domestic suppliers from import competition.

A related approach is to think of the benefits of FDI through their impact on the variety of nontradable inputs available in the host country. Rodríguez-Clare (1996) presents a model to explore this idea. Under full employment, multinationals expand at the expense of domestic firms, so what matters is the demand for domestic inputs per unit of labor hired, a term that Rodríguez-Clare labels the *linkage coefficient*.³¹ If multinationals' linkage coefficient is higher than that of domestic firms, the multinationals' backward linkage effect is positive and FDI would lead to an increase in input variety and host-country productivity.³²

Using firm-level data for several Latin American countries, Alfaro and Rodríguez-Clare (2004) show that although multinationals source a lower share of their inputs domestically than domestic firms, they nevertheless tend to have a higher linkage coefficient thanks to the fact that they use more roundabout production methods, requiring more inputs per unit of labor. Still, to date there is only anecdotal evidence that multinationals indeed lead to an increase in the variety of inputs available in the host country. Moreover, one concern is that if multinationals generate positive backward linkages then this should benefit domestic firms that use similar inputs as multinationals. One would then expect to see positive horizontal externalities, but this is generally not what the literature finds (at least in LDCs).

To summarize, the evidence seems to consistently indicate that:

1. Firms that receive FDI (joint ventures) or are acquired by multinationals generally exhibit higher productivity levels.
2. There is evidence of positive vertical spillovers from foreign buyers to domestic suppliers (backward linkages) and from foreign suppliers to domestic buyers (forward linkages).
3. There are generally insignificant horizontal spillovers to firms within the same industry.

Anecdotal evidence and new research (Alfaro & Charlton, 2008; Chandra & Kolavalli, 2006) suggests that FDI has been particularly important in cases where governments were actively engaged in strategies of technological upgrading in certain sectors, and brought in foreign companies as part of those strategies. Typically, these efforts were part of a set of complementary policies that included increasing the supply of skilled workers in a targeted industry, improving regulation and infrastructure, promoting new activities and innovation, and increasing exports. In many of those cases, the government probably encouraged joint ventures, but the direction of causality is not so clear.

Given these results, do the benefits of fiscal incentives for foreign enterprises outweigh the costs? If the primary reason for giving these incentives is to encourage technology transfer, then the answer should probably be no: if foreign firms are the only ones that use the inputs that benefit from backward spillovers and there are no

horizontal spillovers, then FDI will occur without subsidies. Yet there is clearly a further need to understand the mechanism through which foreign firms generate vertical spillovers. Even if vertical spillovers do exist, Pack and Saggi (2006) argue that “the magnitude of some of the incentives being used seems difficult to justify” (p. 281). They also point out that “investment incentives and tax breaks to multinational investors work against their local competitors. Thus, if there are local firms that could potentially compete with multinationals, the adverse effect on such firms of tax incentives to multinationals needs to be taken into account. The efficacy of investment incentives is also unclear—such policies could easily end up transferring rents to foreign investors without affecting their investment decisions” (Pack & Saggi, p. 281).

5.2 FDI and host-country factor markets

One robust result in the literature on foreign ownership and productivity is that firms with foreign equity have higher productivity levels. Foreign firms are also typically more capital intensive and spend more on worker training. As long as firms do not face a perfectly flat labor supply schedule, it is likely that these different firm characteristics will translate into higher wages. There could also be other imperfections in the labor market, such as search costs for firms seeking skilled workers, or efficiency wage setting, that could result in foreign firms paying higher wages. To the extent that there is worker mobility or productivity spillovers, these wage effects for workers at foreign enterprises could also spill over to other workers.

Almost all studies find that workers in foreign firms are paid higher wages, which may reflect the fact that foreign firms or joint ventures have higher productivity levels. In perfect markets more productive firms would not pay higher wages, but market imperfections (such as search costs or informational asymmetries) could generate such a link. Evidence showing higher wages paid by multinationals includes Aitken, Harrison, and Lipsey (1996) for Mexico and Venezuela, Lipsey and Sjoholm (2004) for Indonesia using a pure cross section, Harrison and Scorse (2004) for Indonesia using a panel, Velde and Morrisey (2003) for a set of African economies, Martins and Esteves (2006) and Poole (2008) for Brazil, and Earle and Telegdy (2007) for Hungary. Studies of industrialized countries also find large wage gaps between domestic and foreign enterprise wages when researchers do not condition for worker or firm characteristics. This includes Almeida (2004) for Portugal, Girma, Greenaway, and Wakelin (2001) and Drifffield and Girma (2003) for the United Kingdom, and Feliciano and Lipsey (1999) for the United States. The unconditional wage gap, which is the gap in wages paid by foreign versus domestically owned enterprises without controlling for worker or firm characteristics, is typically large. It is frequently as high as 40% (for Hungary) or 50% (for Brazil).

When researchers control for worker and firm characteristics, then the wage premium paid by foreign firms declines dramatically. Earle and Telegdy (2007) find that the wage gap between workers employed at foreign and domestic enterprises is not much affected by conditioning on worker characteristics, but that controlling for industry reduces the premia to 34%, and controlling for firm size further reduces it to 28%. (Robert Lipsey has questioned whether it makes sense to control for characteristics such as firm size: if foreign firms are larger and consequently pay higher wages than domestically owned enterprises, workers are nevertheless better off.) When Earle and Telegdy control for unobserved fixed effects by exploiting changes in firm ownership to identify its effect on wages, the premium is further reduced to 7%. Harrison and Scorse (2004) also find that the premium falls to between 5% and 10% when worker and firm characteristics are controlled for. Ibarraran, using data for Mexican maquilas in 1992 and 1999, finds no premium in 1992 but a small premium in 1999.

Martins and Esteves (2006) and Poole (2008) use matched worker and firm panel data for 1995–1999 to analyze the impact of foreign ownership on wages in Brazil. Like Earle and Telegdy (2007), Martins and Esteves (2006) use changes in firm ownership as a way to control for unobserved firm-specific effects that could be correlated with wage premia. They also follow workers who move to or leave foreign enterprises, to control for unobserved worker-specific effects. They find that workers moving from foreign to domestic firms typically take wage cuts when they move, while movers from domestic to foreign firms increase their pay. However, compared to the unconditional wage gaps of 50%, the wage premium associated with working for a foreign firm falls to between 3% and 7% once worker and firm characteristics are controlled for. Both Martins and Esteves (2006) and Poole (2008) conclude that their results support a positive impact of foreign firms in the Brazilian labor market.

To summarize, the evidence suggests that foreign firms pay a small wage premium. While the earlier literature found larger wage premia of more like 20%, these earlier estimates failed to adequately control for individual characteristics of workers, such as education and experience. Consequently, part of the wage gap stems from the fact that foreign firms tend to hire better educated and more skilled workers. Nevertheless, we can safely conclude that foreign enterprises do not pay workers below what their domestic counterparts would pay. In fact, most of the evidence suggests that foreign firms tend to pay higher wages than comparable domestic firms. There is also evidence that foreign firms are more susceptible to pressure from labor groups, leading them to exhibit greater compliance with minimum wages and labor standards. Harrison and Scorse (2008) find that foreign firms in Indonesia were much more likely than domestic enterprises to raise wages and adhere to minimum wages as a consequence of the

antisweatshop campaigns there. They also find that the employment costs of the anti-sweatshop campaigns were minimal, as garment and footwear subcontractors were able to reduce profits to pay the additional wage costs without reducing the number of workers.

While most of the emphasis on the role of FDI in local factor markets is on the relationship between FDI and host-country wages, there is also an emerging literature on FDI's effect on local capital markets. One reason policy makers give for promoting foreign investment in developing countries is the scarcity of capital for new investment. This argument is based on the assumption that incoming foreign investors provide additional capital when they set up new enterprises in local markets. However, as exchange rate volatility has continued to rise many foreign investors have found ways to hedge by borrowing on local capital markets. This increase in local financing for incoming foreign investors may lead to crowding out of domestic firms.

Harrison and McMillan (2003) and Harrison, Love, and McMillan (2004) test whether or not borrowing by local multinationals is crowding out domestic enterprises. The framework in both papers uses an Euler equation approach combined with Generalized Method of Moment estimation. While Harrison and McMillan's work is a country case study which analyzes the behavior of mostly French multinationals operating in Cote d'Ivoire, Harrison et al. use company level data across a panel of countries. The results suggest that in a country such as the Cote d'Ivoire, which was riddled with market imperfections and where access to credit was rationed due to interest rate ceilings, foreign investors did indeed crowd domestic enterprises out of local credit markets. However, Harrison et al. found that foreign investors tended to "crowd in" domestic enterprises—that is, as foreign investment increased, the amount of credit available to domestically owned firms actually rose. The contrasting results again point to the important role of policy complementarities: in a country with credit market imperfections such as the Cote d'Ivoire, FDI exacerbated these problems. These studies using microdata are consistent with macrolevel evidence on the importance of complementarities with financial market development for ensuring productivity gains from FDI (Alfaro et al., 2004).

6. CONCLUDING COMMENTS

Given the varied experiences across countries and time periods, the different interpretations possible, and the difficulties in conducting clean empirical analyses, it is not easy to arrive at strong conclusions regarding the role of industrial, trade, and FDI policy in development. Our survey of the theory and the evidence nevertheless suggests some tentative conclusions.

We do not endorse infant-industry protection in this chapter; yet we do not claim that a somewhat uniform and moderate tariff is a disaster for development. There are

instances where infant-industry protection was successful—particularly in the late nineteenth and early twentieth centuries—and could work today in developing countries. Still, the conditions needed for infant-industry protection to succeed are generally not satisfied. The framework in Section 2 shows that the theoretical justification for intervention requires at a minimum either industry-level rents or a latent comparative advantage, as well as large Marshallian externalities from production. These necessary conditions are not easy to identify for policy makers *ex ante*. Nor is there evidence suggesting that developing countries have generally protected sectors with latent comparative advantage and Marshallian externalities. Most developing countries have the highest tariffs in consumer goods sectors, such as textiles and apparel. We suspect that the types of sectors where there are important Marshallian externalities are not those where developing countries have a latent comparative advantage (i.e., sectors that are intensive in knowledge and human capital). It is also likely that protection has been used as a tool to protect sunset industries instead of sunrise industries; hence, there is an inherent bias against promoting sectors with a latent comparative advantage.

For infant-industry protection to improve welfare it must pass both the Mill and Bastable tests. The Mill test requires that the protected sector can eventually survive international competition without protection, whereas the Bastable test requires that the discounted future benefits compensate the present costs of protection. This means that the dynamic forces which increase industry productivity must operate quickly. In practice, most research assessing the success of IP has ignored these tests. Even if studies could show that protected sectors grew faster, this would not be sufficient evidence to claim that IP is justified from a welfare standpoint. There are other problems as well. In all the models presented in this review, protection is never the first-best policy. In addition, the infant-industry framework typically assumes that the mere expansion of a sector will generate all sorts of positive effects that will increase industry-wide productivity. But this may not happen, and the economy may simply end up with a larger version of the inefficient sector it began with. We argue that it may be better to implement policies designed directly to elicit the investments that will increase productivity. Such investments may not occur without public intervention because of coordination failures.

The hundreds of studies on trade policies, trade shares, productivity, and growth show a strong correlation between increasing trade shares and country performance, and no significant correlation between tariffs on final goods and country outcomes. The only exception is for intermediate or capital goods, where a higher tariff is associated with lower growth. Putting aside the serious problems of reverse causality identified in our review, we interpret this evidence as suggesting that trade and FDI policies are most successful when they are associated with increasing exposure to trade. One implication is that interventions that increase exposure to trade (such as export

promotion) are likely to be more successful than other types of interventions (such as tariffs or domestic content requirements).

We remain skeptical that protection or subsidies to FDI are needed. Nevertheless, new evidence suggests that IP through FDI promotion may be more successful than intervention in trade, in part because FDI promotion policies focus on new activities rather than on protecting (possibly unsuccessful) incumbents. We are confident that if such measures are part of a broader effort to achieve technological upgrading then they may be helpful, whereas if they are implemented in isolation they are likely to fail. This is consistent with Chandra and Kolavalli (2006), who conclude their overview of 10 cases of successful technology upgrading by stating that “the role of government was most effective when its support for specific industries was embedded in institutions and policies that were internally consistent, had an explicit purpose, and were blessed with political commitment” (p. 39). Rodrik refers to this characteristic as “embeddedness.” Chandra and Kolavalli conclude that “in every case, getting it right depended upon the degree of synchronization between institutions and government policies to motivate learning among exporters. The hallmark of these policies was industry specificity, which offers some useful lessons for other developing countries” (p. 41).

The long-running discussion about “picking winners” can be sidestepped by focusing on efforts of “discovery” (as argued by Hausmann & Rodrik, 2003), or by simply working with existing industries and clusters to deal directly with the coordination failures that limit their productivity and expansion. For example, instead of blanket subsidies for exports and FDI, we think of attracting multinationals to produce key inputs or to bring specific knowledge needed by clusters with the ability to absorb them. As Chandra and Kolavalli (2006) have put it, “without host-country policies to develop local capabilities, MNC-led exports are likely to remain technologically stagnant, leaving developing countries unable to progress beyond the assembly of imported components” (p. 19).

We envision an important role for what we refer to as “soft” industrial policy, whose goal is to develop a process whereby government, industry and cluster-level private organizations can collaborate on interventions that can directly increase productivity. The idea is to shift the attention from interventions that distort prices to interventions that deal directly with the coordination problems that keep productivity low in existing or raising sectors. Thus, instead of tariffs, export subsidies, and tax breaks for foreign corporations, we think of programs and grants to, for example, help particular clusters by increasing the supply of skilled workers, encouraging technology adoption, and improving regulation and infrastructure. While “hard” IP is easier to implement than “soft” IP measures, tariffs and subsidies become entrenched and are more easily subject to manipulation by interest groups.

The specific policies that should be pursued as part of this type of IP depend, of course, on the particular coordination failures that affect a cluster. Given the variety of coordination failures that exist, there is a need for a wide set of instruments or policies. An exhaustive list is therefore impossible. Some examples are: regulation to enforce higher quality standards in cases of imperfect information or externalities; public investment in specific infrastructure projects when there are strong investment complementarities (e.g., a regional airport geared to exploit tourism opportunities, or an irrigation project for modern agriculture); attraction of FDI to bring in foreign technologies; scholarships for studies abroad in areas deemed important for growth and diversification of a cluster in cases where thin markets prevent individuals from making such investments; grants for innovative projects proposed by single firms or entrepreneurs, prizes to innovative firms, grants for research projects proposed by organized producers and performed by local research centers, and technical assistance to allow long-term collaborative strategies for education and research between business associations and universities.

It is clearly unreasonable to expect governments to be able to identify the coordination failures affecting different sectors or clusters. A more realistic approach is to invite sector and cluster organizations to come forward with well-justified proposals for government support. It is instructive here to reproduce the practical advice of Altenburg and Meyer-Stamer (1999): “To meet the demands of globalized competition, intra-firm efforts are not sufficient. The business sector has to be able to organize collective action for self-help, and it must be able to articulate its demands *vis-à-vis* political actors. This places great demands on business associations, both in terms of service provision and lobbying. It implies a fundamental upgrading process and the creation of a learning organization. Key features are a professionalization of business associations (e.g., employing more and better qualified professionals) and the implementation of mechanisms to ensure ongoing organizational development.” Perhaps the government should provide support to different sectors that want to start or improve their level of organization. This would be the first line of action in countries where the private sector organizations are weak or are designed for rent seeking or confrontation rather than constructive work.³³

In comparison with the more traditional approach to IP, the soft IP that we propose here has two additional advantages. First, although this requires more research, we conjecture that a soft IP reduces the scope for corruption and rent-seeking associated with hard IP such as protection or selective production subsidies. Second, soft IP is much more compatible with the multilateral and bilateral trade and investment agreements that many LDCs have implemented over the last decades. It is true that if an LDC wants to protect an industry for a period of time, it can always negotiate “space” for that policy when it joins the WTO. This is warranted under the WTO’s rules for

special and differential treatment (SDT), which call for “preferential market access for developing countries, limits reciprocity in negotiating rounds to levels ‘consistent with development needs’ and provides developing countries with greater freedom to use trade policies than would otherwise be permitted by GATT rules” (Hoekman, 2005, p. 1). But if the country has already joined the WTO then this is not possible. Moreover, export subsidies are supposed to be eliminated by all but the poorest countries by 2015 (agreement on subsidies and countervailing measures), local content requirements on multinationals are now WTO-illegal (agreement on trade-related investment measures, TRIMS), and patent laws are supposed to be set according to international standards (agreement on trade-related intellectual property, TRIPS). Such restrictions make it impossible for LDCs to follow some of the policies implemented by South Korea and Taiwan, for example (Rodrik, 2004). Of course, as emphasized by Rodrik, some policies associated with hard IP remain feasible: countries can provide fiscal incentives to particular sectors or to new activities. But clearly the policy space for hard IP has shrunk over the last decades, while that for soft IP remains basically unrestricted.

One area where additional research is urgently needed is on the human cost of adjustment to trade and FDI reforms. Although this was not the focus of our chapter, we note that there is limited research addressing how poor and unskilled workers are affected by trade reform and incoming FDI flows. Some preliminary evidence suggests that trade reforms may push workers toward the informal sectors (Goldberg & Pavcnik, 2007; Muendler, 2007). To the extent that countries lack social safety nets, they may choose to delay liberalizing their trade or FDI policies. Harrison (2007) suggests that globalization is not inherently pro-poor, and that complementary measures are needed to cushion the impact and ensure that the gains from globalization are realized for everyone.

APPENDIX

We first show how size v_m matters for wages. We assume that $x_1 = x_2 = x$. Also, to simplify, we assume that $T_j = L_j = 1$ all j . In this case

$$\Phi_m \equiv \sum_j \hat{x}_{jm}^\theta w_j^{-\theta} \quad (\text{A.1})$$

We also have $\hat{x}_{11} = \hat{x}_{22} = x$, while $v_1 > v_2$ and $\hat{x}_{j1} = \hat{x}_{j2} = z_j$ for any $j = 3, \dots, N$. Then it is clear that $\Phi_1 > \Phi_2 \Leftrightarrow w_1 < w_2$. Assuming to simplify notation that $Y = 1$ then

$$w_j^{1+\theta} = \sum_m \hat{x}_{jm}^\theta v_m / \Phi_m \quad (\text{A.2})$$

Thus,

$$w_1^{1+\theta} = \sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} x^\theta + \frac{v_2}{\Phi_2}$$

and

$$w_2^{1+\theta} = \sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} + \frac{v_2}{\Phi_2} x^\theta$$

In equilibrium Φ_1 and Φ_2 solve the following system of equations

$$\Phi_1 = x^\theta \left(\sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} x^\theta + \frac{v_2}{\Phi_2} \right)^{-\theta/1+\theta} + \left(\sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} + \frac{v_2}{\Phi_2} x^\theta \right)^{-\theta/1+\theta} + A$$

$$\Phi_2 = \left(\sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} x^\theta + \frac{v_2}{\Phi_2} \right)^{-\theta/1+\theta} + x^\theta \left(\sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} + \frac{v_2}{\Phi_2} x^\theta \right)^{-\theta/1+\theta} + A$$

where $A = \sum_{j=3} z_j^\theta w_j^{-\theta}$. If $v_1 = v_2$ then the solution of the system entails $\Phi_1 = \Phi_2$. Also, if $\Phi_1 = \Phi_2$ then necessarily $v_1 = v_2$. Now we show that when $v_1 > v_2$ the solution of the system is such that $\Phi_1 < \Phi_2$. We consider $(\Phi_1 - \Phi_2)'_{v_1}$ at the point $v_1 = v_2$ and $\Phi_1 = \Phi_2$. At the point $v_1 = v_2$ and $\Phi_1 = \Phi_2$ we have

$$(\Phi_1 - \Phi_2)'_{v_1} = \frac{-\theta}{1+\theta} \left(\sum_{m=3} \frac{v_m}{\Phi_m} + \frac{v_1}{\Phi_1} x^\theta + \frac{v_1}{\Phi_1} \right)^{-\theta/(1+\theta)-1} \frac{1}{\Phi_1} (x^{2\theta} + 1 - 2x^\theta) < 0$$

This establishes that a small increase in v_1 implies that $\Phi_1 < \Phi_2$. Now suppose there exist $v_1 > v_2$ such that $\Phi_1 > \Phi_2$. Due to continuity there would exist $v_1 > v_2$ such that $\Phi_1 = \Phi_2$. But this contradicts $v_1 = v_2 \Leftrightarrow \Phi_1 - \Phi_2 = 0$. Thus, the solution of the system above implies that if $v_1 > v_2$, then $\Phi_1 < \Phi_2$. Thus, we have that $\Phi_1 < \Phi_2$ in equilibrium, and this implies $w_1 > w_2$.

Now consider two countries such that country 1 has HP in the first industry and country 2 has HP in the second industry. We assume that $v_1 = v_2 = v$ and (as above) that $\hat{x}_{j1} = \hat{x}_{j2} = z_j$ for any $j = 3, \dots, N$. We will show that $x_1 > x_2$ implies $w_1 > w_2$. From 12 we see that $w_1 > w_2$ is equivalent to $(x_1^\theta - 1)\Phi_2 > (x_2^\theta - 1)\Phi_1$. Obviously, $x_1 = x_2 \Leftrightarrow \Phi_1 = \Phi_2$. We can write $\Phi_1 = f(x_1, x_2)$ and $\Phi_2 = g(x_1, x_2)$. Moreover, $f(x_1, x_2) = g(x_2, x_1)$.

This implies that $(x_1^\theta - 1)\Phi_2 - (x_2^\theta - 1)\Phi_1 = (x_1^\theta - 1)g(x_1, x_2) - (x_2^\theta - 1)f(x_1, x_2) = 0$ if and only if $x_1 = x_2$. We have that $(x_1^\theta - 1)\Phi_2 - (x_2^\theta - 1)\Phi_1$ is equal to

$$\sum_{j=3}^N \hat{z}_j^\theta w_j^{-\theta/1+\theta} (x_1^\theta - x_2^\theta) + (x_1^\theta - 1) \sum_{j=1,2} \hat{x}_{j2}^\theta w_j^{-\theta/1+\theta} - (x_2^\theta - 1) \sum_{j=1,2} \hat{x}_{j1}^\theta w_j^{-\theta/1+\theta}$$

Notice that $\sum_{j=3}^N \hat{z}_j^\theta w_j^{-\theta/1+\theta} (x_1^\theta - x_2^\theta) > 0$, as $x_1 > x_2$. Consider then

$$A \equiv (x_1^\theta - 1) \sum_{j=1,2} \hat{x}_{j2}^\theta w_j^{-\theta/1+\theta} - (x_2^\theta - 1) \sum_{j=1,2} \hat{x}_{j1}^\theta w_j^{-\theta/1+\theta}$$

Using $\hat{x}_{11}^\theta = x_1$, $\hat{x}_{21}^\theta = 1$, $\hat{x}_{12}^\theta = 1$, $\hat{x}_{22}^\theta = x_2$, and 12, and simplifying, we get

$$A = \left(\sum_{m=3,\dots,N} \frac{v_m}{\Phi_m} + \frac{vx_1^\theta}{\Phi_1} + \frac{v}{\Phi_2} \right)^{-\theta/1+\theta} (2x_1^\theta - 1 - x_1^\theta x_2^\theta) \\ - \left(\sum_{m=3,\dots,N} \frac{v_m}{\Phi_m} + \frac{vx_2^\theta}{\Phi_2} + \frac{v}{\Phi_1} \right)^{-\theta/1+\theta} (-x_2^\theta x_1^\theta - 1 + 2x_2^\theta)$$

Differentiating A w.r.t. x_1 and evaluating at $x_1 = x_2 = x$ and $\Phi_1 = \Phi_2 = \Phi$ yields

$$A'_{x_1} = \left(\sum_{m=3,\dots,N} \frac{v_m}{\Phi_m} + \frac{vx_1^\theta}{\Phi_1} + \frac{v}{\Phi_2} \right)^{-\theta/1+\theta} \theta x_1^{\theta-1} \left(2 - x_1^\theta + x_1^\theta - \frac{\frac{\theta}{1+\theta} (2x_1^\theta - 1 - x_1^{2\theta}) \frac{v\theta x_1^{\theta-1}}{\Phi_1}}{\sum_{m=3,\dots,N} \frac{v_m}{\Phi_m} + \frac{vx_1^\theta}{\Phi_1} + \frac{v}{\Phi_2}} \right) \\ = \left(\sum_{m=3,\dots,N} \frac{v_m}{\Phi_m} + \frac{vx_1^\theta}{\Phi_1} + \frac{v}{\Phi_2} \right)^{-\theta/1+\theta} \theta x_1^{\theta-1} \left(2 + \frac{\frac{\theta}{1+\theta} (2x_1^\theta + 1 + x_1^{2\theta}) \frac{v\theta x_1^{\theta-1}}{\Phi_1}}{\sum_{m=3,\dots,N} \frac{v_m}{\Phi_m} + \frac{vx_1^\theta}{\Phi_1} + \frac{v}{\Phi_2}} \right) > 0$$

This implies that if $x_1 > x_2$, then $A > 0$ and, thereby,

$$\frac{x_1^\theta - 1}{\Phi_1} > \frac{x_2^\theta - 1}{\Phi_2}$$

This implies that $w_1 > w_2$.

Now consider two industries with the same size. The only difference is the number of countries that have HP of size x in these two industries. Assume that a set J_i of countries have HP in industry $i = 1, 2$, with $J_2 \subset J_1$. We have

$$\begin{aligned}\Phi_1 &= x^\theta w_1^{-\theta} + w_2^{-\theta} + x^\theta \sum_{j \in J_2} w_j^{-\theta} + x^\theta \sum_{j \in J_1/J_2} w_j^{-\theta} + \sum_{j \notin J_1} w_j^{-\theta} \\ &= x^\theta \left(\frac{x^\theta v}{\Phi_1} + \frac{v}{\Phi_2} + \sum_{m \geq 3} \frac{v_m}{\Phi_m} \right)^{-\theta/(1+\theta)} + \left(\frac{v}{\Phi_1} + \frac{x^\theta v}{\Phi_2} + \sum_{m \geq 3} \frac{v_m}{\Phi_m} \right)^{-\theta/(1+\theta)} \\ &\quad + x^\theta \sum_{j \in J_2} w_j^{-\theta} + x^\theta \sum_{j \in J_1/J_2} w_j^{-\theta} + \sum_{j \notin J_1} w_j^{-\theta}\end{aligned}$$

and

$$\begin{aligned}\Phi_2 &= w_1^{-\theta} + x^\theta w_2^{-\theta} + x^\theta \sum_{j \in J_2} w_j^{-\theta} + \sum_{j \in J_1/J_2} w_j^{-\theta} + \sum_{j \notin J_1} w_j^{-\theta} \\ &= \left(\frac{x^\theta v}{\Phi_1} + \frac{v}{\Phi_2} + \sum_{m \geq 3} \frac{v_m}{\Phi_m} \right)^{-\theta/(1+\theta)} + x^\theta \left(\frac{v}{\Phi_1} + \frac{x^\theta v}{\Phi_2} + \sum_{m \geq 3} \frac{v_m}{\Phi_m} \right)^{-\theta/(1+\theta)} \\ &\quad + x^\theta \sum_{j \in J_2} w_j^{-\theta} + \sum_{j \in J_1/J_2} w_j^{-\theta} + \sum_{j \notin J_1} w_j^{-\theta}\end{aligned}$$

If $J_2/J_1 = \emptyset$ (which implies that $J_1 = J_2$) then $\Phi_1 = \Phi_2$. As J_2/J_1 becomes nonempty, then $\Phi_1 > \Phi_2$. But recall that since $x_1 = x_2 = x$ and $v_1 = v_2 = v$ then

$$w_1 = \left(\sum_m \frac{x^\theta v}{\Phi_1} + \frac{v}{\Phi_2} + \sum_{m \geq 3} \frac{v_m}{\Phi_m} \right)^{1/1+\theta}$$

and

$$w_2 = \left(\sum_m \frac{v}{\Phi_1} + \frac{x^\theta v}{\Phi_2} + \sum_{m \geq 3} \frac{v_m}{\Phi_m} \right)^{1/1+\theta}$$

This implies that $w_2 > w_1$.

Table 1 Cross-country studies on openness and growth

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
1. Tyler (1981)	(1) Annual average real growth rate of total exports; (2) Manufacturing sector export earnings	Cross-sectional data for 55 middle-income developing countries from 1960 to 1977	No identification strategy	Significant positive impact of exports on economic growth
2. Feder (1982)	(1) Export growth \times export/GDP ratio; (2) Growth rate of exports (in specific inter-sectoral externalities regression)	Pooled data from 31 semiindustrialized countries and 17 developed countries between 1964 and 1973	No identification strategy	Significant positive impacts of exports on economic growth (export-oriented policies) bring the economy closer to an optimal allocation of resources
3. Salvatore (1983)	Growth of percentage of exports in GDP	Pooling of cross section and time-series data for 52 developing countries from 1961 to 1978	Full information maximum likelihood (FIML); uses consumer price index (CPI) and index of real GDP of all market economies as instruments for exports	No significant positive impact of exports on economic growth. Trade is very important to the development process, but more in the nature of a handmaiden than as an engine of growth. The policy implication is that excessively pursuing

				a policy of industrialization through import substitution can retard growth
4. Balassa (1985)	(1) Export growth × share of exports in GDP; (2) Trade orientation (the deviation of actual from hypothetical value of per capita exports)	Pooled data for 43 semiindustrialized countries from 1973 to 1979	No identification strategy	The results show that the growth rate of GNP is higher with a greater extent of outward orientation at the beginning of the study period and the greater the extent of reliance on export promotion in response to the external shocks of the period
5. Gupta (1985)	Growth rate of exports	Time-series data for (1) South Korea from the first quarter of 1960 to the last quarter of 1979 and (2) Israel from 1967 to 1981	Sims causality test	Bidirectional causality between exports and economic growth for both countries
6. Jung and Marshall (1985)	Growth rate of real value of exports	Time series, annual data for 37 developing countries from 1951 to 1981	Granger causality test	Export-led growth in only five countries while no causality in 20 countries

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
7. Kavoussi (1985)	(1) RX: growth of a specific country's total export earnings; (2) RW: growth of a country's export earnings due to expansion of world demand for its traditional commodities; (3) RC: a rise in the share of world markets for a country's traditional commodities; (4) RD: an increase in the share of nontraditional commodities in a country's total exports; (5) RCD: an indicator of the effect of trade policy (where $RCD = (1 + RC) \times (1 + RD) - 1$; values greater than or equal to 0 imply an outward trade orientation, while values less than 0 imply an inward trade orientation)	Cross-sectional data for 52 low- and middle-income developing countries from 1967 to 1977	No identification strategy	Firstly, the study shows that a rapid expansion of export earnings requires both favorable external markets and outward-oriented commercial policies. Secondly, the results suggest that free trade appears to enhance growth only when external demand is favorable

8. Chow (1987)	Export growth in manufactured goods	Time series, annual data for eight newly industrializing countries (NICs) from the 1960s and 1970s	Sims causality test	Export-led growth in only Mexico, strong bidirectional causality between export growth and industrial development in other seven countries
9. Darrat (1987)	Growth rate of real value of exports	Time series, annual data for Hong Kong, Korea, Taiwan, and Singapore from 1955 to 1982	White's causality test	Exports and economic growth are significantly and positively correlated. Causality runs from exports to economic growth only in Korea. For the remaining three countries, the export-led growth hypothesis is rejected
10. Goncalves and Richtering (1987)	(1) Annual average growth rate of total export volume in 1975 prices; (2) Average share of exports in GDP; (3) Change in share of exports in GDP in each period	Pooled data for 70 developing countries from 1960 to 1981	No identification strategy	No significant positive impact of exports on GDP net of exports

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
11. Hsiao (1987)	Exports	Time series, annual data for Hong Kong, Korea, Taiwan, and Singapore during the 1960s, 1970s, and 1980s	Granger and Sims causality tests	The Sims' test indicates a feedback relationship while the Granger's test indicates no causal relation between the exports and GDP, except for Hong Kong where both tests indicate a unidirectional causality from GDP to exports
12. Ram (1987)	(1) Export growth rates; (2) Export growth rates \times export GDP shares	Time-series and cross-sectional data for 88 countries from 1960 to 1972 and 1973 to 1982	No identification strategy	Significant positive impact of exports on economic growth
13. Rana (1988)	(1) Export growth rate; (2) Export growth rate \times share of exports in GDP	Pooled data for 43 developing countries	No identification strategy	Supports the view that exports would have less effect on growth when the world environment is unfavorable

14. Singer and Gray (1988)	(1) Growth of country's total export earnings; (2) Trade policy factor $RCD = (1 + RC) \times (1 + RD)$, where RC is the competitiveness factor and RD is the diversification factor. Positive values of RCD represent outward trade orientation	Pooled data for 52 developing countries from 1967 to 1973 and 1977 to 1983	No identification strategy	Positive impact of exports on economic growth under favorable world market conditions; however, this positive impact is weaker for low-income countries
15. Kohli and Singh (1989)	(1) Growth of exports; (2) Export share in GDP; (3) Growth of exports × share of exports in GDP	Pooled data for 30 countries from 1960 to 1970 and 1970 to 1981	No identification strategy	Significant positive impact of exports on economic growth in export-oriented nations. The impact is subject to diminishing returns
16. Kunst and Marin (1989)	Manufacturing sector exports	Time series, annual data for Australia	Granger causality test	Positive causality from productivity to exports
17. Mbaku (1989)	Growth rate of exports	Pooled data for 37 African countries from 1970 to 1981	No identification strategy	The results show that export expansion significantly enhances economic growth in Africa

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
18. Fosu (1990)	Annual export growth	Cross sectional data for 28 least-developed countries (LDCs) in Africa from 1960 to 1970 and 1970 to 1980	No identification strategy	Significant positive impact of exports on economic growth, but this positive relationship is smaller compared to other LDCs
19. Grabowski, Sharma, and Dhakal (1990)	Growth of exports	Time series, annual data for Japan from 1885 to 1939 and 1952 to 1980	Granger causality test	No export-led growth in pre-World War II period but export-led growth in post-World War II
20. Sheehey (1990)	Export growth × share of exports in GDP	Pooled data for 36 countries from 1960 to 1970	No identification strategy	This is a critique of Balassa and Feder's work particularly. The author argues that, since the link between sectoral growth and growth of GDP is common to all sectors, the growth clearly cannot be due to relative productivity differences and externality effects. By showing the same tests that support the

				“promotion” of all major components of GDP, the author argues that the previous tests have no bearing at all on the export-promotion/import-substitution controversy
21. Afxentiou and Serletis (1991)	Growth rate of real value of exports	Time series, annual data for 16 developed countries	Granger causality test	Growth led exports in three nations and was bidirectional in the United States. In general, there is no significant causality between export growth and GNP growth in these 16 countries
22. Ahmad and Kwan (1991)	(1) Total real value of exports; (2) Total real value of manufactured exports; (3) Share of manufactured exports in total exports	Pooled time-series and cross-sectional data for 47 developing African countries from 1981 to 1987	Granger causality test	No support for export-led growth or growth led exports
23. Alam (1991)	(1) Export growth; (2) Qualitative evaluation of trade orientation (World Bank Development	Pooled data for 41 developing countries from 1965 to 1973 and 1973 to 1984	No identification strategy	Significant positive relationship between outward orientation and exports and output

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
	Report 1987), including effective rates of protection, export incentives, exchange rate alignments and direct controls			growth rates for both time periods. Moreover, the impact of trade policies on growth rates acted more strongly through increases in productivity rather than increases in investment rates
24. Dodaro (1991)	(1) Manufactured exports as a percentage of total merchandise exports; (2) Dummy variable equal to one if over 50% of a country's exports are makeup of fuels, minerals, and metals, 0 otherwise	Pooled data for 41 developing countries	Instrumental variable technique; uses nominal per capita GNP and population size as instruments for manufactured exports	The level of development is an important determinant of the degree of manufacturing and processing in a country's export basket. Moreover, the results show that the composition of exports affects economic growth. Concentration of exports in manufactures and greater levels of processing allow for greater diversification and

				a reduction in concentration both with respect to commodities and market areas. This reduction in concentration may reduce export earnings instability, fostering further economic growth
25. Esfahani (1991)	(1) Share of imports in $GDP \times$ growth rate of imports; (2) Share of exports in $GDP \times$ growth rate of exports	Cross-sectional data for 31 semiindustrialized countries from 1960 to 1973, 1973 to 1981, and 1980 to 1986	Simultaneous equations; uses log GDP per capital and log labor as instruments for export and import variables	Export promotion is particularly important for countries that cannot obtain sufficient foreign aid or capital
26. Kwan and Cotsomotis (1990)	Ratio of exports to national income (the growth rate of exports to national income is used as a robustness check)	Time series, annual data for China from 1952 to 1985	Granger causality test	Bidirectional causality between exports and economic growth for the postreform period (since 1978), but no causality for the prereform period
27. Oskooee, Mohtadi, and Shabsigh (1991)	Export growth	Time series, annual data for 20 LDCs from the 1950s to the 1980s	Granger causality test with Akaike's optimal lag criterion	The study indicates positive causality from economic growth to export growth in 4 out of 20 countries (Korea, Nigeria, South Africa,

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
				and Thailand) and negative causality in Indonesia. Korea and Thailand exhibit positive causality in both directions
28. Salvatore and Hatcher (1991)	Growth rate of real value of exports	Time series and pooled, annual data for 26 developing countries from 1963 to 1973 and 1973 to 1985	No identification strategy	The econometric results only partially support the hypothesis that international trade benefits most developing countries and that an outward orientation leads to a more efficient use of resources and growth
29. Sharma, Norris, and Cheung (1991)	Export growth	Time series, quarterly data for five industrialized countries from 1960 to 1987	Granger causality test	Export-led growth for Germany and Japan, while growth led exports for the United States and the United Kingdom. There was no causality in either direction for Italy

30. Dollar (1992)	(1) Dollar measure of outward orientation: $RPL_i = 100 \times eP_i/P_{u.s.}$, where RPL_i is the index of country i 's relative price level, e is the exchange rate, and P_i is the consumption price index for country i ; (2) Variability of the real exchange rate	Cross-sectional data for 95 countries from 1976 to 1985	No identification strategy	Significant negative relationship between distortion in the real exchange rate and growth of per capita GDP after controlling for the effects of real exchange rate variability and investment level. Trade liberalization, devaluation of the real exchange rate and maintenance of a stable real exchange rate could improve growth performance in poor countries
31. Edwards (1992)	(1) Index of trade intervention; (2) INVERV1, overall intervention index obtained when a homoskedastic (unscaled) model is used to predict trade flows for the 183 commodities in the sample; (3) INVERV2, overall intervention index obtained when a heteroskedastic (scaled) model with residuals proportional to GNP is	Cross-sectional data for 30 developing countries from 1970 to 1982	Instrumental variable technique; uses the average and coefficient of variation of the black-market premium to identify the level of trade restrictions	A strong positive relationship exists between trade orientation and economic performance. The relationship is robust to the method of estimation, to correction for errors in variables and to the deletion of outliers

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
	used; (4) OPEN1, overall openness index obtained from the unscaled trade model; (5) OPEN2, overall openness index obtained from the residuals of the scaled trade model; (6) OPENM1, manufacturing sector openness index obtained from the unscaled trade model; (7) OPENM2, manufacturing sector openness index obtained from the scaled heteroskedastic trade model			
32. Egwaikhide (1992)	(1) Natural log of real value of oil exports; (2) Natural log of real value of oil imports	Time series, annual data for Nigeria from 1973 to 1978	Instrumental variable technique; uses the natural logs of the lags of real imports and the price for crude oil as instruments for real exports	Crude oil exports have only marginally stimulated the growth of output in Nigeria

33. Giles, Giles, and McCann (1992)	Real value of exports	Time series, annual data for New Zealand from 1963 to 1991	Granger causality test	The study finds mixed evidence for the export-led growth hypothesis when they use growth rates of exports and output. While they reject the hypothesis at the aggregate level, there is some support for the ELG of certain exporting sectors (such as minerals, chemicals, plastics, metal, metal products, live animals, and meat). However, replacing levels of exports and output with growth rates, there is causality from real exports of manufactured goods, meat, and live animals to real GDP
34. Hutchison and Singh (1992)	Growth rate of real value of exports	Time series, annual data for 34 developing countries from the 1950s to the 1980s	Granger causality test	11 countries showed evidence of significantly positive export externality effects contributing to nonexport sector domestic growth. Three countries showed evidence of externalities

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
				in the nonexport sector contributing to export growth and three countries showed evidence of negative externalities
35. Levine and Renelt (1992)	(1) Exports as a percentage of GDP; (2) Imports as a percentage of GDP; (3) Leamer's (1988) openness measure based on factor-adjusted trade; (4) Leamer's (1988) trade-distortion measure based on H-O deviations; (5) Black-market exchange-rate premium; (6) Dollar's (1992) real exchange-rate distortion for SH benchmark countries	Pooled data for 119 countries from 1960 to 1989	No identification strategy	Significant positive correlation between share of investment in GDP and share of trade in GDP. None of the measures of openness are robustly correlated with growth when other explanatory variables are introduced
36. Marin (1992)	Log of exports of manufacturing goods (quarterly observations)	Time-series data for the United States, the United Kingdom, Japan, and Germany	Granger causality test	Exports cause labor productivity in all the countries studied. The hypothesis of export-led growth cannot be rejected

37. Serletis (1992)	(1) Log of exports; (2) Log of imports	Time series, annual data for Canada from 1870 to 1985	Granger causality test	Causality from exports to growth but no causality from imports to growth
38. Sheehey (1992)	(1) Share of exports in GDP; (2) Growth rate of share of exports in GDP; (3) Average annual growth rate of exports	Pooled data for 53 nonoil exporting and six developed countries from 1960 to 1981	No identification strategy	Significant positive impact of exports only during favorable world demand and in industrialized countries
39. Wilbur and Haque (1992)	Natural log of exports	Time series, annual data for 11 developing and six developed countries from 1960 to 1972 and 1973 to 1987	No identification strategy	Supports the “export expansion” hypothesis, emphasizing the significant importance of exports in the formation of savings for both highly developed and less developed countries
40. Alege (1993)	(1) Real value of total exports; (2) Real value of oil exports	Time series, annual data for Nigeria from covering the period of 1960-1985	Granger causality test	The results show the existence of strict econometric exogeneity between exports and GDP and a unidirectional causality from GDP to oil exports. No causality from total export growth (or oil export growth) to GDP growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
41. Dodaro (1993)	Manufactured exports as a percentage of total merchandise exports	Time series, annual data for a wide range of LDCs	Granger causality test	The causality test shows very little support for the basic export-promotion contention. There are two main results of this study: first, the level of development is an important determinant of the degree of manufacturing and processing in a country's export basket; second, the efficiency of export-led growth depends on the composition of exports and the stage of development
42. Ghartey (1993)	Log of exports	Time series, seasonally adjusted quarterly data for Taiwan, Japan, and the United States	Stepwise Granger causality test (Hsiao's version of Granger causality test)	Exports cause economic growth in Taiwan; economic growth causes exports in the United States; and there is bidirectional causality in Japan

43. Gordon and Sakyi-Bekoe (1993)	Real value of exports	Time series, annual data for Ghana from 1955 to 1987	Granger causality test, Sims model, modified Sims model, Akaike final prediction (error model) and the nonparametric multiple rank F-test model	Causality results vary depending upon the model used for testing and the lag specification defined. The violation of the normality assumption leads the Granger model to the apparently incorrect conclusion that GDP causes exports
44. Khan and Saqib (1993)	Natural log of real value of total exports, manufactured exports and primary exports	Time series, annual data for Pakistan from 1972 to 1988	Simultaneous equations; uses foreign income and the index of domestic export prices to world export prices to identify the export demand function, and uses the effective exchange rate index and the relative price index to identify the export supply function	A strong positive association between export performance and economic growth is found, but more than 90% of the contribution of exports to economic growth is indirect in nature
45. Oskooee and Asle (1993)	Real value of exports	Time series, quarterly data for nine LCDs	Granger causality tests with the new techniques of cointegration and error-correction models	The results indicate bidirectional causality for all countries (except for Malaysia for which there is no cointegration) between export growth and economic growth when the cointegrating properties of the time series are incorporated into the analysis

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
46. Oxley (1993)	Natural log of real value of exports	Time series, annual data for Portugal from 1833 to 1985	Granger causality test	The export-led growth hypothesis is rejected in favor of reverse causality
47. Sengupta (1993)	(1) Export growth; (2) Export growth × share of exports in GDP	Time series, annual data, mainly focused on newly industrialized countries (NICs) in Asia including Japan, the Philippines, South Korea, and Taiwan, from the 1960s to the 1980s	No identification strategy	The empirical evidence supports the basic premises of the new growth theory in the case of the successful NICs in Asia and for Korea, in particular
48. Atesoglu (1994)	Growth rate of real value of exports	Time series, annual data for the United States from the 1960s to the 1980s	Instrumental variables technique; uses the rate of growth of external prices and the rate of growth of the world's real income to instrument for exports	The export-led Kaldor growth model interprets and predicts the short-run fluctuation in the growth rate of the US economy. Moreover, the dynamic Harrod foreign trade multiplier rule of Thirlwall is not preferable to the Kaldorian model for explaining annual changes in the US rate of growth

49. Coppin (1994)	(1) Real exports; (2) Growth rate of real exports; (3) Share of manufactured exports in total exports; (4) Share of the growth rate of manufactured exports in total exports	Pooled data for 59 low- and middle-income countries during the 1980s	No identification strategy	Positive impact of exports (but not manufactured exports) on economic growth. The results also show that the growth in the labor force was an important factor in explaining growth
50. Dutt and Ghosh (1994)	Real value of exports	Time series, annual data for 26 low-, middle-, and high-income countries from 1953 to 1991	No identification strategy	For the majority of the countries in the sample, export growth and economic growth have moved together
51. Greenaway and Sapsford (1994a)	Growth rate of share of exports in GDP	Cross-sectional data for 104 countries from 1960 to 1988	No identification strategy	The results support the hypothesis that exports and growth are positively correlated. This relationship is robust to variations in the composition of the country samples. Moreover, the strength of the association varies from subperiod to subperiod, but the positive association becomes stronger over time. Export growth does not always follow reform

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
52. Greenaway and Sapsford (1994)	Growth rate of share of exports in GDP	Time series, annual data for five developing countries from 1956 to 1988	No identification strategy	Significant positive impact of exports on economic growth for Sri Lanka, but negative for Turkey
53. Hansen (1994)	(1) Growth rate of exports \times percentage of exports in GDP; (2) Growth rate of exports of manufactured goods and services \times percentage of manufacturing and service exports in GDP; (3) Growth rate of “other” commodity exports \times percentage of “other” commodity exports in GDP	Time series, annual data for New Zealand from 1968 to 1991	Engle-Granger cointegration test	Exports of manufactures and services have had a greater effect on the economy's recent growth experience than exports of primary products and raw materials
54. Hotchkiss, Moore, and Rockel (1994)	(1) Annualized growth rate of exports weighted by the proportion of exports in GNP; (2) Annualized growth rate of exports (unweighted)	Annual data for 85 countries from 1960 to 1986	No identification strategy	The results indicate that exports contribute to growth through both a sector-externality effect and a factor-productivity effect for middle income countries, but only through a factor-productivity effect for low-income countries

55. Love (1994)	Export growth	Time series, annual data for 20 developing countries from the 1960s to the 1990s	Granger causality test	Substantial support for the hypothesis that exports cause growth. In 14 out of 20 countries there is evidence of causality from export growth to GDP growth, with causality being positive for 10 countries and negative for 4. The export-led results are highly sensitive to the definition of economic growth
56. Sengupta and Espana (1994)	(1) Time derivative of exports; (2) Average share of exports in GDP; (3) Real export growth	Time series, annual data for Japan, the Philippines, Taiwan, Thailand, and South Korea from the 1960s to the 1980s	Engle-Granger cointegration test	The externality effect of exports and their productivity growth played the role of a catalyst for other sectors' growth. Exports and output are cointegrated and there is evidence of a significant positive impact of exports on economic growth, particularly in South Korea
57. Sharma and Dhakal (1994)	Growth rate of real value of exports	Time series, annual data for 30 developing countries from 1960 to 1988	Granger causality test	Causal relationship between export growth and output growth is found in five countries; export-led growth is found in six; output

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
				growth causes export growth in eight; no causal relationship for remaining countries. No causal relationship between export and output growth for South Korea
58. Ukpolo (1994)	Annual percentage growth rates of exports of fuel products	Time series, annual data for eight low-income African countries from 1969 to 1988	No identification strategy	Significant positive impact of nonfuel primary exports on economy
59. Yaghmaian (1994)	(1) Exports in constant prices; (2) Average annual growth rate of exports	Pooled and time-series data for 66 developing countries from 1971 to 1980 and 1981 to 1990	No identification strategy	A positive and statistically significant association was obtained between export growth and the growth of output when population statistics were used for the labor variable in the neoclassical growth model. However, when substituting employment for population, there was no statistical support for export-led growth

60. Ahmad and Harnhirun (1995)	Real exports evaluated in per capita terms in 1980 prices	Time series, annual data for five Asian countries from 1966 to 1990	Granger causality test with error-correction model	Causality between exports and economic growth in both directions for Singapore
61. Amirkhalkhaili and Dar (1995)	Growth rate of real value of exports	Time series, annual data for 23 developing countries from the 1960s to the 1990s	No identification strategy	There is evidence of export-led growth for all except the strongly inward-oriented group of countries. No major difference between moderately inward-oriented countries and moderately and strongly outward-oriented countries in terms of the impact of export expansion on economic growth
62. Clark (1995)	(1) Qualitative evaluation of trade orientation (World Bank Development Report 1987), including effective rates of protection, export incentives, exchange rate alignments, and direct controls; (2) Real export growth	Time series and pooled, annual data for 35 developing countries from 1973 to 1985	No identification strategy	Significant positive impact of exports and outwardly oriented trade policy on industrial diffusion rather on output growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
63. Coe and Helpman (1995)	(Ratio of imports of goods and services to GDP) \times (Natural log of foreign R&D capital stock)	Pooled data for 21 OECD countries and Israel from 1971 to 1990	No identification strategy	Demonstrates the extent to which a country's total factor productivity depends not only on domestic R&D capital but also on foreign R&D. Foreign R&D has beneficial effects on domestic productivity and the result is stronger when the country is more open to foreign trade
64. Holman and Graves (1995)	First difference of log of exports	Time series, annual data for South Korea from 1953 to 1990	Granger and Sims causality tests	The Sims results found consistent two-way exogeneity between exports and GNP growth. The Granger causality results indicate that there is bidirectional causality between exports and economic growth
65. Jin (1995)	(1) Natural log of real value of goods exported; (2) Natural log of world commodity price level for all exports (used as a proxy for foreign price shocks); (3) Natural log	Time series, seasonally adjusted quarterly data for Hong Kong, Singapore, South Korea, and Taiwan from 1973 to 1993	Granger causality tests	The results from the variance decompositions indicate significant feedback relations between exports and output for Hong Kong, Singapore, and South

	of industrial production index for all industrial countries (used as a proxy for foreign output shocks)	Korea, which can be interpreted to suggest that in the short run, economic growth and export growth reinforce each other. However, the Granger cointegration test suggests that no long-run equilibrium exists between exports and the level of output		
66. Jin and Yu (1995)	Exports of goods and services	Time series, seasonally adjusted quarterly data for Korea, Japan, Canada, and the United States	Granger causality test	The results do not support the causal implication of the export-led growth hypothesis. For Korea and Japan, there is bidirectional causality between export growth and GNP growth; for Canada and the United States, growth led exports
67. Kwan and Kwok (1995)	Export growth	Time series, annual data for China from 1952 to 1985	Granger causality test	Exports lead to growth, including instantaneous causality
68. Lee (1995)	(1) Ratio of imported to domestic capital goods; (2) Ratio of imports to investment; (3) Share of imports in GDP	Cross-sectional data for 79 countries from 1960 to 1985	Instrumental variable technique; uses land size, distance from trade distortion as instruments for openness measures	The ratio of imported to domestically produced capital goods in the composition of investment has a significant positive impact on per capita

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
				income growth rates across countries, and for developing countries in particular. Moreover, the ratio of foreign to domestic components of investment is an important factor in economic growth. The policy implication is that any trade distortions that restrict the importance of capital goods hurt the economy in the long run
69. McCarville and Nnadozie (1995)	Real value of exports	Time series, annual data for Mexico from 1926 to 1988	Granger causality test	Supports the export-led growth argument
70. Paul and Chowdhury (1995)	Natural log of growth rate of real exports	Time series, annual data for Australia from 1949 to 1991	Granger causality test	There is evidence of Granger causality running from exports to GDP growth, implying that expansion of exports promotes economic growth in Australia

71. Rashid (1995)	Growth of the real value of exports	Time series, annual data for India from 1960 to 1989	2SLS, four equations multivariate simultaneous model; uses RER and LW to instrument for the growth rate of exports. Note: RER is the nominal exchange rate \times producer prices in the United States divided by India's CPI as a percentage change; LW is world income	The results support that idea that trade is the handmaiden of growth and suggest that domestic factors play a more important role in increasing growth rates. Furthermore, the study shows that liberalization has not had a significant effect on industrialization and investment in the study period
72. Abhayaratne (1996)	Logs of real exports and imports (measured in 1985 prices)	Time series, annual data for Sri Lanka from 1960 to 1992	Granger causality test	The results reject the notion that foreign trade stimulates economic growth
73. Amoateng and Amoako-Adu (1996)	Growth rate of real value of exports	Pooled (including time-series and cross-sectional data) data for 35 African countries from 1970 to 1990	Granger causality test (trivariate model, introducing external debt servicing as a third variable)	Both the export-driven GDP growth and GDP growth-led export-promotion hypotheses are supported. During the 1983-1990 subperiod, the structural adjustment programs that removed some of the economic distortions and encouraged regular repayment of the external debt also improved economic outcomes in the countries studied

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
74. Bodman (1996)	Logs of manufacturing goods exports	Time series, seasonally adjusted quarterly data for Australia and Canada from 1960 to 1995	No identification strategy	Export-led growth hypothesis holds for both economies. The reverse causality is rejected for both countries, except for the Canadian manufacturing sector, for which there is a small, significant positive effect of labor productivity on manufactured exports
75. Boltho (1996)	Export growth	Time series, annual data for Japan from 1913 to 1937, 1952 to 1973, and 1973 to 1990	Granger causality test (in both directions)	For the period 1952-1973, not one of the five tests supports the idea of export-led growth. For the remaining two periods, the export-led growth hypothesis seems to be rejected
76. Burney (1996)	Export growth	Cross-sectional data from 1965 to 1980 and 1980 to 1990, which consists of 89 and 95 countries, respectively	No identification strategy	The study confirms that there is a positive relationship between exports and economic growth. However, the relationship is significant only for the period

				1980-1990. Moreover, the level of significance of the impact of exports on economic growth is relatively higher for high-income economies
77. Cheng and Chu (1996)	Natural log of exports in constant dollars	Time series, annual data for the United States from 1940 to 1990	Granger causality test	The results show bidirectional causality between exports and economic growth
78. Doraisami (1996)	Growth rate of real values of exports	Time series, annual data for Malaysia from 1963 to 1993	Engle-Granger causality test	The study provides strong empirical support for bidirectional causality between exports and output and a positive long-run relationship between exports and growth
79. Dutt and Ghosh (1996)	Real value of exports	Time series, annual data for 26 developed and developing countries	Engle-Granger cointegration and causality tests with error-correction model	The study provides evidence of cointegration in only 14 countries. It shows that export causes growth in five countries, while growth causes exports in two countries. Bidirectional causality is found in three countries, and no causality is found in four countries

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
80. Fosu (1996)	(1) Growth rate of real exports; (2) Mean annual growth rate of nonfuel primary exports; (3) Average annual nonfuel primary exports as a percentage of total exports	Pooled, time-series and cross-sectional data for 76 less developed countries from 1967 to 1986	No identification strategy	The study shows that exports have a positive effect on GDP growth. Moreover, this positive impact generally remains positive when nonexport GDP is the dependent variable. Lastly, the growth of nonfuel primary exports exhibits little or no external impact on the nonexport sector (casting some doubt on the sustainability of primary product export-led growth)
81. Harrison (1996)	(1) Trade reform (1960–1984); (2) Trade reform (1979–1988); (3) Black-market premium; (4) Price distortion; (5) Trade shares; (6) Unprotected agriculture; (7) Movement toward international prices	Pooled, time-series and cross-sectional data for developing countries	Granger causality test	Significant positive impact of openness on economic growth. The Granger causality test shows that causality between exports and growth runs in both directions. However, the choice of time period for analysis is critical

82. Henriques and Sadorsky (1996)	Natural log of real value of exports	Time series, annual data for Canada from 1870 to 1991	Granger causality test	The results tend to favor the one-way Granger causal relationship that the growth rate of GDP influences export growth
83. Jin and Yu (1996)	Real value of exports	Time series, quarterly data for the United States from 1959 to 1992	No identification strategy	Based on the VAR techniques, no significant causal impacts are found between exports and output. This suggests that export expansion is neutral with respect to the growth of the US economy
84. Karunaratne (1996)	Growth rate of value of exports	Time series, seasonally adjusted quarterly data from Australia from 1979 to 1994	Granger causality test with error correction	Trade operated as an engine of growth in Australia during the study period
85. Kwan, Cosotmitis, and Kwok (1996)	Growth rate of real value of exports	Time series, annual data for Taiwan from 1953 to 1988	No identification strategy	The result shows that while the weak exogeneity assumption appears to be valid, the super exogeneity assumption is rejected. Thus, the results cast doubt on policy recommendations based on the export-led growth hypothesis

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
86. Mallick (1996)	Growth rate of real value of exports	Time series, annual data for India from 1950 to 1992	Engle-Granger causality tests with error-correction model	Cointegration and causality test results indicate the existence of strong cointegration and Granger causality between income and exports growth. Causality runs from income to export growth
87. Mollik (1996)	Growth rate of real value of exports	Time series, annual data for Bangladesh from 1972 to 1992	Granger and Sims causality tests	Exports cause economic growth but not the opposite
88. Mulaga and Weiss (1996)	Growth rate of effective protection (ERP): $ERP_i = (t_i - a_{ij} \times i_j) / (1 - a_{ij})$ where t_i and i_j are the tariffs or tariff-equivalents for goods i and j , respectively, and a_{ij} is the share of input j in the value of output i at world prices	Firm-level data for Malawi from 1970 to 1991	No identification strategy	The impact of a decline in protection on total factor productivity depends upon the way productivity growth is measured. The study finds a positive relationship for the simple TFP estimates. However, the growth seems unassociated with changes in protection when capital input is measured by capital services rather than capital stock

89. Piazolo (1996)	(1) Real value of exports of goods and services; (2) Real value of imports of goods and services; (3) Trade orientation dummy variable (import-substitution dummy variable is 0 from 1965 to 1982; export diversification dummy variable is 1 from 1983 to 1992)	Time series, annual data for Indonesia from 1965 to 1992	Engle-Granger causality test	Indonesian long-run economic growth is more investment driven than trade- or export driven. However, the short-run effects of exports are strongly positive
90. Pomponio (1996)	Manufactured goods exports	Time series, annual data for 66 developing and OECD countries from 1960 to 1988	Granger causality test	The bivariate causality tests show weak evidence supporting the idea that manufactured exports lead to output growth. However, when the investment factor is included in the trivariate causality test, there is mixed evidence in support of the trivariate causal relationships
91. Riezman, Whiteman, and Summers (1996)	(1) Growth rate of exports; (2) Growth rate of imports; (3) Share of exports + imports in GDP	Time series, annual data for 126 countries from 1960 to 1985	Granger causality test	30 out of 126 countries support the export-led growth hypothesis; 25 out of 126 countries support the growth-led exports hypothesis. Using a weaker notion of "support," 65 out of 126

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
				countries support the export-led growth hypothesis. For the “Asian Tiger” countries, the relationship between export growth and output growth becomes clearer when conditioned on human capital and investment growth as well as import growth
92. Thornton (1996)	Natural log of real value of exports	Time series, annual data for Mexico from 1895 to 1992	Granger causality test	Positive Granger causal relationship running from exports to economic growth
93. Al-Yousif (1997)	(1) Growth rate of value of exports; (2) Growth rate of exports × share of exports in GDP	Time series, annual data for Saudi Arabia, Kuwait, UAE, and Oman from 1973 to 1993	No identification strategy	Significant positive impact of exports on economic growth. However, economic growth is highly affected by oil
94. Berg (1997)	(1) Growth rate of real value of exports (in Granger causality test); (2) Growth rates of real values of exports and imports (in simultaneous equations)	Time-series data for Mexico from 1960 to 1991	Granger causality test; Simultaneous equations; uses the real exchange rate and the growth rate of real GDP of the United States to	During the period 1960-1991, trade and real output growth were positively related in Mexico. Simultaneous equation regressions suggest that both exports

			instrument for exports; uses the real exchange rate and a measure of foreign capital inflows to instrument for imports	and imports significantly explain the growth of total factor productivity in Mexico. However, this study also points out the limitations of empirical exercises due to the changing composition of exports and shifting policy regimes, and emphasizes the importance of theoretical work
95. Coe, Helpman, and Hoffmaister (1997)	(1) Change of the share of machinery and equipment imports from industrial countries in GDP in each developing country (denoted by M); (2) M log of foreign R&D capital stock	Pooled data for 77 developing countries and 22 industrialized countries from 1971 to 1990	No identification strategy	The results imply that a developing country's total factor productivity is larger the larger is its foreign R&D capital stock, the more open it is to machinery and equipment imports from industrial countries and the more educated is its labor force
96. Gani (1997)	Growth rate of the (real) share of exports in GDP	Time series, annual data for Papua New Guinea from 1970 to 1992	No identification strategy	Exports and stable exchange rates have positively contributed to economic growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
97. Ghatak, Milner, and Utkulu (1997)	(1) Natural log of the real value of exports (in Granger causality test); (2) Real value of exports of manufactured goods, fuel, and nonfuel primary products (in Johansen maximum likelihood procedure)	Time series, annual data for Malaysia from 1955 to 1990	Granger causality test	The results provide support for the export-led growth hypothesis; aggregate exports Granger-cause real GDP and nonexport GDP. This relationship is found to be driven by manufactured exports rather than by traditional exports
98. Gokkceus (1997)	Natural log of the industry-specific annual total protection rate, including all charges on imports such as customs duty, municipality tax, stamps, funds, etc.	Panel data with annual observations for 29 four-digit rubber industry plants in Turkey from 1983 to 1986	No identification strategy	This study reveals that trade liberalization or declining protection have a significant positive effect on productivity growth
99. Greenaway, Morgan, and Wright (1997)	(1) Growth in real value of merchandise exports; (2) Percentage change in exports	Pooled data for 74 developing countries from the 1980s to the 1990s	No identification strategy	In general, liberalization seems to have been associated with deterioration in growth for this sample; the result was robust for both the full and the restricted samples

100. Liu, Song, and Romilly (1997)	Natural logs of exports and imports	Time series, quarterly data for China from 1983 to 1995	Granger causality test	Bidirectional causality between economic growth and exports plus imports, which is consistent with China's development strategy of protected export-promotion
101. Park and Prime (1997)	(1) Growth rate of exports; (2) Share of exports in GDP; (3) Growth rate of exports × share of exports in GDP	Pooled and cross-sectional data for 26 inland provinces and 11 coastal provinces in China	No identification strategy	The results support the hypothesis that exports have contributed to the growth of provincial incomes in China for the period examined with both the cross-sectional and pooled analyses, with the results being primarily driven by the comparison between the coastal provinces and the inland areas
102. Pineres and Ferrantino (1997)	(1) GEXP, the growth rate of the real value of exports; (2) RXR, the real exchange rate; (3) TRAD7 is the variance of the traditional index calculated across industries; (4) SpecI, a static measure of specialization; (5) CSX, a measure of the change in export composition	Time series, annual data for Chile from 1962 to 1992	Granger causality test	In general, growth periods in Chile have been associated with stability in the composition of exports and an acceleration of export diversification. There is a positive relationship between export diversification and economic growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
103. Begum and Shamsuddin (1998)	(1) Level of exports; (2) Share of exports in GDP; (3) Growth rate of exports; (4) Export growth rate weighted by the share of exports in GDP. Note: OLS results use the last measure as the exports variable	Time series, annual data for Bangladesh from 1961 to 1992	Instrumental variable technique; uses the annual growth rate of world income and the difference between foreign and domestic inflation rates to instrument for exports	Export growth has significantly increased economic growth through its positive impact on total factor productivity in the economy. The contribution of exports to economic growth was more pronounced during 1982–1990, when the government pursued a policy of trade liberalization and structural reform and political turmoil was not persistent
104. Biswal and Dhawan (1998)	Natural log of total real value of exports and total real value of manufactured goods exports	Time series, annual data for Taiwan from 1960 to 1990	Granger causality test with error-correction model	The study finds evidence of bidirectional causality between exports and growth
105. Clerides et al. (1998)	A dummy variable that equals 1 if the plant exports, 0 otherwise	Plant-level, panel data for Colombia (1981–1991), Morocco (1984–1990), and Mexico (1984–1990)	(1) FIML; simultaneously estimates export market participation patterns and marginal cost	The study addresses the question of whether the association between exporting and efficiency reflects causation flowing from exporting

			realizations; (2) GMM; used to test the robustness of the marginal cost estimate	experience to improvements in performance. The actual data suggest an inconsistent causality pattern (no-learning-by-exporting scenario). The FIML and GMM tests confirm that the association between exporting and efficiency is most plausibly explained as low-cost producers choosing to become exporters
106. Doyle (1998)	Natural log of the real value of exports	Time series, annual data for Ireland from 1953 to 1993	Granger causality test with error correction	Exports and GDP are cointegrated. Evidence of short-run and long-run causality runs from exports to growth
107. Iscan (1998)	Share of exports in GDP	Panel, sector-level data for Mexico from 1970 to 1990	Granger causality test with VAR model	The study found no evidence that exports lead to capital accumulation or vice versa. However, there is evidence that common determinants, such as the real exchange rate, may be the driving force for the capital accumulation

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
108. Islam (1998)	Value of exports	Time-series data for 15 Asian countries from 1967 to 1991	Multivariate error-correction model to test the Granger causality between exports and growth	Causality test results indicate that export expansion causes growth in two third of these countries, corrected for simultaneity between the causal factors. A country with a large public sector, higher level of economic development, and lower vulnerability to external economic shocks, is more likely to reap the benefits of export-promotion strategies
109. Keller (1998)	A foreign knowledge stock variable which is a weighted sum of the cumulative R&D expenditures of a country's trading partners (the weights are given by the bilateral import shares)	Panel data for 21 OECD countries from 1971 to 1990	Monte Carlo-based robustness test; R&D spillovers among randomly matched trade partners investigated	Randomly matched trade patterns give rise to greater positive R&D spillovers than do true bilateral trade patterns, suggesting that the Coe and Helpman (1995) model of trade-related international R&D spillovers is not robust

110. McNab and Moore (1998)	Dummy variables for inwardly oriented trade policy and outwardly oriented trade policy	Pooled data for 41 countries	Simultaneous estimations; openness measures are used as instruments	Outward trade policy increased annual GDP, and trade policy is a robust determinant of growth
111. Onafowora and Owoye (1998)	Real value of exports and trade policy dummies	Data from 12 sub-Saharan African (SSA) countries from 1963 to 1993	Johansen's cointegration tests and VECM	Changes in trade policies and exports have positive effects on growth for 10 out of 12 SSA countries
112. Shan and Sun (1998a)	Growth rate of real value of exports	Quarterly, seasonally adjusted data for Australia from 1978 to 1996	Granger causality test using Toda and Yamamoto method	Results indicate one-way causality from industrial growth to export growth with a 1-year lag
113. Shan and Sun (1998b)	Natural log of exports	Monthly, seasonally adjusted data for China from 1978 to 1996	Granger causality test using Toda and Yamamoto method	Results indicate bidirectional causality between exports and real industrial output in China
114. Shan and Sun (1998c)	(1) Natural log of the growth rate of exports; (2) Natural log of the growth rate of imports	Quarterly, seasonally adjusted data for Hong Kong, South Korea, and Taiwan from 1978 to 1996	Granger causality test using Toda and Yamamoto method, followed by sensitivity analysis	Principal results from the paper cannot offer support for the export-led growth hypothesis. Hong Kong and Korea show bidirectional Granger causality between exports and growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
115. Pineres and Ferrantino (1999)	(1) Natural log of real value of exports; (2) The mean of a cumulative export experience function for each commodity (traditionally scores); (3) A static measure of specialization which approaches 1 implying a high degree of specialization and nears 0 suggesting a degree of export diversification	Annual data for Colombia from 1962 to 1993	(1) Cointegration and error-correction modeling used to test relationship between GDP and exports; (2) simultaneous equation system: using the price of coffee, oil, lagged log of real exports, world interest rates as instruments for openness measures	Traditional Granger causality tests reveal little evidence of export-led growth. However, analysis of structural export change and export diversification sheds significant light on the trade-growth linkages. Results reveal a positive interaction between structural changes in the export sector and Colombian GDP growth. Results indicate that increased export diversification leads to more rapid growth in real exports; and that more rapid structural change in exports is associated with accelerated growth in Colombian GDP
116. Clark, Sawyer, and Spriukle (1999)	Trade policy orientation, measured using Dollar's index of real exchange rate distortion	Pooled data for 94 developing countries	No identification strategy	Outward-oriented trade strategies are found to have an important effect on industrialization

117. Dhawan and Biswal (1999)	(1) Natural log of real value of exports; (2) Terms of trade	Annual data for India from 1961 to 1993	VAR model; Johansen test for cointegration; Engle and Granger's error-correction approach (to determine the direction of causal flow in the short run as well as in the long run)	The results suggest that there is one long-run equilibrium relationship among real GDP, real exports and terms of trade, and the causal relationship flows from GDP growth and terms of trade to export growth. The causality from exports to GDP appears to be a short-run phenomenon
118. Frankel and Romer (1999)	Trade is measured using instrumental variables that take account of countries' geographic locations	Cross-sectional data for 150 countries	OLS and instrumental variables techniques	Trade has a positive effect on income growth
119. Greenaway, Morgan, and Wright (1999)	(1) Real value of exports; (2) Decomposition of exports into different categories	Panel data for 69 countries from 1975 to 1993	No identification strategy	Results report a strong positive relationship between real export growth and real output growth. The study also finds that export composition does matter
120. Moosa (1999)	Sum of exports of goods and services	Annual data for Australia from 1900 to 1993	Cointegration and causality tests	Results fail to detect either a long-run or short-run relationship between exports and growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
121. Vamvakidis (1998)	(1) Share of exports plus imports in GDP; (2) Economy judged "open" if it meets five conditions from Sachs and Warner: (a) average tariff less than 40%; (b) average nontariff barriers less than 40%; (c) black-market premium less than 20% of official exchange rate; (d) government is not communist; (f) no state monopoly on major exports	Data cover the period from 1950 to 1992. Data for real GDP <i>per capita</i> , investment share, and population growth are from the Penn World Table; trade shares come from the World Table (World Bank, 1994), and school enrollment ratio are from Barro and Lee (1994). The analysis focuses on 109 countries engaged in regional trade agreements (RTA) and 51 countries engaged in broad liberalization between 1958 and 1989	No identification strategy	Economies grow faster after broad liberalization, in both the short run and the long run; however, they grow more slowly after an RTA
122. Weinhold and Rauch (1999)	Share of exports plus imports in GDP	Data from 39 countries 1960–1990	No identification strategy	Increased trade may lead to growth

123. Xu and Wang (1999)	Foreign R&D spillovers weighted by total imports, capital goods imports, and noncapital goods imports	Data from 21 OECD countries from 1983 to 1990	No identification strategy	Trade in capital goods was found to be a significant channel of R&D spillovers
124. Anoruo and Ahmad (1999)	Growth rate of the share exports plus imports in GDP	Time series, annual data for Indonesia, Malaysia, the Philippines, and Thailand from 1960 to 1997	Granger causality test	The results show bidirectional causality between economic growth and openness
125. Al-Marhubi (2000)	(1) Average share of exports in GDP from 1961 to 1988; (2) Export structure, measured as the ratio of manufactured good exports to total exports from the World Development Indicators (WDI) 1997	Cross-sectional data for 91 countries from 1961 to 1988	No identification strategy	The results show that export diversification is associated with faster growth. This relationship is economically large and is robust to different model specifications and different measures of export diversification. Moreover, distortions to international trade and market-oriented resource allocation that run counter to a country's comparative advantage can have adverse effects on economic efficiency and growth performance

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
126. Anwer and Sampath (2000)	Natural logs of goods and nonfactor service exports	Time series, annual data for 97 countries from 1960 to 1992	Granger causality test	In 97 countries, GDP and exports are integrated of different orders for 36 countries. Among the other 61 countries, 17 countries exhibit no long-run relationship between the two variables; 35 countries show causality in at least one direction (10 countries show unidirectional causality from GDP to exports, 5 show unidirectional causality from exports to GDP, and 20 show bidirectional causality), and 9 countries do not show any causality
127. Choudhri and Hakura (2000)	Sector openness index, which depends on how international trade affects the technology transfer process	Panel data for 44 countries (including 33 developing countries) from 1970 to 1993	No identification strategy	The effect of increased openness on productivity growth differs across sectors. In traditional (low-growth) manufacturing sectors, no

				effect or little effect exists; for medium-growth sectors, import competition has a significant growth-enhancing effect; and there is some evidence to show that export expansion in high-growth sectors leads to an increase in productivity growth
128. Darrat, Hsu, and Zhong (2000)	Real value of exports	Data for Taiwan from 1953 to 1988	VECM model after Johansen cointegration tests, instead of simple Engel-Granger procedure	The exports are both weakly and super-exogenous to output
129. Erlat (2000)	Export and import flows, divide into three categories: net exporting, import competing, noncompeting	Annual data for Turkey from 1963 to 1994, divided into four subperiods	No identification strategy	First, trade plays a more significant role in employment changes during the post-1980 periods; this is observed more in the net exporting and noncompeting categories rather than the import-competing category. Second, the switch to export-oriented growth in 1980 did not lead export-based employment to be

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
				dominant in employment changes, but has acted as a buffer in the sense that employment may either have grown much less or declined more severely if the post-1980 expansion of exports had not occurred
130. Jin (2000)	Share of imports in GDP	Annual data for Japan, South Korea, Japan, the Philippines, Thailand, Malaysia, and Singapore from 1955 to 1995	VAR model	The results do not strongly support the idea that openness leads to growth
131. Kim (2000)	Direct policy measures such as legal rates of tariff, quota ratios, and nominal rates of protection	Panel data for 36 Korean manufacturing industries over nine subperiods from 1966 to 1988	No identification strategy	Total factor productivity estimates based on the standard assumption of perfect competition and constant returns are biased and show a spurious relationship with changes in trade regimes. When both imperfect

				competition and nonconstant returns are taken into account, the growth accounting approach yields estimates of total factor-productivity growth that are quite low
132. Morley and Perdikis (2000)	Growth rate of exports	Annual data for Egypt from 1955 to 1996	VECM and Granger causality tests	There appears no evidence of either direction of causality between exports and growth
133. Moreira and Najberg (2000)	Value of exports and imports	National accounts series for Brazil from 1990 to 1997	No identification strategy	Trade liberalization has a negative short-run impact on employment
134. Pereira and Xu (2000)	Growth rate of exports	Data for 39 countries from 1960 to 1992	Multivariate VAR approach test to Granger relations	Results strongly support the theory of export-led growth
135. Francisco and Ramos (2001)	Real value of exports and imports	Annual data for Portugal from 1865 to 1998	Granger causality in the ECM-VAR	The empirical results reject unidirectional causality between exports, imports, and economic growth. However, there are feedback effects between exports, output growth and import growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
136. Ghirmay, Grabowski, and Sharma (2001)	Real value of exports	Time-series data for 19 least developed countries (each with at least 30 annual observations)	Granger causality tests based on error-correction modeling	Export expansion leads to economic growth in 15 out of 19 countries by either increasing the volume of investment, improving efficiency, or both
137. Khalafalla and Webb (2001)	Real value of exports and imports (export data are broken down into two additional series: total value of primary commodity exports, and total value manufacturing exports)	Quarterly data for Malaysia from 1965 to 1996	VAR model, VECM, Granger causality test	Statistical tests confirm export-led growth for the full period and for the period to 1980, but tests on the 1981–1996 period show growth causing exports. Primary exports have a stronger direct impact on economic growth than manufactures
138. Madden, Savage, and Bloxham (2001)	Share of imports in $GDP \times \text{foreign R\&D}$ capital stock	Panel data for 15 OECD countries and six Asian countries from 1980 to 1995	No identification strategy	Total factor productivity and domestic R&D are positively related. Domestic R&D has relatively large impact on total factor productivity

139. Abdulnasser (2002)	Value of exports	Quarterly data for Japan from 1966 to 1999	Granger causality test in VAR model, LR test for causality and bootstrap simulation technique	There is bidirectional causality between export and output growth in Japan
140. Chuang (2002)	A set of trade-induced learning variables that take into account trade partners and learning characteristics	Cross-country panel data for 78 countries from 1960 to 1985	Panel regression. The general problem of causality is relatively minor in this study as only traded goods in a subset are taken	Holding other variables constant, trade-induced learning has a positive and significant effect on growth
141. Greenaway, Morgan, and Wright (2002)	(1) A dummy variable, equal to 1 if it is activated at the time of a country's first World Bank Structural Adjustment Loans (SALs) (2) a dummy variable based on tariffs, quotas, export impediments and promoters, and exchange rate misalignment; (3) a dummy variable for whether an economy is open or closed based on five criteria from Sachs and Warner: nontariff barrier coverage; average tariffs; the black-market exchange rate premium;	Panel data for 73 countries from 1975 to 1993	No identification strategy	Liberalization may favorably impact growth of real GDP per capita. However, the effect would appear to be lagged and relatively modest

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
	whether the economy is socialist or not; whether a state monopoly exists over major exports			
142. Jin (2002)	Real value of exports	Monthly data for Korea's four largest provinces (Seoul, Kyunghee, Kyungnam, and Pusan) from 1987 to 1996	Engle and Granger tests for cointegration, Granger causality tests, four-variable VAR model, variance decompositions and impulse responses	Export growth has a significant impact on output growth for all provinces, although a feedback effect from output to export growth appears in Seoul and the Kyungnam province
143. Ahmed (2003)	(1) Share of exports in GDP; (2) Average tariff collection rate	Times series data for Bangladesh from 1974 to 1996	VECM model after cointegration tests	Exports and average tariff collection rates are significant in explaining GDP growth
144. Connolly (2003)	Sum of high technology imports and nonhigh technology imports, which are measured as total imports from the world, excluding high technology goods imported from developed countries	Annual panel data for 86 countries from 1965 to 1995 (data are not available for all countries in all years)	No identification strategy	High technology imports positively affect domestic imitation and innovation. Moreover, their role is greater in developing nations. Finally, foreign technology embodied in imports plays a greater role in growth than domestic technology

145. Dar and Amirahalkhali (2003)	(1) Real value of exports; (2) Share of exports plus imports in GDP	Data for 19 countries from 1971 to 1999	No identification strategy	Results generally indicate that trade openness is important for economic growth, but the magnitude of the relationship varies significantly across countries
146. Ferreira and Rossi (2003)	(1) Effective rate of protection; (2) Nominal tariff	Annual data for 16 industrial sectors in Brazil from 1985 to 1997	Panel regression includes IV (the measurement of effective rate of protection and mean nominal tariff), 2SLS FE, and FE, no causality tests used	There is significant evidence that reducing trade barriers improves total factor productivity and labor productivity
147. Singh (2003)	Share of real exports in real GDP for each industry	Panel data for 10 manufacturing industries in India from 1973/1974 to 1993/1994	No identification strategy	Exports do not induce convergence, but instead seem to accentuate the process of divergence among industries. The study provides some evidence for the significant effects of exports on the level of output per capita and total factor productivity in the manufacturing sector. The effects of exports on total factor productivity are significant in half of the sample industries, but are statistically insignificant in the remaining half

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
148. Yanikkaya (2003)	(1) Trade volumes (import penetration; export share in GDP; trade with OECD countries; trade with non-OECD countries; US bilateral exports; US bilateral imports and population density), (2) Trade restrictions (import duties, export duties, taxes on international trade, bilateral payment arrangements, and other measures of trade barriers)	Panel data for over 100 developed and developing countries from 1970 to 1997	Cross-country regression, including OLS, SUR, 3SLS; instrumental variables technique, using log of average GDP per capita for the previous 5 years, 5-year lagged value of life expectancy, actual value of telephone mainlines, the access to international water, war deaths, tropical climate, and political regime as instruments for trade measures	The regression results for trade volumes provide substantial support for the hypothesis that trade promotes growth through a number of channels such as technology transfer, scale economies, and comparative advantage. Interestingly, all measures of trade barriers used in the study are significantly and positively correlated with growth except for restrictions on current account payments, which are negatively but insignificantly correlated with growth. Thus, the results provide considerable evidence for the hypothesis that restrictions on trade can promote growth, particularly for developing countries and under certain conditions

149. An and Iyigun (2004)	Skill-intensive exports	Panel data for 86 countries from 1970 to 1990	Panel regression, followed by robustness tests including dropping outliers, causality test for direction of relationship between rate of <i>per capita</i> GDP growth and subsequent skill content of exports	After controlling for GDP per capita, education, openness to foreign trade, and political and macroeconomic stability, a higher export content of skill-intensive goods generates higher per capita GDP growth rates, but the reverse does not hold
150. Balaguer and Cantavella-Jorda (2004)	(1) Real value of exports; (2) Export composition (in relative terms), calculated as share of consumption goods, semimanufactured and capital goods in total exports	Annual data for Spain from 1961 to 2000	Johansen's cointegration test followed by Granger causality tests	Exports and growth are cointegrated, and there is bidirectional causality between exports and growth
151. Bugsten et al. (2004)	A dummy variable, equal to 1 if a firm exports more than a certain amount (depending on company size, labor productivity, capital productivity, etc.), 0 otherwise	Firm-level, panel data for the manufacturing sectors of four African countries: Kenya (1992–1994); Ghana (1991–1993); Zimbabwe (1992–1994); Cameroon (1992/1993–1994/1995)	Simultaneous estimation of a dynamic production function and a dynamic discrete choice model for the decision to export	The study finds that exporting increases productivity

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
152. Chuang and Hsu (2004)	(1) Natural log of share of exports to OECD in total industry sales; (2) Natural log of share of imports in total industry sales	Firm-level data for China for 1995	No identification strategy	The presence of foreign ownership has a positive and significant effect on the productivity of domestic firms. Moreover, trading with more advanced countries helps China gain access to new technology and information, which improves China's productivity and enables it to compete in international markets
153. Dawson and Hubbard (2004)	(1) Growth rate of exports; (2) growth rate of exports \times the share of exports in GDP	Annual panel data for 14 Central and East European countries CEECs from 1994 to 1999	First stage, aggregate production into export production and nonexport production; in the second stage, random effects and fixed effects model applied to test relationship between GDP growth and export production variants	Export growth is a significant determinant of GDP growth

154. Dritsakis (2004)	(1) Investment, measured as gross fixed capital adjusted by the GDP deflator; (2) Real export revenue	Quarterly data for Bulgaria and Romania from 1991 to 2001	Johansen cointegration test and Granger causality test based on a vector error-correction model	There is significant evidence to show cointegration between exports and growth, as well as investment and growth. Exports promote growth in both countries; growth also promotes exports in both countries. Investment leads to growth in both countries
155. Dutta and Ahmed (2004)	(1) Real value of exports; (2) Import tariff collection rate	Time-series data for Pakistan from 1973 to 1995	VAR model to test long-run effects, VECM model to test short-run effects, both with human capital included in the model framework	There is a unique long-run relationship between industrial growth and its major determinants including real exports. The short-run relationship between industrial growth and real exports is also significant
156. Dodzin and Vamvakidis (2004)	Share of exports plus imports in GDP	Panel data for 92 developing countries from 1960 to 2000	Granger causality test	The impact of trade liberalization on different sectors of production in developing economies shows increases in the production share of the industrial sector, at the expense of agricultural production

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
157. Falvey, Foster, and Greenway (2004)	Trade-mediated knowledge spillovers (export spillovers and import spillovers)	Data for 21 OECD countries from 1975 to 1990	Griliches' (1979) perpetual inventory method, weighted by imports or exports	Results support the existence of spillovers through imports, but the evidence of spillovers through exports is less compelling
158. Lee, Ricci, and Rigobon (2004)	(1) The share of trade in GDP; (2) Tariff; (3) Import duties; and (4) Black-market premium	A panel data of eight periods of 5 years each, is spanning from 1961 to 1965 and 1996 to 2000, including 100 countries	“Identification through Heteroskedasticity” (IH): exploit plausible differences in variances of structural innovations (error terms) across subsamples of the data	The results suggest that openness has a small positive effect on growth, which is not particularly robust
159. Thangavelu and Rajaguru (2004)	Value of exports and imports	Time-series data for Hong Kong, India, Indonesia, Japan, Malaysia, Philippines, Singapore, Taiwan, and Thailand from 1960 to 1996	VECM approach	The long-run result shows that there is no causal effect of exports on labor productivity growth for Hong Kong, Indonesia, Japan, Taiwan, and Thailand, thereby suggesting that there is no

				export-led productivity growth in these countries. However, significant causal effects were found from imports to productivity growth, suggesting import-led productivity growth in India, Indonesia, Malaysia, the Philippines, Singapore, and Taiwan. In addition, the results indicate that imports tend to have greater positive impacts on productivity growth in the long run
160. Awokuse (2005a)	Real value of exports	Quarterly time-series data from 1963 to 2001 for Korea	VECM approach and augmented levels VAR modeling with integrated and cointegrated processes (of arbitrary orders) used to test Granger causality	There is bidirectional causality between exports and GDP growth

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
161. Awokuse (2005b)	Real value of exports	Quarterly time-series data from 1960 to 1991 for Japan	Augmented VAR methodology used to test Granger causality; Direct acyclic graphs (DAG) used to further restrict VAR	The causal path between exports and GDP growth in Japan is bidirectional; other variables such as capital and foreign output are also significant determinants of productivity growth in Japan
162. Van Biesebroeck (2005)	Value of exports	Panel survey data for approximately 200 manufacturing firms in each of nine African countries from 1992 to 1996	(1) GMM; uses lagged exports to test unidirectional causality from exports to productivity improvement; (2) Instrumental variables technique; uses ethnicity of the owner and the state ownership	There is a causal link from exporting to productivity.

163. Fu (2005)	Value of exports	(1) Panel data of 26 manufacturing industries for the period 1990–1997 for China; (2) Pooled data for 358 subindustries, including 179 state-owned enterprises (SOEs) industries and 179 comparable township and village enterprises (TVE) industries from 1995	Uses a two stage process. First stage uses a nonparametric Malmquist TFP approach to decompose TFP growth into technical change and efficiency improvements. Second stage uses regressions to test the impact of exports on TFP growth	No significant evidence was found in favor of significant productivity gains caused by exports at the industry level
164. Dollar and Kraay (2004)	(1) Trade volumes; (2) Tariffs; (3) Membership in the World Bank; (4) Presence of capital controls	Cross-sectional data for 92 countries and spanning the 1960s to the 1990s	No identification strategy	The study tests whether openness has systematic effects on the share of income accruing to the poorest in society. It finds little evidence to support such systematic effects, even when allowing the effects of openness to depend on the level of development and differences in factor endowments

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
165. Dollar and Kraay (2004)	(1) Decade-over-decade changes in the volume of trade; (2) Reductions in average tariff rates; (3) Share of trade in GDP	Cross-sectional data for 101 countries and covering periods from the 1970s to the 1990s	Instrumental variables technique; uses lagged trade volumes as instruments for current trade volumes	Changes in growth rates are highly correlated with changes in trade volumes, controlling for lagged growth and addressing a variety of econometric difficulties. However, there is no significant correlation between changes in inequality and changes in trade volumes, controlling for changes in average incomes
166. Wacziarg and Welch (2003)	Composite measure of openness including exchange rate reforms			Positive relationship between a composite measure of economic reforms and economic growth, but not significant for the 1990s
167. Alcalá and Ciccone (2004)	The natural log of the share of the sum of imports and exports in purchasing power parity GDP (real openness)	Cross-sectional data from Penn World Tables for 1985	IV using a geography-based instrument for trade (obtained from aggregating bilateral trade shares)	Trade is a significant and robust determinant of aggregate productivity when real openness is used

			<p>predicted by the gravity equation). This method was first used by Frankel and Romer (1999)</p>	
168. Rigobon and Rodrik (2004)	Log of the share of trade in GDP	Cross-sectional dataset from Penn World Tables and World Development Indicator, including 86 countries	<p>“Identification through Heteroskedasticity” (IH): exploit plausible differences in variances of structural innovations (error terms) across subsamples of the data</p>	The share of trade in GDP has a negative impact on income levels and democracy, but a positive impact on the rule of law
169. Chang et al. (2005)	Share of real exports plus real imports in real GDP	Pooled cross-country and time-series data for 82 countries from 1960 to 2000	<p>Nonlinear growth regression specification that interacts a proxy of trade openness with proxies of educational investment, financial depth, inflation stabilization, public infrastructure, governance, labor-market flexibility, ease of firm entry, and ease of firm exit</p>	The growth effects of openness are positive and economically significant if certain complementary reforms are undertaken

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
170. Hausmann et al. (2005)	A measure of the income level of a country's exports, which also captures the level of sophistication of those exports	There are two sources of data. (1) The first is the United Nations Commodity Trade Statistics Database (COMTRADE) covering over 5000 products at the Harmonized System 6-digit level for the years 1992–2003 (available for 124 countries over 1999–2001); (2) the real <i>per capita</i> GDP data from the WDI, which is available for 113 countries	Instrumental variables technique; uses log population and log land area as instruments for the export sophistication measure	Export sophistication positively affects growth
171. Freund and Bolaky (2008)	The log of the share of the sum of imports and exports in GDP	Cross-country data from 126 countries covering periods from 2000 to 2004	IV using a geography-based instrument (followed by Frankel and Romer (1999)) and also using remoteness from other markets as an additional instrument for trade	Openness interacts with regulation and only positively affects growth in unregulated economies with minimal restrictions on entry

172. Romalis (2007)	Share of trade (evaluated at either current or constant price) in GDP	Data for developing countries from 1961 to 2000	Instrumental variables technique; uses US MFN tariffs as instruments for developing country trade shares	Openness positively affects per capita GDP growth
173. Estevadeordal and Taylor (2008)	Tariffs on consumption, capital, and intermediate goods from primary sources	Data on growth rate come from the Penn World Table database (version 6.2); data on average tariff come from the Economic Freedom in the World 2005 database, which are available every 5 years from 1970 to 2000, plus annually for 2001, 2002, and 2003, and the sample size grows from 77 countries in 1970 to 122 in the year 2000	IV using two “GATT Potential” instruments (the interaction of an indicator of GATT membership in 1975 with the pre-Uruguay Round tariff level and the interaction of Great Depression intensity with the initial period tariff level) to identify for trade liberalization (reflected by tariff policy)	The results show that liberalizing tariffs on imported capital and intermediate goods did lead to faster GDP growth, but the negative effects of tariff on growth is not as strong for consumption goods tariffs
174. Donaldson (2009)	Introduction of railroads within Indian states	Historical district by district (239 districts) data within India for 1861-1930	Uses the introduction of railroads as an exogenous shifter of trade costs	

Continued

Table 1 Cross-country studies on openness and growth—Cont'd

Study author and date	Measure of openness	Data description	Identification strategy	Does openness increase growth?
175. Feyrer (2009)	Trade shares	Annual cross country data between 1960 and 1995	Uses geography as an instrument for trade shares in a first stage gravity equation. Differences over time in the importance of air and sea distance lead to time variation.	Trade has a significant effect on income with an elasticity of roughly one half. Differences in predicted trade growth can explain roughly 17 percent of the variation in cross country income growth between 1960 and 1995.

Table 7 Foreign investment and technology spillovers

	Countries	Years	Data	Technology spillovers?	Details
1. Globerman and Steven (1979)	Canada	1972	Industry level	Yes (positive horizontal spillovers)	Positive horizontal spillovers
2. Blomstrom and Persson (1983)	Mexico	1970	Industry level (the Mexican manufacturing industry 1970)	Yes (positive horizontal spillovers)	Positive horizontal spillovers (the existence of spillovers efficiency benefits from the foreign-owned plants to the domestically owned ones)
3. Blomström (1986)	Mexico	Mexican census of manufactures 1970 and 1975	Industry level	Yes (positive horizontal spillovers)	Positive horizontal spillovers (a result of the fact that input market start to operate better because of foreign entry rather than technology transfer)
4. Haddad and Harrison (1993)	Morocco	1985-1989	Firm level	No	No evidence of spillovers (The results reject the hypothesis that foreign presence accelerated productivity growth in domestic firms)
5. Blomstrom and Wolff (1994)	Mexico	1970/1975	Industry level	Yes (positive horizontal spillovers)	Positive horizontal spillovers (suggested by the fact that productivity levels of locally owned firms in Mexico have converged on those of foreign-owned firms)

Continued

Table 7 Foreign investment and technology spillovers—Cont'd

	Countries	Years	Data	Technology spillovers?	Details
6. Caves (1996)	Australia and Canada	Canada (1965–1967) and Australia (1963–1963, 1966–1967)	Industry level	Yes (positive horizontal spillovers)	Positive horizontal spillovers for both Canada and Australia (foreign investment impels higher technical efficiency in competing domestic firms and speeds the transfer of new technology of them). However, this result is not that significant for Canada probably because of the existence of tariffs
7. Kokko (1994)	Mexico	1970	Industry level (the Mexican manufacturing industry 1970)	Large technology and high foreign shares are harmful for spillovers.	The results suggest that advanced MNC technologies or large technology gaps alone do not constitute unsurmountable obstacles to spillovers, but that spillovers are less likely in industries when large technology gaps and high foreign shares coincide.
8. Kokko (1996)	Mexico	1970	Industry level (the Mexican manufacturing industry 1970)	This chapter examines the signs of productivity spillovers	This chapter examines if there are signs of productivity spillovers from competition between local firms and foreign affiliates. The results indicate that spillovers from competition are not determined by foreign presence alone, but rather by the simultaneous interactions between foreign and local firms

9. Kokko, Tansini, and Zejan (1996)	Uruguay	1988-1990	Firm level	No	No spillovers in a general sense. But there is a positive horizontal spillover in the subsample of plants with moderate technology gaps <i>vis-à-vis</i> foreign firms
10. Borensztein et al. (1998)	69 LDC	1970-1989	Industry level (cross section)	Yes (positive horizontal spillovers)	Positive horizontal spillovers from FDI only when a sufficient absorptive capability of the advanced technologies is available in the host economy
11. Aitken and Harrison (1999)	Venezuela	1976-1989	Firm level (panel)	No	No evidence supports the existence of technology "spillovers" from foreign firms to domestically owned firms
12. Blomstrom and Sjoholm (1999)	Indonesia	1991	Plant level (Indonesian Central Bureau of Statistics)	Yes (positive horizontal spillovers)	Positive horizontal spillovers (Labor productivity is higher in establishments with foreign equity than in purely domestically owned firms and the latter benefit from spillovers from FDI). But these spillovers were restricted to nonexporting local firms
13. Chuang and Lin (1999)	Taiwan	1991	Firm level (Taiwan's manufacturing census data)	Yes (positive horizontal spillovers)	Positive horizontal spillovers effect on firm's productivity
14. Sjoholm (1999)	Indonesia	1980 and 1991	Plant level	Yes (positive horizontal spillovers)	Positive horizontal spillovers (suggested by the fact that high competition to increase the degree of spillovers from FDI) and high technology gaps give rise to large spillovers

Continued

Table 7 Foreign investment and technology spillovers—Cont'd

	Countries	Years	Data	Technology spillovers?	Details
15. Sjoholm (1999)	Indonesia	1980 and 1991	Plant level	Yes (both positive horizontal and vertical)	Both interindustry and intraindustry spillovers are found, which indicates an evidence of positive horizontal as well as vertical spillovers
16. Djankov and Hoekman (2000)	Czech Republic	1992–1996	Firm level	Yes (positive horizontal spillovers)	Positive horizontal spillovers (foreign investment has the predicted positive impact on TFP growth of recipient firms. FDI appears to have a greater impact on TFP growth than do joint ventures, suggesting that parent firms are transferring more knowledge to affiliates than joint venture firms obtain from their partners)
17. Kathuria (2000)	India	1976–1989	Firm level	No (negative horizontal spillovers)	There is an efficiency gain for foreign-owned in 13 out of 26 sectors. For these 13 sectors, there is a negative horizontal spillover from the presence of foreign firms in the sector, but foreign technical capital stock has a positive impact

18. Liu, Siler, Wang, and Wei (2000)	UK	1991–1995	Industry level	Yes (positive horizontal spillovers)	Positive horizontal spillovers. The study also shows that the extent to which local firms benefit from the introduction of advanced technology depends largely on their own technological capabilities
19. Bosco (2001)	Hungary	1993–1997	Firm level	Insufficient evidence	The evidence for technological spillovers is weak and does not allow clear-cut conclusions. Foreign presence in high-tech industries does seem to have a positive effect on both local and foreign firms
20. Damijan et al. (2001)	Bulgaria, Czech Republic, Estonia, Hungary, Poland, Romania, Slovakia Republic, and Slovenia	1994–1998	Firm level	No	There are no significant positive horizontal spillovers for domestic firms from FDI. (Note: trade is an important alternative source of international R&D spillovers to firms without foreign participation in the Czech Republic, Poland, Romania, and Slovenia)
21. Driffeld (2001)	UK	1989–1992	Industry level (report on the census of production)	No	No evidence of investment or output spillovers occurring as a result of inward investment. However, catching up does occur in foreign-owned sector
22. Girma, Greeway, and Wakelin (2001)	UK	1991–1996	Firm level	No	No aggregate evidence of horizontal (intraindustry) spillovers. Specifically, no wage and productivity spillovers to domestic firms as a result of foreign presence

Continued

Table 7 Foreign investment and technology spillovers—Cont'd

	Countries	Years	Data	Technology spillovers?	Details
23. Kinoshita (2001)	Czech Republic	1995-1998	Firm level	Yes (positive horizontal spillovers)	First, technology spillovers from FDI occur for firms that are more R&D intensive; second, spillovers from foreign joint ventures are insignificant for Czech manufacturing firms; lastly, the extent of technology spillovers is greater in oligopolistic sectors than in nonoligopolistic sectors
24. Kokko, Zejan, and Tansini (2001)	Uruguay	1988	Firm level	Yes (positive horizontal spillovers)	Positive horizontal spillovers from FDI to a subsample of locally owned manufacturing plants with moderate technology gaps <i>vis-à-vis</i> foreign firms
25. Konings (2001)	Bulgaria, Romania, and Poland	1993-1997	Firm level	Yes (negative horizontal spillovers) for Bulgaria and Romania, No for Poland	On average, there are negative horizontal spillovers to domestic firms in Bulgaria and Romania and there is no evidence of spillovers in Poland (a negative competition effect that dominates a positive technology effect)
26. Kugler (2001)	Colombia	1974-1998	Industry level	Yes (positive vertical spillovers)	This study distinguishes intra- and interindustry spillovers. He finds widespread evidence for positive vertical (interindustry) spillovers, while horizontal spillovers appear only to be important in machinery equipment sector

27. Li, Liu, and Parker (2001)	China	1995 (1995 industrial census)	Industry level	Yes (positive horizontal spillovers)	Positive horizontal spillovers for SOEs. Collective-and-private-owned enterprises benefit from demonstration and contagion effects from foreign presence
28. Barrios and Strobl (2002)	Spain	1990-1998	Firm level	Yes (positive horizontal spillovers)	Positive horizontal spillovers only when domestic firms had the appropriate “absorptive capacity.”
29. Castellani and Zanfei (2002)	France, Italy, Spain	1992-1997	Firm level	Yes (positive horizontal spillovers)	Positive and significant horizontal productivity spillovers for Italian firms, negative spillovers on Spanish firms and nonsignificant effects on French firms
30. Carkovic and Levine (2002)	72 countries	1960-1995	Industry level (cross section)	No	No evidence of spillovers
31. Dimelis and Louri (2002)	Greece	1997	Firm level	Yes (positive horizontal spillovers)	Positive horizontal spillovers. Moreover, the spillovers stemming from minority-owned foreign establishments are larger than those from majority-owned ones
32. Haskel, Pereira, and Slaughter (2007)	UK	1973-1992	Plant level	Yes (positive horizontal spillovers)	Positive horizontal spillovers from inward FDI to domestic plants
33. Girma (2002)	UK	1989-1999	Firm level	No	No evidence of productivity spillovers

Continued

Table 7 Foreign investment and technology spillovers—Cont'd

Countries	Years	Data	Technology spillovers?	Details	
34. Girma and Gorg (2002)	UK	1980–1992	Firm level	This chapter focuses on the role of absorptive capacity in benefiting from spillovers from FDI	
35. Girma and Wakelin (2002)	UK	1988–1996	Firm level (OneSource Database)	Yes (positive horizontal spillovers)	Positive horizontal spillovers occur in the domestic firms (with the presence of multinational firms in the same sector and region)
36. Grog and Strobl (2002)	Ghana	1991–1997	Firm level	Yes (positive horizontal spillovers)	Positive horizontal spillovers from labor mobility
37. Gorg and Strobl (2002)	Ireland	1973–1996	Plant level	Yes (positive horizontal spillovers)	Positive horizontal spillovers for indigenous plants in high-tech industries

38. Ruane and Ugur (2002)	Ireland	1991-1998	Plant level	No	No evidence of productivity spillovers (when the measurement of foreign presence is the MNC employment as a percentage of total employment). However, a weak evidence of productivity spillovers is found when foreign presence measures as the level of employment in foreign companies in the relevant sector
39. Smarzynska (2002)	Lithuania	1996-2000	Firm level	Yes (positive vertical spillovers)	Positive vertical spillovers (the study shows that the existence of productivity spillovers from FDI taking place through contacts between foreign affiliates and their local suppliers in upstream sectors but there is no indication of spillovers occurring within the same industry)
40. Zukowska- Gagelmann (2002)	Poland	1993-1997	Firm level	No (negative horizontal spillovers)	Negative horizontal spillovers for the most productive local firms in high competition industries and positive horizontal spillovers for the least productive local firms in low competition industries
41. Blalock and Gertler (2003)	Indonesia	1988-1996	Firm level (panel)	Yes (positive vertical spillovers)	Strongly positive vertical spillovers from FDI

Continued

Table 7 Foreign investment and technology spillovers—Cont'd

	Countries	Years	Data	Technology spillovers?	Details
42. Javorcik (2003)	Lithuania	1996–2000	Firm level	Yes (positive vertical spillovers)	Positive vertical productivity spillovers. Moreover, the results indicate that spillovers are associated with projects with shared domestic and foreign ownership but not with fully owned foreign investment
43. Keller and Yeaple (2003)	USA	1987–1996	Firm level	Yes (positive horizontal spillovers)	Positive horizontal spillovers from FDI (accounts for 14% of productivity growth of US firm, FDI leads to significant productivity gains for domestic firms), also evidence of imports- related spillovers
44. Liu (2008)	China	1995–1999	Firm level	Yes (no horizontal but positive vertical spillovers)	No positive horizontal spillovers (negative short-run effects) but strong positive backward linkages
45. Hu and Jefferson (2002)	China	1995–1999	Firm level	No (negative horizontal spillovers)	Sample restricted to China's electronic and textile industries, drawn from the survey of large and medium enterprises conducted by the Chinese National Statistical Bureau. The authors find negative horizontal spillovers, although the impact is positive for firms receiving foreign investment

46. Liu et al. (2008)	China	1998-2005	Firm level	Yes (positive forward spillovers, net zero horizontal spillovers)	They find that Hong Kong, Macao, and Taiwan (HMT) invested firms generate negative horizontal spillovers, while Non-HMT foreign firms tend to bring positive horizontal spillovers. These two opposing horizontal effects seem to cancel out at the aggregate level. They also find forward spillover effects from all types of FDI
47. Du, Harrison and Jefferson (2008)	China	1998-2005	Firm level	Yes large backward spillovers and zero horizontal spillovers	They find significant firm-level effects of Non-HMT investment

End Notes

- *. We thank Dani Rodrik for encouragement and very useful comments and suggestions. We are also grateful to Gordon Hanson and participants at a conference where this and other chapters of the handbook were presented. We would also like to thank Luosha Du, Shanthi Nataraj, and Alexander Tarasov for excellent research assistance. All errors are our own.
- 1. Multiple Pareto-ranked equilibria can also arise in a closed economy, although the behavior of prices makes this less likely. This is because as one sector is expanding, the relative price moves against this sector, and this may rule out multiplicity. See Murphy, Shleifer, and Vishny (1989).
- 2. There are other rationales that have been discussed for industrial policy, including financial market imperfections. We ignore these arguments in this chapter (see Baldwin (1969) for an early criticism, and Pack and Saggi (2006) for a recent survey).
- 3. This section follows Rodríguez-Clare (2007). Models of multiple equilibria in a small open economy include Ciccone and Matsuyama (1996), Okuno-Fujiwara (1988), and Rodríguez-Clare (1996), Rodrik (1996).
- 4. In the traditional model, \bar{L} , is infinite, so labor productivity is simply $\lambda_2(1 + \alpha L_2)$. The alternative assumption that these aggregate externalities are bounded is not only more realistic, but also leads to a simpler analysis. Moreover, this assumption allows us to focus on the issue of latent comparative advantage, as opposed to advantages arising from differences in size or scale. For an analysis where scale (but not infant-industry protection) takes center stage, see Ethier (1982), which formalizes the discussion relating to Frank Graham's argument for protection (Graham, 1923). Scale effects could be captured in the model presented here by assuming that \bar{L} is large; in this case small countries could not exhaust the Marshallian externalities even if they specialized completely in industry 2.
- 5. The result of multiple Pareto-ranked equilibria can be converted into one of development traps by introducing proper dynamics in the model (see Krugman 1991; Matsuyama, 1991). Under some conditions, the economy may be specialized in the good in which it does not have latent comparative advantage, and there may be no equilibrium taking it to specialization in the other good. Government intervention in this case would require more than simple coordination to select the good equilibrium.
- 6. If Marshallian externalities take time to be realized, then one could talk about a *dynamic comparative advantage*. See Redding (1999).
- 7. We assume here that the South is sufficiently large that under autarky it would have $L_2 \geq \bar{L}$.
- 8. One problem with this idea as a way to think about income differences across countries is that it would imply that poor countries have a lower physical capital share, which is not consistent with the data (see Gollin, 2002).
- 9. Formally, we would say that a country has a latent or dynamic comparative advantage in a good if its opportunity cost *given the realization of all static and dynamic Marshallian externalities* is lower than the international price.
- 10. Mill (1848, reference 1909) is generally credited for being the first to express this idea in a clear and simple way, although it was List, Matile, Richelot, & Colwell (1856) who vigorously argued for the adoption of infant-industry protection of manufacturing in European countries. See Corden (1997) for a discussion of the different arguments for and against infant-industry protection, and Irwin (1996) for an excellent treatment of its intellectual history.
- 11. Another way to think about rents is if South were no longer a small economy. Consider Figure 1 again and imagine that there are no productivity differentials between South and North. Imagine further that these two regions have equal size, and that demand for good 2 is sufficiently high that the equilibrium entails one country fully specialized in good 2 and the other fully specialized in good

1. Then, if condition (1) is satisfied, the country that specializes in good 2 is better off than the country that specializes in good 1.
12. Yet another way to have rents would be through the existence of pure profits, as in the literature on “strategic trade policy” (Eaton & Grossman, 1989; Spencer & Brander, 1983).
13. We have assumed here that there is no wage premium in South. If there were a wage premium in South of equal magnitude as in North, then in the absence of a latent comparative advantage in good 2, there would not be multiple equilibria, but it could be advisable for South to implement a policy to specialize in sector 2. This would no longer be a case of Marshallian externalities and infant-industry protection, but rather a standard application of the theory of domestic distortions and trade policy.
14. One difference with the case analyzed above, where rents arise from lack of FPE, is that with wage premia we no longer have the restriction that $R < \theta$. The reason is that now the opportunity cost of good 2 given that region is specialized in good 1 is $R\lambda_{1N}/\lambda_{2N}$, and we need this to be higher than p^* , which is possible even if $\lambda_{1N}/\lambda_{2N} < p^*$. Note that in this case there is no equilibrium with specialization in good 1 in South.
15. Perhaps the most important type of inter-industry externalities is generated by industries that supply specialized inputs that are used intensively by firms in many other sectors. According to Wade (1990), this kind of reasoning was quite important in Taiwan, where the government promoted several sectors that were deemed to provide critical inputs for many other industries. See Noland and Pack (2003) disagree with this view.
16. The condition for specialization in good 1 to be an equilibrium is $(\lambda_{2S}/\lambda_{1S})/(\lambda_{2N}/\lambda_{1N}) < \theta_2/\theta_1$, which is clearly satisfied given $(\lambda_{2S}/\lambda_{1S})/(\lambda_{2N}/\lambda_{1N}) < 1$ together with $1 < \theta_2/\theta_1$. On the other hand, specialization in good 2 implies that $w/\theta_2\lambda_{2S} = p_2^*$. This is an equilibrium if $w/\theta_1\lambda_{1S} > p_1^*$, or $\lambda_{2S}/\lambda_{1S} > \lambda_{2N}/\lambda_{1N}$, which cannot be satisfied if South has CA in good 1.
17. This may explain the existence of cases of geographic concentration of sectors that failed to experience significant agglomeration economies (e.g., concentrations of footwear and textile producers). Perhaps these are cases of clusters that failed to achieve Marshallian externalities (see Altenburg & Meyer-Stamer, 1999).
18. An appropriate policy could be to subsidize production but only to the extent that it is done using modern technologies.
19. Another example, also from Ecuador, concerns the development of new exports of broccoli and mangoes, where finding the best seeds and meeting international phytosanitary standards presented producers with significant coordination problems. As stated by Hernández et al. (2007), collective action fostered and implemented by several private, public, and mixed agencies was important in dealing with such problems and in facilitating the development of these new sectors. Similar cases are documented for Chile in Agosín and Bravo-Ortega (2007) and for several countries in Chandra and Kolavalli (2006).
20. Formally, if $s_m > v_m$ for all m then there is an equilibrium with $w_j = w$ for all j . To see this, simply assume that $w_j = 1$ all j , and $p_m = 1/x_m$, and $L_{jm} > 0$ only if $k_{jm} = 1$. We only need to check that $(1/x_m)\sum_j x_m L_{jm} = v_m$ with $\sum_j L_{jm} = v_m$ holds for all m . One can choose L_{jm} in such a way that this holds as long as $s_m \geq v_m$ all m .
21. The equilibrium could have goods in industry 1 produced by both country 1 and other countries, in which case it is obvious that $w_1 > 1$. Otherwise, if country 1 is the only country with positive production in industry 1, then the equilibrium condition $p_1 L_1 x_1 = v_1 (1 - L_1 + w_1 L_1)$ together with $p_1 x_1 = w_1$ and $s_1 = L_1$ imply that $w_1 = v_1 (1 - s_1)/s_1(1 - v_1)$, hence $v_1 > s_1$ implies $w_1 > 1$. If $s_1 = L_1 > v_1$ then $w_1 > 1$ cannot be sustained, because country 1 cannot specialize completely in the good in which it has a superior productivity.
22. To simplify the discussion, we assume that $T_j = L_j$ for all j . Otherwise, countries with a higher ratio T_j/L_j would tend to have higher wages for reasons that do not relate to IP (although see next section).

23. If the costs of collective action differ across industries, then this clearly would have to be taken into account in this choice. To simplify the analysis, we assume that the cost of achieving collective action is included in the x_m .
24. One could imagine a situation in which countries understand this model and implement IP to maximize their income. If collective action had no cost, then obviously all countries would simply achieve HP in all industries. But imagine that collective action is costly. In the resulting equilibrium sectors with higher complexity and/or larger demand would be matched by high prevalence in such a way that the return to collective action would be the same in all sectors.
25. This is not consistent with what happens for rich countries, where higher productivity goes together with falling diversification. One way to think about this is that at some point a third sector emerges, a high-tech sector, which is a simple CRS sector where research implies increasing productivity. When T and productivity in high-tech is sufficiently high, the agriculture sector disappears, and increasing productivity in high tech leads now to falling exports of manufactures and increasing concentration.
26. A key ingredient in generating this result is that country A 's opportunity cost of good 2 falls below that of country B , a result of A 's latent comparative advantage in good 2.
27. It is important to note that this is not infant-industry protection, however, since the protected industry is not presumed to become competitive at some future period.
28. Beason and Weinstein (1996) perform this exercise for Japan while Lee (1996) does it for Korea. See Noland and Pack (2003) for a comprehensive review of this literature.
29. Although we have focused on trade policies, the use of local-content requirements on foreign producers can also be justified by appealing to Marshallian externalities. In fact, such requirements force multinationals to buy local inputs just as protection induces domestic consumers (and firms) to buy from local producers. A major difference, of course, is that local-content requirements are quantitative restrictions rather than explicit tariffs or subsidies.
30. One interesting exception in Poole (2008), who uses a matched establishment-worker database from Brazil to present evidence consistent with the existence of positive multinational spillovers through worker mobility in Brazil.
31. This is in contrast to most of the literature on multinationals and linkages, which focused instead on the share of inputs that multinationals source domestically.
32. The positive productivity effect relies on the presence of love of variety for inputs, as in Ethier (1982) and Romer (1990).

References

- Abdunasser, H. I. (2002). Export performance and economic growth nexus in Japan: A bootstrap approach. *Japan and the World Economy*, 14(1), 25–33.
- Abhayaratne, A. S. P. (1996). Foreign trade and economic growth: Evidence from Sri Lanka. *Applied Economic Letters*, 13, 567–570.
- Acemoglu, D., Antràs, P., & Helpman, E. (2007). Contracts and technology adoption. *American Economic Review*, 97(3), 916–943.
- Afxentiou, P. C., & Serletis, A. (1991). Exports and GNP causality in the industrial countries: 1950–1985. *Kyklos*, 44(2), 167.
- Agosín, M., & Bravo-Ortega, C. (2007). *The emergence of new successful export activities in Latin America: The case of Chile*. Chile: Department of Economics, Universidad de Chile.
- Ahmad, J., & Harnhirun, S. (1995). Unit roots and cointegration in estimating causality between exports and economic growth: Empirical evidence from the ASEAN countries. *Economics Letters*, 49, 329–334.
- Ahmad, J., & Kwan, A. C. C. (1991). Causality between exports and economic growth: Empirical evidence from Africa. *Economics Letters*, 37, 397–416.
- Ahmed, N. (2003). Trade liberalization and endogenous growth of manufacturing industries in Bangladesh: An empirical investigation. *Applied Economics*, 35(3), 305–314.

- Aisbett, E., Harrison, A., & Zwane, A. (2005). *Globlization and poverty: What is the evidence?* University of Florida: Gainesville, Florida. Prepared for a conference in Honor of Jagdish Bhagwati's 70th Birthday.
- Aitken, B., Hanson, G. H., & Harrison, A. E. (1997). Spillovers, foreign investment, and export behavior. *Journal of International Economics*, 43(1–2) (8), 103–132.
- Aitken, B. J., & Harrison, A. E. (1999). Do domestic firms benefit from direct foreign investment? *Evidence from Venezuela*. *American Economic Review*, 89, 605–618.
- Aitken, B., Harrison, A., & Lipsey, R. E. (1996). Wages and foreign ownership A comparative study of mexico, venezuela, and the united states. *Journal of International Economics*, 40(3–4) (5), 345–371.
- Alam, M. S. (1991). Trade orientation and macro economic preformance in LDCs: An empirical studies. *Economic Development and Cultural Change*, 39, 839–848.
- Alcalá, F., & Ciccone, A. (2004). Trade and productivity. *Quarterly Journal of Economics*, 119(2), 613–646.
- Alege, S. P. (1993). Export and growth in the Nigerian economy: Causality test. *Indian Journal of Economics*, 73, 397–416.
- Alfaro, L., Chanda, A., Kalemli-Ozcan, S., & Sayek, S. (2004). FDI and economic growth: The role of local financial markets. *Journal of International Economics*, 64(1), 89–112.
- Alfaro, L., & Charlton, A. (2008). *Growth and the quality of foreign direct investment: Is all fdi equal?*. CEP Discussion Paper No. 830.
- Alfaro, L., & Rodriguez-Clare, A. (2004). Multinationals and linkages: An empirical investigation. *Economia (Journal of LACEA)*, Spring, 4(2).
- Al-Marhubi, F. (2000). Export diversification and growth: An empirical investigation. *Applied Economics Letters*, 7, 559–562.
- Almeida, R. (2004). *The labor market effects of foreign-owned firms*. World Bank Policy Research Working Paper No. 3300.
- Altenburg, T., & Meyer-Stamer, J. (1999). How to promote clusters. *Policy Experiences from Latin America, World Development*, 27(9).
- Alvarez, F., & Lucas, R. (2005). *General equilibrium analysis of the Eaton-Kortum model of international trade*. NBER Working Paper No. 11764.
- Al-Yousif, Y. K. (1997). Exports and economic growth: Some empirical evidence from the Arab golf countries. *Applied Economics Letters*, 29, 263–267.
- Amirkhalkhali, S., & Dar, A. (1995). A varing coefficients model of export expansion factor accumulation and economic growth: The case of Columbia. *Review of Development Economics*, 208–280.
- Amiti, M., & Konings, J. (2005, July). *Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia* (pp. 1–35). IMF Working Paper.
- Amoateng, K. A., & Amosko-Adu, B. (1996). Economic growth. Export and external debt causality: The Case of African countries. *Applied Economics*, 28, 21–27.
- Amsden, A. H. (1989). *Asia's next giant: South Korea and late industrialization*. Oxford: Oxford University Press.
- An, G., & Lyingun, M. F. (2004). The export skill content, learning by exporting and economic growth. *Economic Letters*, 84(1), 29–34.
- Anderson, K. Y. M. (1996, August). *Social policy dimensions of economic integration: Environmental and labour standards*. NBER Working Paper No. W5702.
- Andrew, A., & Burstein, A. (2007). *Pricing-to-market, trade costs, and international relative prices*. Working Paper Series 2007–26, Federal Reserve Bank of San Francisco.
- Anoruo, E., & Ahmad, Y. (1999). Openness and economic growth: Evidence from selected Asian countries. *Indian Economic Journal*, 47(3), 110–116.
- Antras, P., & Helpman, E. (2007). *Contractual frictions and global sourcing*. CEPR Discussion Paper No. 6033.
- Anwer, M. S., & Sampath, R. K. (2000). Exports and economic growth. *Indian Economic Journal*, 47(3), 79–88.
- Arkolakis, C., Demidova, S., Klenow, P., & Rodríguez-Clare, A. (2008). *Endogenous variety and the gains from trade American economic review papers and proceedings*.
- Arnold, J. M., Javorcik, B. S., & Mattoo, A. (2006, September 21). *Does services liberalization benefit manufacturing firms? evidence from the czech republic*. CEPR Discussion Paper No. 5902.

- Atesoglu, H. S. (1994). An application of a kaldorian export-led model of growth to the United States. *Applied Economics*, 26, 479–483.
- Atkeson, A., & Burstein, A. (2007a). Pricing-to-market in a ricardian model for international trade. *American Economic Review*, 97(2), 362–367.
- Atkeson, A., & Burstein, A. (2007b). *Innovation, firm dynamics, and international trade*. NBER Working Papers 13326, National Bureau of Economic Research, Inc.
- Atkeson, A., & Burstein, A. (2007c). *Pricing-to-market, trade costs, and international relative prices*. Working Paper Series 2007–26, Federal Reserve Bank of San Francisco.
- Aw, B. Y., Chung, S., & Roberts, M. J. (2000). Productivity and turnover in the export market: Micro-level evidence from the republic of Korea and Taiwan (China). *The World Bank Economic Review*, 14(1), 65.
- Aw, B. Y., Roberts, M. J., & Winston, T. (2007). Export market participation, investments in R D and worker training, and the evolution of firm productivity. *The World Economy*, 30(1), 83.
- Aw, B. Y., Roberts, M., & Xu, D. (2009). *R&D investment, exporting, and productivity dynamics*. Working Paper, Pennsylvania State University.
- Awokuse, T. O. (2005a). Exports, economic growth and causality in Korea. *Applied Economic Letters*, 12 (11), 693–696.
- Awokuse, T. O. (2005b). Export-led growth and the Japanese economy: Evidence from VAR and directed acyclic graphs. *Applied Economic Letters*, 12(14), 849–858.
- Bagwell, K., & Staiger, R. W. (1998). Will preferential agreements undermine the multilateral trading system? *The Economic Journal*, 108(449), 1162.
- Bahmani-Oskooee, M., Mohtadi, H., & Shabsigh, G. (1991). Exports, growth and causality in LDCs: A re-examination. *Journal of Development Economics*, 36(2), 405–415.
- Bailey, M. N., Hulten, C., & Campbell, D. (1992). The distribution of productivity. *Brookings Papers on Economic Activity*, 1, 187.
- Balaguer, J., & Cantavella-Jorda, M. (2004). Structural change in exports and economic growth: Cointegration and causality analysis for Spain (1961–2000). *Applied Economics*, 36(5), 473–477.
- Balassa, B. (1985). Exports, Policy choices, and economic growth in developing countries after the 1973 oil shock. *Journal of Development Economics*, 18, 23–35.
- Balasubramanyam, V. N., Salisu, M., & Sapsford, D. (1996). Foreign direct investment and growth in EP and IS countries. *Economic Journal, Royal Economic Society*, 106(434), 92–105.
- Baldwin, R. E. (1969). The case against infant industry protection. *Journal of Political Economy*, 77(May/June), 295–305.
- Baldwin, R., & Krugman, P. R. (1986, September). *Persistent trade effects of large exchange rate shocks*. NBER Working Paper Series 2017, National Bureau of Economic Research.
- Baldwin, R., & Krugman, P. (1989, November). Persistent trade effects of large exchange rate shocks. *The Quarterly Journal of Economic*, 104(4), 635–654.
- Bardhan, P. (1970). *Economic growth, development, and foreign trade: A study in pure theory*. New York: Wiley Interscience.
- Bardhan, P. (1971). On optimum subsidy to a learning industry: An aspect of the theory of infant-industry protection. *International Economic Review*, 12, 54–70.
- Barrios, S., & Strobl, E. (2002). Foreign direct investment and productivity spillovers: Evidence from the spanish experience. *Weltwirtschaftliches Archiv*, 138, 459–481.
- Bartel, A. P., & Harrison, A. E. (2000). *Ownership versus environment: Disentangling the sources of public sector inefficiency*. World Bank Policy Research Working Paper No. 2272.
- Bartel, A. P., & Harrison, A. E. (2005). Ownership versus environment: Disentangling the sources of public-sector inefficiency. *Review of Economics and Statistics*, 87(1)(02/01), 135–147.
- Beason, R., & Weinstein, D. E. (1996). Growth, economies of scale, and targeting in Japan (1955–1990). *The Review of Economics and Statistics*, 78(2), 286–295.
- Begum, S., & Shamsuddin, A. F. M. (1998). Exports and economic growth in Bangladesh. *Journal of Development Studies*, 35(1), 89–114.
- Berg, H. (1997). The relationship between international trade and economic growth in Mexico. *The North American Journal of Economics and Finance*, 8(1), 1–21.

- Bernard, A. B., & Jensen, J. B. (1999). Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics*, 47(1)(2/1), 1–25.
- Bernard, A. B., Redding, S. J., & Schott, P. K. (2006). *Multi-product firms and trade liberalization*. NBER Working Papers 12782, National Bureau of Economic Research.
- Bernard, A. B., & Wagner, J. (1997). Exports and success in german manufacturing. *Weltwirtschaftliches Archiv*, 133(1), 134.
- Bhagwati, J. (2004). *In defense of globalization*. New York: Oxford University Press.
- Bhagwati, J., & Ramaswami, V. K. (1963). Domestic distortions, tariffs and the theory of optimum subsidy. *The Journal of Political Economy*, 71(1), 44.
- Bigsten, A., Collier, P., Dercon, S., Fachamps, M., Gauthier, B., Gunning, J. W., et al. (2004). Do African manufacturing firms learn from exporting? *The Journal of Development Studies*, 40(3), 115–141.
- Biswal, B., & Dhawan, U. (1998). Export-led growth hypothesis: Cointegration and causality analysis for Taiwan. *Applied Economics Letters*, 5, 699–701.
- Blalock, G., & Gertler, P. (2003). *Technology from foreign direct investment and welfare gains through the supply chain*. Mimeo, Cornell University.
- Blalock, G., & Gertler, P. J. (2004). Learning from exporting revisited in a less developed setting. *Journal of Development Economics*, 75(2)(12), 397–416.
- Blomström, M. (1986). Multinationals and market structure in Mexico. *World Development*, 14(4) (4), 523–530.
- Blomstrom, M., & Persson, H. (1983). Foreign investment and spillovers efficiency in an underdeveloped economy: Evidence from the Mexican manufacturing industry. *World Development*, 11, 493–501.
- Blomstrom, M., & Sjoholm, F. (1999). Multinational corporations and productive convergence in Mexico. In W. J. Baumol, R. R. Nelson & E. N. Wolff (Eds.), *Convergence of productivity: Cross national studies and historical evidence* (pp. 263–283). Oxford: Oxford University Press.
- Blomstrom, M., & Wolff, E. N. (1994). *Multinational corporations and productivity convergence in mexico*. NBER Working Papers 3141, National Bureau of Economic Research, Inc.
- Bodman, P. M. (1996). On export-led growth in Australia and Canada: Cointegration. *Causality and Structural Stability, Australian Economic Paper*, 35, 282–299.
- Bolaky, B., & Freund C. (2004). *Trade, regulations, and growth*, The World Bank. Policy Research Working Paper Series: 3255.
- Boltho, A. (1996). Was Japaness growth export-led? *Oxford Economic Papers*, 48, 415–432.
- Borensztein, E., De Gregorio, J., & Lee, J-W. (1998). How does foreign direct investment affect economic growth? *Journal of International Economics*, 45, 115–135.
- Borrus, M., Tyson, L., & Zysman, J. (1986). Creating advantage: How government policies shape high technology trade. In P. Krugman (Ed.), *Strategic trade policy and the new international economics*. New York: MIT Press.
- Borrus, M., Tyson, L., & Zysman, J. (1996). Creating advantage: How government policies shape international trade in the semiconductor industry. In P. R. Krugman (Ed.), *Strategic trade policy and the new international economics*. New York: MIT Press.
- Bosco, M. G. (2001). Does FDI contribute to technological spillovers and growth? A panel data analysis of Hungarian firms. *Transitional Corporations*, 10, 43–68.
- Bosworth, B., & Collins, S. M. (1999). Capital flows to developing countries: Implications for saving and investment. *Brookings Papers on Economics*, 1, 143–180.
- Brander, J. (1995). Strategic trade policy. In G. Grossman & K. Rogoff (Eds.), *Handbook of international economics* (Vol. 3). Amsterdam: North-Holland.
- Broda, C., Greenfield, J., & Weinstein, D. (2006). *From groundnuts to globalization: A structural estimate of trade and growth*. NBER Working Papers 12512. National Bureau of Economic Research, Inc.
- Broda, C. M., Limão, N., & Weinstein, D. E. (2006, February). *Optimal tariffs: The evidence* (Vol. w12033). NBER Working Paper Series.
- Broda, C., & Weinstein, D. E. (2006). Globalization and the gains from variety. *Quarterly Journal of Economics*, 121(2) (05/01), 541–585.
- Burney, N. A. (1996). Exports and economic growth: Evidence from cross-country analysis. *Applied Economic Letters*, 3, 369–373.

- CAF. (2007). *Reporte economía y desarrollo 2006: Un Camino Hacia la Transformación Productiva*. Caracas, Venezuela: Corporación Andina de Fomento.
- Carkovic, M., & Levine, R. (2002). Does foreign direct investment accelerate economic growth? Working Paper, University of Minnesota.
- Castellani, D., & Zanfei, A. (2002a). *Technology gaps, absorptive capacity and the impact of inward investments on productivity of European firms*. Mimeo, University of Urbino (Italy).
- Caves, R. E. (1996). *Multinational enterprise and economic analysis* (2nd ed.). Cambridge: Cambridge University Press.
- Chandra, V., & Kolavalli, S. (2006). Technology, adaptation, and exports: How some countries got it right, chapter 1. In V. Chandra (Ed.), *Technology, adaptation and exports*. Washington, DC: World Bank Publications.
- Chang, H.-J. (2002). Kicking away the ladder: Development strategy in historical perspective. Anthem Press, 187 Pages, London.
- Chang, R., Kaltani, L., & Loayza, N. (2005). *Openness can be good for growth: The role of policy complementarities*. NBER Working Papers: 11787, National Bureau of Economic Research, Inc.
- Cheng, B. S., & Chu, O. (1996). U.S. exports and economic growth causality. *Atlantic Economic Journal*, 34, 113–115.
- Choudhri, E. U., & Hakura, D. S. (2000). International trade and productivity growth: Exploring the sectoral effects on developing countries. *IMF Staff Papers*, 47(1), 30–53.
- Chow, P. O. Y. (1987). Causality between export growth and industrial development. *Journal of Development Studies*, 26, 55–63.
- Chuang, Y. C. (2002). The trade-induced learning effect on growth: Cross-country evidence. *Journal of Development Studies*, 39(2), 137–154.
- Chuang, Y. C., & Hsu, P. F. (2004). FDI, trade, and spillover efficiency: Evidence from China's manufacturing sector. *Applied Economics*, 36(10), 1103.
- Chuang, Y.-C., & Lin, C.-M. (1999). Foreign direct investment. R&D and spillover efficiency: Evidence from Taiwan's manufacturing firms. *Journal of Development Studies*, 35, 117–137.
- Ciccone, A., & Matsuyama, K. (1996). Start-up costs and pecuniary externalities as barriers to economic development. *Journal of Development Economics*, 49(1).
- Clark, D. P. (1995). Trade orientation and industrial sector growth in developing countries. *Journal of Developing Areas*, 30, 1–10.
- Clark, D. P., Sawyer, W. C., & Sprukle, L. (1999). Openness and Industrialization in developing countries. *Applied Economics Letters*, 6, 161–164.
- Clemens, M. A., & Williamson, J. G. (2001, September). *A tariff-growth paradox? Protection's impact the world around, 1875–1997*. NBER Working Paper No. 8459.
- Clerides, S. K., Lach, S., & Tybout, J. R. (1998). Is learning by exporting important? Micro-dynamics evidence from Columbia, Mexico, and Morocco. *The Quarterly Journal of Economics*, 113(3), 903–947.
- Coe, D. T., & Helpman, E. (1995). International R&D spillovers. *European Economic Review*, 39(5), 859–896.
- Coe, D. T., Helpman, E., & Hoffmaister, W. A. (1997). North-south R&D spillovers. *Economic Journal*, 107, 134–149.
- Connolly, M. (2003). The dual nature of trade: Measuring its impact on imitation and growth. *Journal of Development Economics*, 72(1), 31–55.
- Costantini, J., & Melitz, M. (2007). The dynamics of firm level adjustment to trade liberalization. CEPR.
- Coppin, A. (1994). Determinants of LDC output growth during the 1980s. *Journal of Developing Areas*, 219–228.
- Corden, M. (1997). *Trade policy and economic welfare* (2nd ed.). Oxford: Oxford University Press.
- Damijan, L. P., Majcen, B., Knell, M., & Rojec, M. (2001). *the role of FDI. Absorptive capacity and trade in transferring technology to transition countries: Evidence from firm panel data for eight transition countries*. Mimeo, UN Economic Commission for Europe, Geneva.
- Dar, A., & Amirkhalkali, S. (2003). On the impact of trade openness on growth: Further evidence from OECD countries. *Applied Economics*, 35(16), 1731–1766.
- Darrat, A. F. (1987). Are exports an engine of growth? Another look at the evidence. *Applied Economics*, 19, 277–283.

- Darrat, A. F., Hsu, M. K., & Zhong, M. (2000). Testing export exogeneity in Taiwan: Further evidence. *Applied Economics Letters*, 7(9), 563–568.
- Dawson, P. J., & Hubbard, L. D. (2004). Exports and economic growth in Central and East European countries during transition. *Applied Economics*, 36(16), 1819–1824.
- De Ferranti, D., Perry, G., Lederman, D., & Maloney, W. (2001). *From natural resources to the knowledge economy*. Washington, DC: World Bank.
- DeJong, D. N., & Ripoll, M. (2006). Tariffs and growth: An empirical exploration of contingent relationships. *Review of Economics and Statistics*, 88(4) (11/01), 625–640.
- Delgado, M. A., Fariñas, J. C., & Ruano, S. (2002). Firm productivity and export markets: A non-parametric approach. *Journal of International Economics*, 57(2) (8), 397–422.
- De Loecker, J. (2007). Do exports generate higher productivity? Evidence from slovenia. *Journal of International Economics*, 73(1) (9), 69–98.
- Demidova, S., & Rodriguez-Clare, A. (2008). *Trade policy under firm-level heterogeneity in a small economy*. NBER Working Paper No. 13688.
- De Pineres, S. A., & Ferrantino, G. (1999). Export sector dynamics and domestic growth: The case of Colombia. *Review of Development Economics*, 3(3), 268–280.
- Dhawan, U., & Biswal, B. (1999). Re-examining export-led growth hypothesis: A Multivariate cointegration analysis for India. *Applied Economics*, 33, 525–530.
- Diamond, P. A., & Mirrlees, J. A. (1971). Optimal taxation and public production I: Production efficiency. *The American Economic Review*, 61(1), 8.
- Dimelis, S., & Louri, H. (2002). Foreign direct investment and effiency benefits: A conditional quantile analysis. *Oxford Economic Papers*, 54, 449–469.
- Dixit, A. (1985). Tax policy in open economies. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of public economics* (1st ed. Vol. 1, pp. 313–374). Amsterdam: Elsevier.
- Djankov, S., & Hoekman, B. (2000). Foreign investment and productivity growth in czech enterprises. *The World Bank Economic Review*, 14(1), 49.
- Dodaro, S. (1991). Comparative advantage, trade and growth: Exported growth revisited. *World Development*, 19(9), 1153–1165.
- Dodaro, S. (1993). Exports and growth: A reconsideration of causality. *Journal of Developing Areas*, 27, 245–278.
- Dodzin, S., & Vamvakidis, A. (2004). Trade and industrialization in developing economies. *Journal of Development Economics*, 75(1), 319–328.
- Dollar, D. (1992). Ourward-oriented developing economics really do grow more rapidly: Evidence from 95 LDCs, 1976–85. *Economic Development and Cultural Change*, 40(3), 523–544.
- Dollar, D., & Kraay, A. (2004). Trade, growth, and poverty. *The Economic Journal*, 114(493), F22.
- Donaldson, D. (2009). *Railroads of the Raj: Estimating the impact of transportation infrastructure*. Working Paper. London School of Economics.
- Doraisami, A. (1996). Export growth and economic growth: A re-examination of some time series evidence of the malaysian experience. *Journal of Developing Areas*, 30, 223–230.
- Doyle, E. (1998). Export-output causality: The Irish case, 1953–1993. *Atlantic economic Journal*, 26, 147–161.
- Driffield, N. (2001). The impact on domestic productivity of inward investment in the UK. *The Manchester School*, 69, 103–119.
- Driffield, N., & Girma, S. (2003). Regional foreign direct investment and wage spillovers: Plant level evidence from the UK electronics industry. *Oxford Bulletin of Economics and Statistics*, 65(4), 453.
- Dritsakis, N. (2004). Exports, investment and economic development of pre-accession countries of the EU: An empirical investigation of Bulgaria and romania. *Applied Economics*, 36(16), 1831–1838.
- Du, L., Harrison, A., & Jefferson, G. (2008). Testing for horizontal and vertical foreign investment spillovers in China, Working Paper, University of California Berkeley. 1998–2005.
- Duankov, S., & Heokman, B. (2002). Foreign investment and productivity growth in Czech enterprises. *World Bank Economic Review*, 14, 49–64.
- Dutt, S. D., & Ghosh, D. (1994). Trade libralization and industrial growth in Pakistan: A cointegration analysis. *Applied Economics*, 36(13), 1421–1429.

- Dutt, S. D., & Ghosh, D. (1996). An empirical investigation of the export growth-Economic growth relationship. *Applied Economics Letters*, 1, 44–48.
- Dutta, D., & Ahmed, M. (2004). An aggregate import demand function for India: A cointegration analysis. *Applied Economics Letters*, 11(10), 607.
- Earle, J. S., & Telegdy, A. (2007, March). *Ownership and wages: Estimating public-private and foreign-domestic differentials using leed from hungary, 1986–2003*. NBER Working Paper No. W12997.
- Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: How endowments influence economic development. *Journal of Monetary Economics*, 50(1), 3–39.
- Eaton, J., & Grossman, G. (May 1986). Optimal trade and industrial policy under oligopoly. *The Quarterly Journal of Economics*, 101(2), 383–406.
- Eaton, J., & Kortum, S. (1999). International technology diffusion: Theory and measurement. *International Economic Review*, 40(3), 537.
- Eaton, J., & Kortum, S. (2001). Technology, Trade and growth: A unified framework. *European Economic Review*, 45, 742–755.
- Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70, 1741–1780.
- Edwards, S. (1998). Openness, productivity and growth: What do we really know? *The Economic Journal*, 108(447), 383.
- Ederington, J., & McCalman, P. (2008). Endogenous firm heterogeneity and the dynamics of trade liberalization. *Journal of International Economics*, 74, 422–440.
- Edwards, S. (1992). Trade orientation, Distortions, and growth in developing countries. *Journal of Development Economics*, 39, 31–57.
- Egwaihede, F. O. (1992). Oil export and economic growth in Nigeria: A preliminary investigation. *Indian Journal of Economics*, 73, 221–232.
- Erlat, G. (2000). Measuring the impact of trade flows on employment in the Turkish manufacturing industry. *Applied Economics*, 32(9), 1169–1180.
- Ernesto, L.-C. J., & Meissner, C. M. (2003). Exchange-rate regimes and international trade: Evidence from the classical gold standard era. *American Economic Review*, 93(1), 344–353.
- Esfahani, F. I. S. (1991). Exports, Imports and economic growth in semi-industrialized countries. *Journal of Development Economics*, 35, 93–116.
- Estevadeordal, A., & Taylor, A. M. (2008). *Is the Washington consensus dead? Growth, openness, and the great liberalization, 1970s–2000s*. Working Paper ed. National Bureau of Economic Research 1050 Massachusetts Avenue Cambridge, MA 02138.
- Ethier, W. J. (1982a). Decreasing costs in international trade and frank graham's argument for protection. *Econometrica*, 50(5), 1243–1268.
- Ethier, W. J. (1982b). National and international returns to scale in the modern theory of international trade. *American Economic Review*, 72(3), 389–405.
- Ethier, W. J. (1982c). The general role of factor intensity in the theorems of international trade. *Economics Letters*, 10(3–4), 337–342.
- Ethier, W. J. (1982d). Dumping. *Journal of Political Economy*, 90(3), 487–506.
- Falvey, R., Foster, N., & Greenaway, D. (2004). Imports, exports, knowledge spillovers and growth. *Economics Letters*, 85(2), 209–213.
- Feder, G. (1982). On exports and economic growth. *Journal of Development Economics*, 12, 59–73.
- Feenstra, R. (1994). *New goods and index numbers: U.S. import prices*. Working Paper ed. NBER Working Papers 3610 National Bureau of Economic Research, Inc.
- Feliciano, Z., & Lipsey, R. E. (1999). *Foreign ownership and wages in the united states, 1987–1992*. NBER Working Paper No. W6923.
- Fernandes, A., & Isgut, A. (2007). *Learning-by-exporting effects: Are they for real?* MPRA Paper 3121. University Library of Munich. Germany.
- Ferreira, P. C., & Rossi, J. L. (2003). New evidence from Brazil on trade liberalization and productivity growth. *International Economic Review*, 44(4), 1383–1405.
- Feyrer, J. (2009). *Trade and income-exploiting time series in geography*. Working Paper. NBER and Dartmouth College.
- Findlay, R. (1995). *Factor proportions, trade and growth*. New York: MIT Press.

- Fosu, A. X. (1990). Exports and economic growth: The African case. *World Development*, 18, 531–535.
- Fosu, A. K. (1996). Primary exports and economic growth in developing countries. *World Economy*, 19, 466–475.
- Francisco, F., & Ramos, R. (2001). Exports, imports and economic growth in Portugal: Evidence from causality and cointegration analysis. *Economic Modelling*, 18(4), 613–623.
- Frankel, J. A., & Romer, D. (1999). Does trade cause growth. *American Economic Review*, 89(3), 379–399.
- Freund, C., & Bolaky, B. (2008, October). Trade, regulations, and income. *Journal of Development Economics*, 87(2), 309–321.
- Fu, X. (2005). Exports technical progress and productivity growth in a transition economy: A non-parametric approach for China. *Applied Economics*, 37(7), 735–739.
- Gani, A. (1997). Economic growth in Papua New Guinea: Some empirical evidence. *Pacific Economic Bulletin*, 12(1), 41–52.
- Gawande, K., Krishna, P., & Olarreaga, M. (2005). *Lobbying competition over trade policy*, Working Paper No. 11371. NBER.
- Gharte, E. E. (1993). Cause relationship between exports and economic growth: Some empirical evidence in Taiwan. *Japan and the US, Applied Economics*, 25, 1141–1152.
- Ghartery, E. (1993). Causal relationship between exports and economic growth: Some empirical evidence in Taiwan, Japan and the US. *Applied Economics*, 25(9), 1145.
- Ghatak, S., Miner, C., & Utkulu, U. (1997). Exports, exports expansion and growth: Cointegration and causalit: Evidence for Malaysia. *Applied Economics*, 29(2), 213–223.
- Girma, T., Grabowski, R., & Sharma, S. C. (2001). Exports, investment, efficiency and economic growth in LDC: An empirical investigation. *Applied Economics*, 33(6), 599–700.
- Giles, D. E. A., Giles, J. A., & McCann, E. (1992). Causality unit roots and export-led growth: The New Zealand experience. *Journal of International Trade and Economic Development*, 1, 195–218.
- Girma, S. (2002). *Geographic proximity, absorptive capacity and productivity spillovers from FDI: A threshold regression analysis*. GEP Research Paper 02/06. University of Nottingham.
- Girma, S., & Gorg, H. (2002). *Foreign direct investment, spillovers and absorptive capacity: Evidence from quantile regressions*. GEP Research Paper 02/14. University of Nottingham.
- Girma, S., Greenaway, D., & Wakelin, K. (2001). Who benefits from foreign direct investment in the UK. *Scottish Journal of Political Economy*, 48, 119–133.
- Girma, S., & Wakelin, K. (2002). Are there regional spillovers from FDI in the UK? In D. Greenaway, R. Upward & K. Wakelin (Eds.), *Trade, investment, migration and labor markets*. Basingstoke: Macmillan.
- Globerman, S. (1979). Foreign direct investment and ‘spillover’ efficiency benefits in Canadian manufacturing industries. *Canadian Journal of Economics*, 12, 42–56.
- Gokceus, O. (1997). Trade liberalization and productivity growth: New evidence from the Turkish rubber industry. *Applied Economics*, 29(5), 639–645.
- Goldberg, P. K., Amit, K., Pavcnik, N., & Topalova, P. (2008). *Imported intermediate inputs and domestic product growth: Evidence from India*. NBER Working Paper Series 14416.
- Goldberg, P., Khandelwal, A., Pavcnik, N., & Topalova, P. B. (2008, June). *Multi-product firms and product turnover in the developing world: Evidence from India*. NBER Working Paper No. W14127.
- Goldberg, P., & Maggi, G. (1999). Protection for sale: An empirical investigation. *American Economic Review*, 89(5), 1135–1155.
- Goldberg, P., & Pavcnik, N. (2007). Distributional effects of globalization in developing countries. *Journal of Economic Literature*, 45(1), 39.
- Gonçalves, R., & Richtering, J. (1987). Intercountry comparison of export performance and output growth. *The Developing Economies*, 25(1).
- Gordon, D. V., & Sakyi, K. (1993). Testing the export-growth hypothesis: Some parametric and non-parametric results for Ghana. *Applied Economics*, 25(4), 553–563.
- Gorg, H., & Greenway, D. (2004). Much ado about nothing? do domestic firms really benefit from foreign direct investment? *The World Bank Research Observer*, 19(2), 171–197.
- Gorg, H., & Strobl, E. (2002a). *Spillovers from foreign firms through worker mobility: An Empirical Investigation*. GEP Research Paper 02/13, University of Nottingham.

- Gorg, H., & Strobl, E. (2002b). Multinational companies and indigenous development: An empirical analysis. *European Economic Review*, 46, 1305–1322.
- Gorodnichenko, Y., Svejnar, J., & Terrell, K. (2008). *Globalization and innovation in emerging markets*. NBER Working Paper No. 14481. National Bureau of Economic Research Inc.
- Grabowski, R., Sharma, S. C., & Dhakal, D. (1990). Exports and Japanese economic Development. *Economics Letters*, 32(2), 127–132.
- Graham, F. D. (1923). Some aspects of protection further considered. *The Quarterly Journal of Economics*, 37 (2), 199.
- Greenaway, D., Morgan, W., & Wright, P. (1997). Trade liberalization and growth in developing countries: Some new evidence. *World Development*, 25(1), 1885–1897.
- Greenaway, D., Morgan, W., & Wright, P. (1999). Exports, export composition and growth. *Journal of International Trade and Economic Development*, 8(1), 41–51.
- Greenaway, D., Morgan, W., & Wright, P. (2002). Trade liberalisation and growth in developing countries. *Journal of Development Economics*, 67(1), 229–244.
- Greenaway, D., & Sapsford, D. (1994). What does liberalisation do for exports and growth? *Review of World Economics (Weltwirtschaftliches Archiv)*, 130(1), 152–174.
- Greenwald, B., & Stiglitz, J. E. (2006). Helping infant economies grow: Foundations of trade policies for developing countries. *The American Economic Review*, 96(2).
- Grossman, G. M. (1992). *Imperfect competition and international trade*. Cambridge MA: MIT Press.
- Grossman, G., & Helpman, E. (1991). *Innovation and growth in the global economy*. Cambridge MA: MIT Press.
- Gupta, S. (1985). Export growth and economic growth revisited. *Indian Economic Journal*, 32, 52–129.
- Haddad, M., & Harrison, A. (1993). Are there positive spillovers from direct foreign investment? Evidence from panel data for Morocco. *Journal of Development Economics*, 42, 51–74.
- Hallward-Driemeier, M., Iarossi, G., & Sokoloff, K. L. (2002). *Exports and manufacturing productivity in east asia: A comparative analysis with firm-level data*. NBER Working Paper Series W8894, National Bureau of Economic Research, Inc.
- Halpern, L., Koren, M., & Szeidl, A. (2005, July). *Imports and productivity*. CEPR Discussion Paper No. 5139.
- Hansen, P. (1994). The government, exporters and economic growth in New Zealand. *New Zealand Economic Papers*, 28(2), 133–142.
- Hansen, J., Jensen, D., & Madsen, E. (2003). The establishment of the Danish windmill industry—Was it worthwhile? *Review of World Economics*, 139, 324–347.
- Hanson, G. H. (2001). U.S.-Mexico integration and regional economies: Evidence from border-city pairs. *Journal of Urban Economics*, 50(2), 259–287.
- Harris, R., & Robinson, C. (2002). *Spillovers from foreign ownership in the United Kingdom: Estimates for UK manufacturing using the ARD*. Mimeo. University Durham.
- Harrison, A. E. (1994a). An empirical test of the infant industry argument: Comment. *The American Economic Review*, 84(4), 1090–1095.
- Harrison, A. E. (1994b). Productivity, imperfect competition, and trade reform. Theory and evidence. *Journal of International Economics*, 36(1–2), 53–73.
- Harrison, A. E. (1996). Openness and growth: A time-series. Cross-country analysis for developing countries. *Journal of Development Economics*, 48, 419–447.
- Harrison, A. (2007). *Globalization and poverty*. University of Chicago Press.
- Harrison, A., & Hanson, G. (1999). Who gains from trade reform? Some remaining puzzles. *Journal of Development Economics*, 59, 125–154.
- Harrison, A. E., Inessa, L., & McMillan, M. S. (2004). Global capital flows and financing constraints. *Journal of Development Economics*, 75(1), 269–301.
- Harrison, A. E., Love, I., & McMillan, M. S. (2004). Global capital flows and financing constraints. *Journal of Development Economics*, 75(1), 269–301.
- Harrison, A. E., & McMillan, M. (2003). Does direct foreign investment affect domestic firm credit constraints? *Journal of International Economics*, 61, 73–100.
- Harrison, A. E., & McMillan, M. (2007). On the links between globalization and poverty. *Journal of Economic Inequality*, 5, 123–134.

- Harrison, A., & Scorse, J. (2004). Globalization's impact on compliance with labor standards. In S. Collins & D. Rodrik (Eds.), *Brookings trade forum 2003*. Washington, DC: Brookings Institution Press.
- Harrison, A. E., & Scorse, J. (2008). *Multinationals and anti-sweatshop activism American Economic Review*. forthcoming.
- Haskel, J. E., Pereira, S. C., & Slaughter, M. J. (2007). Does inward foreign direct investment boost the productivity of domestic firms? *The Review of Economics and Statistics*, 89(3), 482–496.
- Haussmann, R., Hwang, J., & Rodrik, D. (2005). *What you export matters*. NBER Working Papers: 11905. National Bureau of Economic Research, Inc.
- Haussmann, R., & Klinger, B. (2006). *Structural transformation and patterns of comparative advantage in the product space*. Working Paper. University of Harvard.
- Haussmann, R., Rodríguez-Clare, A., & Rodrik, D. (2005). Inter-American development bank, economic and social study series, February: Towards a strategy for economic growth in Uruguay
- Haussmann, R., & Rodrik, D. (2003). Economic development as self-discovery (with Ricardo Hausmann). *Journal of Development Economics*, 72.
- Haussmann, R., & Rodrik, D. (2006). *Doomed to choose: Industrial policy as predicament*. John F. Kennedy School of Government, Harvard University.
- Head, K. (1994). Infant industry protection in the steel rail industry. *Journal of International Economics*, 37, 141–165.
- Helpman, E., & Krugman, P. R. (1985). *Market structure and foreign trade* (p. 26). Cambridge: The MIT Press.
- Henriques, I., & Sadorsky, P. (1996). Export-led growth or growth-driven exports? The Canadian case. *Canadian Journal of Economics*, 29(3), 540–555.
- Hernández, I., Cely, N., González, F., Muñoz, E., & Prieto, I. (2007). *The discovery of new export products in Ecuador*. Ecuador: Stratega.
- Hoekman, B. (2005). Operationalizing the concept of policy space in the WTO: Beyond special and differential treatment. *Journal of International Economic Law*, 8(2), 405–424.
- Holman, J. A., & Graves, P. E. (1995). Korean exports economic growth: An econometric reassessment. *Journal of Economic Development*, 20(2), 45–56.
- Hotchkiss, L., Julie, R. E., & Moore, M. R. (1994). Export expansion and growth at different stages of development. *Journal of Economic Development*, 19, 87–105.
- Hsiao, M.-C. W. (1987). Tests of causality and exogeneity between exports and economic growth: The case of the Asian NICs. *Journal of Economic Development*, 12(2), 143–159.
- Hsieh, C.-T., & Klenow, P. J. (2007). *Misallocation and manufacturing TFP in China and India*. NBER Working Paper No. 13290.
- Hu, A., & Jefferson, G. (2002). FDI impact and spillover: Evidence from China's electronic and textile industries. *The World Economy*, 25, 1063–1076.
- Huchet, J.-F. (1997). The China circle and technological development in the Chinese electronics industry. In N. Barry (Ed.), *The China circle: Economics and electronics in the PRC, Taiwan, and Hong Kong* (pp. 254–285). Washington, DC: Brookings Institution Press.
- Hutchison, M., & Singh, N. (1992). Exports, non-exports and externalities: A granger causality approach. *International Economic Journal*, 6(2), 79–94.
- Imbs, J., & Wacziarg, R. (2003). Stages of diversification. *American Economic Review*, 93(1), 63–86.
- Irwin, D. (1996). *Against the tide: An intellectual history of free trade*. Princeton: Princeton University Press.
- Irwin, D. (2000). Did late nineteenth century U.S. tariffs promote infant industries? Evidence from the tinplate industry. *Journal of Economic History*, 60, 335–360.
- Irwin, D. (2002). Interpreting the tariff-growth correlation in the late nineteenth century. *Papers and proceedings (May): American economic review*.
- Irwin, D. (2007). Tariff incidence in America's gilded age. *Journal of Economic History*, 67, 582–607.
- Iscan, T. (1998). Exports and capital accumulation: Some empirical evidence from the Mexican manufacturing industry. *Applied Economics Letters*, 5(6), 355–360.
- Islam, M. N. (1998). Export expansion and economic growth: Testing for cointegration and causality. *Applied Economics*, 30(3), 415–425.
- Javorcik, B. S. (2004). Does Foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review*, 94(3).

- Jin, J. C. (1995). Export led growth and the four little dragons. *Journal of International Trade and Economics Development*, 4(2), 203–215.
- Jin, J. C. (2000). Openness and growth: An interpretation of empirical evidence from East Asian countries. *Journal of International Trade and Economic Development*, 9(1), 5–17.
- Jin, J. C., & Yu, E. S. H. (1995). The causal relationship between exports and income. *Journal of Economic Development*, 20(1), 131–140.
- Jin, J. C., & Yu, E. S. H. (1996). Export-led growth and the US economy: Another look. *Applied Economics Letters*, 3(5), 341–344.
- Jung, W., & Marchall, P. (1985). Exports, growth, and causality in developing countries. *Journal of Development Economics*, 18, 241–250.
- Karunaratne, N. D. (1996). Growth and trade dynamics under regime shifts in Australia. *Journal of Economic Studies*, 23(2), 55–69.
- Kasahara, H., & Rodrigue, J. (2008). Does the use of imported intermediates increase productivity? Plant-level evidence. *Journal of Development Economics*, 87, 106–118.
- Kathuria, V. (2000). Productivity spillovers from technology transfer to Indian manufacturing firms. *Journal of International Development*, 12, 343–369.
- Kavoussi, R. M. (1985). International trade and economic development: The recent experience of developing countries. *Journal of Developing Areas*, 19(3), 379–392.
- Keller, W. (1998). Are International R&D spillovers trade-related? Analyzing spillovers among randomly matched trade parters. *European Economic Review*, 42, 1469–1481.
- Keller, W. (2004). International technology diffusion. *Journal of Economic Literature*, 42, 752–782.
- Keller, W., & Yeaple, S. (2003). *Multinational enterprises, international trade, and productivity growth: Firm-level evidence from the United States*. GEP Research Paper 03/13. University of Nottingham.
- Kemp, M. C. (1960). The mill-bastable infant industry dogma. *Journal of Political Economy*, 68(1), 65–67.
- Ketels, C., Lindqvist, G., & Solvell, O. (2006). *Cluster initiatives in developing and transition economies*. Stockholm: Center for Strategy and Competitiveness.
- Khalafalla, K. Y., & Webb, A. J. (2001). Export-led growth and structural change: Evidence from Malaysia. *Applied Economics*, 33(13), 1703–1715.
- Khan, A. H., & Saqib, N. (1993). Exports and economic growth: The Pakistan experience. *International Economic Journal, Korean International Economic Association*, 7(3), 53–63.
- Kim, E. (2000). Trade liberalization and productivity growth in Korean manufacturing industries: Price protection, market power, and scale efficiency. *Journal of Development Economics*, 62(1), 55–83.
- Kinoshita, Y. (2001). *R&D and technology spillovers through FDI: Innovation and pacify*. CEPR Discussion Paper DP2775.
- Kohli, I., & Singh, N. (1989). Exports and growth: Critical minimum effort and dimishing returns. *Journal of Development Economics*, 30(2), 391–400.
- Kokko, A. (1994). Technology. Market characteristics, and spillovers. *Journal of Development Economics*, 43, 279–293.
- Kokko, A. (1996). Productivity spillovers from competition between local firms and foreign affiliates. *Journal of International Development*, 8, 517–530.
- Kokko, A., Tansini, R., & Zejan, M. (1996). Local technological capability and productivity spillovers from FDI in the uruguayan manufacturing sector. *Journal of Development Studies*, 32, 602–611.
- Kokko, A., Zejan, M., & Tansini, R. (2001). Trade regimes and spillover effects of FDI: Evidence from Uruguay. *Weltwirtschaftliches Archiv*, 137, 124–149.
- Konings, J. (2001). The effects of foreign direct investment on domestic firms: Evidence from firm level panel data in emerging economies. *Economics of Transition*, 9, 619–633.
- Krueger, A. O., & Tuncer, B. (1982). An empirical test of the infant industry argument. *American Economic Review*, 72(5), 1142–1152.
- Krugman, P. R. (1987a). *Trigger strategies and price dynamics in equity and foreign exchange markets*. NBER Working Paper No. 2459.
- Krugman, P. (1987b). The narrow moving band, the Dutch disease, and the competitive consequences of Mrs. Thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of Development Economics*, 27, 41–55.

- Krugman, P. (1991). *Geography and trade*. New York: The MIT Press.
- Krugman, P., & Obstfeld, M. (2008). *International economics: Theory and policy* (8th ed.). New York: Harper Collins College Publisher.
- Kugler, M. (2001). *The diffusion of externalities from foreign direct investment: The sectoral pattern of technologies spillovers*. Mimeo, University of Southampton.
- Kunst, R. M., & Marin, D. (1989). On exports and productivity: A causal analysis. *Review of Economics and Statistics*, 71(4), 699–703.
- Kwan, A. C. C., & Cotsomitis, J. A. (1990). Economic growth and the expanding export sector: China 1952–1985. *International Economic Journal*, 5(1), 105–117.
- Kwan, A. C. C., Cotsomitis, J. A., & Kwok, B. (1996). Exports, economic growth and exogeneity: Taiwan 1953–1988. *Applied Economics*, 28(4), 467–471.
- Kwan, A. C. C., & Kwok, B. (1995). Exogeneity and the export-led growth hypothesis: The case of China. *Southern Economic Journal*, 61(4), 1158–1166.
- Lee, J. W. (1995). Capital goods imports and long-run growth. *Journal of Development Economics*, 48, 91–110.
- Lee, J. (1996). Government interventions and productivity growth in Korean manufacturing industries. *Journal of Economic Growth*, 1(3), 391–414.
- Lee, H. Y., Ricci, L. A., & Rigobon, R. (2004). Once again, is openness good for growth? *Journal of Development Economics*, 75(2), 451–472.
- Lehmann Sibulle, H., & O'Rourke, K. H. (2008). *The structure of protection and growth in the late 19th century*. NBER Working Paper No. 14493.
- Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82(4), 942–963.
- Levinsohn, J. (1993). Testing the imports-as-market-discipline hypothesis. *Journal of International Economics*, 35, 1–22.
- Lewis, S. A. (1955). *The theory of economic growth*. London: Allen and Unwin.
- Li, X., Liu, X., & Parker, D. (2001). Foreign direct investment and productivity spillovers in the Chinese manufacturing sector. *Economic Systems*, 25, 305–321.
- Lileeva, A. (2004). *Import competition and selection*. Working Paper. York University.
- Lileeva, A., & Trefler, D. (2007). *Improved access to foreign markets raises plant-level productivity*!for some plants. NBER Working Paper No. 13279.
- Lipsey, R. E., & Sjoholm, F. (2004). Foreign direct investment, education and wages in Indonesian manufacturing. *Journal of Development Economics*, 73(1), 415–422.
- List, F., Matile, G.-A., Richelot, H., & Colwell, S. (1856). *National system of political economy*. Philadelphia: J.B. Lippincott & Co.
- Liu, Z. (2008). Foreign direct investment and technology spillovers: Theory and evidence. *Journal of Development Economics*, 85, 176–193.
- Liu, P., Liu, Z., & Yifan, Z. (2008). *Do Chinese domestic firms benefit from FDI inflow? Evidence of horizontal and vertical spillovers*. Working Paper. Lingnan University, Hong Kong.
- Liu, X., Siler, P., Wang, C., & Wei, Y. (2000). Productivity spillovers from foreign direct investment: Evidence from UK industry level panel data. *Journal of International Business Studies*, 31, 407–425.
- Liu, X., Song, H., & Romilly, P. (1997). An empirical investigation of the causal relationship between openness and economic growth in China. *Applied Economics*, 29(12), 1679–1686.
- Lopez-Cordova, J. E. (2002). *NAFTA and Mexico's manufacturing productivity: An empirical investigation using micro-level data*. Mimeo, International American Development Bank.
- Love, J. (1994). Engines of growth: The export and government sectors. *World Economy*, 17(2), 203–218.
- Lucas, R. E. Jr., (1988). On the mechanics of development planning. *Journal of Monetary Economics*, 22, 3–42.
- Luzio, E., & Greenstein, S. (1995). Measuring the performance of a protected infant industry: The case of Brazilian microcomputers. *Review of Economics and Statistics*, 77, 622–633.
- Madden, G., Savege, S. J., & Bloxham, P. (2001). Asian and OECD international R&D spillovers. *Applied Economics Letters*, 8(7), 431–435.

- Mallick, S. K. (1996). Causality between exports and economic growth in India: Evidence from cointegration based error correction models. *Indian Journal of Economics*, 76, 307–320.
- Manova, K. (2006). *Credit constraints, heterogenous firms, and international trade*. Harvard University.
- Marin, D. (1992). Is the export-led growth hypothesis valid for industrialized countries. *Review of Economics and Statistics*, 74(4), 678–688.
- Markusen, J. R. (1989). Trade in producer services and in other specialized. Intermediate inputs. *American Economic Review*, 79, 85–95.
- Marshall, A. (1920). *Principles of economics*. London: MacMillan.
- Martins, P. S., & Esteves, L. A. (2006). *Is there rent sharing in developing countries?* Institute for the Study of Labor (IZA). Discussion Papers: Matched-Panels Evidence from Brazil, 2317.
- Matsuyama, K. (1991). Increasing returns, industrialization, and indeterminacy of equilibrium. *The Quarterly Journal of Economics*, 106, 617–650.
- Mbaku, J. M. (1989). Export growth and economic performance in developing countries: Further evidence from Africa. *Journal of Economic Development*, 14(2), 127–142.
- McCarville, M., & Nnadozie, E. (1995). Causality tests of export-led growth: The case of Mexico. *Atlantic Economic Journal*, 23(2), 140–145.
- McNab, R. M., & Moore, R. E. (1998). Trade policy, export expansion, human capital and growth. *Journal of International Trade and Economic Development*, 7(2), 237–256.
- Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695–1725.
- Melitz, M. (2005). When and how should infant industries be protected? *Journal of International Economics*, 66, 177–196.
- Mill, J. S. (1909). *Principles of political economy*. Green: London Longmans.
- Mobarak, A. M., & Purbasari, D. P. (2006). *Do multinational corporations seek out politically connected firms?*. Working Paper, Firm Level Evidence on Local Partner Search by Multinational Corporations.
- Moosa, I. A. (1999). Is the export-led growth hypothesis valid for Australia? *Applied-Economics*, 31(7), 903–906.
- Moreira, M. M., & Najberg, S. (2000). Trade liberalisation in Brazil: Creating or exporting jobs? *Journal of Development Studies*, 36(3), 78–99.
- Morley, B., & Perdikis, N. (2000). Trade liberalisation, government expenditure and economic growth in Egypt. *Journal of Development Studies*, 36(4), 38–54.
- Moschos, D. (1989). Export expansion. Growth and the level of economic development. *Journal of Development Economics*, 30, 93–102.
- Muendler, M.-A. (2001). *Trade and growth revisited: Managing to converge, agreeing to diverge*. University of California, San Diego. unpublished manuscript.
- Muendler, M.-A. (2004). *Trade, technology and productivity: A study of Brazilian manufacturers 1986–1998*. CESifo Working Paper Series No. 1148. University of California, San Diego.
- Muendler, M.-A. (2007). *Trade and Workforce Changeover in Brazil*. NBER Working Papers No. 12980.
- Mulaga, G., & Weiss, J. (1996). Trade reform and manufacturing performance in Malawi. *World Development*, 24(7), 1267–1278.
- Murphy, K. M., Shleifer, A., & Vishny, R. W. (1989). Industrialization and the Big Push. *Journal of Political Economy*, 97(5).
- Nishimizu, M., & Page, J. (1982). Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity. *The Economic Journal*, 92, 920–936.
- Noland, M., & Pack, H. (2003). *Industrial policy in an era of globalization: Lessons from Asia*. Washington DC: Institute for International Economics.
- Nunn, N., & Trefler, D. (2006). Putting the Lid on lobbying: Tariff Structure and Long-Term Growth when Protection is for Sale, NBER Working Paper No. 12164.
- Okuno-Fujiwara, M. (1988). Interdependence of industries, coordination failure and strategic promotion of an industry. *Journal of International Journal*, 25, 25–43.
- Olley, S., & Pakes, A. (1993). *A limit theorem for a smooth class of semiparametric estimators*. Cowles Foundation Discussion Papers, No. 1066. Yale University.

- Onafowora, O. A., & Owoye, O. (1998). Can trade liberalization stimulate economic growth in Africa? *World Development*, 26(3), 497–506.
- O'Rourke, K. H. (2000). Tariffs and growth in the late 19th century. *Economic Journal*, 110, 456–483.
- Oskooee, B. M., & Also, J. (1993); Export growth and economic growth: An application of cointegration and error-correction modeling. *Journal of Developing Areas*, 27(4), 535–542.
- Oxley, L. (1993). Cointegration, causality and export-led growth in Portugal, 1865–1985. *Economics Letters*, 43(2), 163–166.
- Pack, H., & Saggi, K. (2006). Is there a case for industrial policy? A critical survey. *The World Bank Research Observer*, 21(2), 267–297.
- Park, J. H., & Prime, P. B. (1997). Export performance and growth in China: A cross-provincial analysis. *Applied Economics*, 29(13), 53–63.
- Park, A., Yang, D., Shi, X., & Jiang, Y. (2006). *Exporting and firm performance: Chinese exporters and the Asian financial crisis*. Working Paper No. 549. University of Michigan.
- Paul, S., & Chowdhury, K. (1995). Export-led growth hypothesis: Some empirical testing. *Applied Economics Letters*, 2(6), 177–179.
- Pavcnik, N. (2002). Trade liberalization, Exit, and productivity improvement: Evidence from chilean plants. *Review of Economic Studies*, 69(1), 245–276.
- Pelzman, J., & Shoham, A. (2006). De-linking the Relationship between trade liberalization and reduced domestic fiscal budgets: The experience of the Israeli economy: 1984–2005. *Global Economy Journal*, 6, Article 2.
- Pereira, A. M., & Xu, Z. (2000). Export growth and domestic performance. *Review of International Economics*, 8(1), 60–73.
- Piazolo, M. (1996). Determinants of indonesian economic growth, 1965–1992. *Seoul Journal of Economics*, 9(4), 269–298.
- Pineres, A.-S., & Ferrantino, M. J. (1997). Export diversification and structural dynamics in the growth process: The case of Chile. *Journal of Development Economics*, 52(2), 375–391.
- Pomponio, X. Z. (1996). A causality analysis of growth and export performance. *Atlantic Economic Journal*, 24(2), 168–176.
- Poole, J. (2008). *Multinational spillovers through worker turnover*. UC Santa Cruz.
- Porter, M. (1990). *The competitive advantage of nations*. New York: Free Press. (republished with a new introduction, 1998).
- Porter, M. (1998). *Clusters versus industrial policy in on competition*. Boston, MA: HBS Press.
- Pritchett, L. (1996). Measuring outward orientation in LDCs: Can it be done? *Journal of Development Economics*, 49, 307–335.
- Rajan, R. G., & Zingales, L. (1998). Power in a theory of the firm. *The Quarterly Journal of Economics*, 113(2), 387–432.
- Ram, R. (1987). Exports and economic growth in developing countries: Evidence from time-series and cross-section data. *Economic Development and Cultural Change*, 36, 51–72.
- Ramondo, N. (2008). *Size, geography, and multinational production*. Working Paper.
- Ramondo, N., & Rodríguez-Clare, A. (2008). *Trade, multinational production, and the gains from openness*. Working Paper. Pennsylvania State University.
- Rana, P. B. (1988). Exports, policy changes, and economic growth in developing countries after the 1973 oil Shock: Comments. *Journal of Development Economics*, 28(2), 261–264.
- Rashid, A. I. (1995). Trade, growth, and liberalization: The Indian experience, 1977–1989. *Journal of Developing Areas*, 29(3), 355–370.
- Redding, S. (1999). *Dynamic comparative advantage and the welfare effects of trade*, Oxford economic papers (Vol. 51(1), pp. 15–39). Oxford: Oxford University Press.
- Rhee, Y. W., & Belot, T. (1990). *Export catalysts in low-income countries*. World Bank Discussion Paper No. 72.
- Riezman, R. G., Whiteman, C. H., & Summers, P. M. (1996). The engine of growth or its handmaiden? A time-series assessment of export-led growth. *Empirical Economics*, 21(1), 77–110.
- Rodriguez, F. (2008). *Policymakers beware: The use and misuse of regressions in explaining economic growth*, policy research brief 5. International Poverty Centre.

- Rodríguez, F., & Rodrik, D. (1999). *Trade policy and economic growth: A skeptic's guide to cross-national evidence*. NBER Working Paper No. W7081.
- Rodríguez-Clare, A. (1996a). Multinationals, linkages and economics development. *American Economic Review*, 86, 852–873.
- Rodríguez-Clare, A. (1996b). The division of labor and economic development. *Journal of Development Economics*, 49(1).
- Rodríguez-Clare, A. (2007). Clusters and comparative advantage: Implications for industrial policy. *Journal of Economic Development*, 82(1), 43–57.
- Rodrik, D. (1996). Understanding economic policy reform. *Journal of Economic Literature*, XXXIV, 9–41.
- Rodrik, D. (2004). *Industrial policy for the twenty-first century*. Harvard University.
- Rodrik, D. (2006). What's so special about China's exports? *China and World Economy*, 14, 1–19.
- Rodrik, D. (2007). *Industrial development: Stylized facts and policies*, in United Nations, *Industrial Development for the 21st Century*. New York: U.N.
- Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: The primary of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9, 131–165.
- Romalis, J. (2007). *Market access, openness and growth*. NBER Working Papers: 13048, National Bureau of Economic Research, Inc.
- Romer, P. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94, 1002–1037.
- Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, 98, s71–s102.
- Romer, P. (1994). The origins of endogenous growth. *Journal of Economic Perspectives*, 8, 3–22.
- Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4). Amsterdam: Elsevier, North-Holland.
- Rosenthal, S. S., & Strange, W. C. (2005). The micro-empirics of agglomeration economies. In R. Arnott & D. McMillen (Eds.), *Blackwell companion to urban economics*. Malden, MA: Blackwell.
- Ruane, F., & Ugur, A. (2002). *Foreign direct investment and productivity spillovers in Irish manufacturing industry: Evidence from firm level panel data*. Trinity Economic Papers 02/06. Trinity College, Dublin.
- Sachs, J. D., & Warner, A. M. (1995). *Natural resource abundance and economic growth*. NBER Working Paper No. W5398.
- Salvatore, D. (1983). A simultaneous equations model of trade and development with dynamic policy simulations. *A Model of Trade and Development*, 36, 66–90.
- Salvatore, D., & Hatcher, T. (1991). Inward oriented and outward oriented trade strategies. *Journal of Development Studies*, 27(3), 7–25.
- Sauré, P. (2007). Revisiting the infant industry argument. *Journal of Development Economics*, 84, 104–117.
- Sengupta, J. K. (1993). Growth in NICs in Asia: Some tests of new growth theory. *Journal of Development Studies*, 29(2), 45–60.
- Sengupta, J., & Espana, J. R. (1994). Exports and economic growth in Asian NICs: An econometric analysis for Korea. *Applied Economics*, 26, 41–51.
- Serletis, A. (1992). Export growth and canadian economic development. *Journal of Development Economics*, 38, 133–145.
- Shan, J., & Sun, F. (1998a). Domestic saving and foreign investment in Australia: A Granger Causality Test. *International Economics Journal*, 12(4), 79–87.
- Shan, J., & Sun, F. (1998b). On the export-led growth hypothesis: The econometric evidence from China. *Applied Economics*, 30(8), 1055–1065.
- Shan, J., & Sun, F. (1998c). On the export-led growth hypothesis for the little dragons: An empirical re-investigation. *Atlantic Economic Journal*, 26(4), 353–371.
- Sharma, S. C., & Dhakal, D. (1994). Causal analysis between exports and economic growth in developing countries. *Applied Economics*, 26, 1145–1157.
- Sharma, S. C., Norris, M., & Cheung, D. W. W. (1991). Exports and economic growth in industrialized countries. *Applied Economics*, 23, 697–707.
- Sheehey, E. (1990). Exports and growth: A flawed framework. *Journal of Development Studies*, 111–116.
- Sheehey, E. J. (1992). Exports and growth: Additional evidence. *Journal of Development Studies*, 28(4), 730–734.

- Singer, H. W., & Gray, P. (1988). Trade policy and growth of developing countries: Some new data. *World Development*, 16(3), 395–403.
- Singh, T. (2003). Effects of exports on productivity and growth in India: An industry-based analysis. *Applied Economics*, 35, 741–749.
- Sjoholm, F. (1999a). Technology gap. Competition and spillovers from direct foreign investment: Evidence from establishment data. *Journal of Development Studies*, 36, 53–73.
- Sjoholm, F. (1999b). Productivity growth in Indonesia: The role of regional characteristics and direct foreign investment. *Economic Development and Cultural Change*, 47, 559–584.
- Smarzynka, B. K. (2002). *Does foreign direct investment increase the productivity of domestic firms?* World Bank Policy Research Working Paper: In search of Spillovers through Backward Linkages 2923.
- Smarzynska-Javorcik, B. (2004). Does foreign direct investment increase productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review*, 94(3), 605–627.
- Spencer, B. J., & Brander, J. A. (1983, August). *International R&D rivalry and industrial strategy*. (Vol. w1192). NBER Working Paper Series.
- Succar, P. (1987). The need for industrial policy in LDCs—A re-statement of the infant industry argument. *International Economic Review*, 28(2).
- Thangavelu, S. M., & Rajaguru, G. (2004). Is there an export or import-led productivity growth in rapidly developing Asian countries? A multivariate VAR analysis. *Applied Economics*, 36(10), 1083–1093.
- Thornton, J. (1996). Cointegration, causality and export-led growth in Mexico, 1895–1992. *Economic Letters*, 50, 413–416.
- Topalova, P. (2004). *Trade liberalization and firm productivity: The case of India*. International Monetary Fund Working Paper No. 04/28.
- Tybout, J. (1992). Linking trade and productivity: New research directions. *World Bank Economic Review*, 6(May), 198–211.
- Tybout, J. R., & Westbrook, M. D. (1995). Trade liberalization and the dimensions of efficiency change in Mexican manufacturing industries. *Journal of International Economics*, 39, 53–76.
- Tyler, W. G. (1981). Growth and export expansion in developing countries. *Journal of Development Economics*, 9, 121–130.
- Ukpolo, V. (1994). Export composition and growth of selected low-income African countries: Evidence from time-series data. *Applied Economics*, 26, 445–449.
- Vamvakidis, A. (1998). *Regional trade agreements versus broad liberalization: Which path leads to faster growth? Time-series evidence*. international monetary fund. IMF Working Papers: 98/40.
- Van Bieseboeck, J. (2003). *Revisiting some productivity debates*. NBER Working Papers No. W10065.
- Van Bieseboeck, J. (2005). Exporting raises productivity in sub-Saharan African manufacturing firms. *Journal of International Economics*, 67(2), 373–391.
- Velde, D. W., & Morrissey, O. (2003). Do workers in africa get a wage premium if employed in firms owned by foreigners? *Journal of African Economics*, 12(1), 41–73.
- Waciarg, R., & Welch, K. H. (2003). *Trade liberalization and growth: New evidence*. NBER Working Papers No. 10152.
- Wade, R. (1990). *Governing the market: Economic theory and the role of government in East Asian industrialization*. Princeton: Princeton University Press.
- Weinhold, D., & Rauch, J. E. (1999). Openness, specialization, and productivity growth in less developed countries. *Canadian Journal of Economics*, 32(4), 1009–1027.
- Wilbur, W. L., & Haque, M. Z. (1992). An investigation of the export expansion hypothesis. *Jounral of Development Studies*, 28(2), 297–313.
- World Bank. (1993). *The East Asian miracle: Economic growth and public policy*. Oxford: Oxford University Press.
- Xu, B., & Wang, X. (1999). Capital goods trade and R&D spillovers in the OECD. *The Canadian Journal of Economics*, 32(5), 1258–1274.
- Yaghmaian, B. (1994). An empirical investigation of exports, development, and growth in developing countries: Challenging the neoclassical theory of export-led growth. *World Development*, 22(12), 1977–1995.

- Yanikkaya, H. (2003). Trade openness and economic growth: A cross-country empirical investigation. *Journal of Development Economics*, 72(1), 57–89.
- Young, A. (1991). Learning by doing and the dynamic effects of international trade. *Quarterly Journal of Economics*, 106, 369–405.
- Zagha, R., & Nankani, G. T. (2005). *Economic growth in the 1990s: Learning from a decade of reform*. Washington, DC: World Bank Publication.
- Zukowska-Gagelmann, K. (2002). Productivity spillovers from foreign direct investment in Poland. *Economic Systems*, 24, 223–256.

Monetary and Exchange Rate Policies

Eduardo Levy Yeyati and Federico Sturzenegger¹

Escuela de Negocios, Universidad Torcuato Di Tella, Minones 2177 (1428), Buenos Aires, Argentina

Contents

1. Introduction	4216
1.1 Do nominal variables matter for development?	4217
2. What Do We Talk About When We Talk About MERP?	4220
2.1 Classifying exchange rate policies	4224
2.2 Exchange rate policy trends in the post-Bretton Woods era: A casual glance at the distribution of regime choices	4230
2.3 Inside the nonfloat group	4233
2.4 Monetary policy: From exchange rate anchors to inflation targeting	4236
3. Why Do We Talk About MERP?	4240
3.1 The link between exchange rates and growth revisited	4241
3.2 ERR and output volatility	4246
3.3 MERP on price stability	4247
3.4 MERP on income distribution	4248
3.5 Indirect links: The integration channel	4249
3.6 Indirect links: The financial channel	4251
3.7 Summing up	4254
4. The Making of a Policy	4254
4.1 Exchange rate anchors in the 1980s	4254
4.2 Financial integration and financial crises in the 1990s	4258
4.3 Float cum inflation targeting (FIT)	4260
4.4 The comeback of exchange rate regimes: Leaning against the appreciation wind	4263
5. Where Do We Stand?	4267
End Notes	4269
References	4273

Abstract

To the extent that they exert a critical influence on the macroeconomic environment, monetary and exchange rate policies (MERP) are relevant for development. However, the analytical economic literature often sees nominal variables as being irrelevant for the real economy, while the multiplicity of channels examined by the empirical literature complicates the task of deriving

usable policy implications. Specifically, this chapter attempts to answer the following question: What exchange rate regime and monetary policy framework is more conducive to achieving development policy objectives in a particular country today, and why? We map the direct and indirect links from MERP to key development objectives, and discuss the main findings and how it relates with the empirical evidence to provide an up-to-date perspective of the policy debate and derive criteria for policy choices.

JEL classifications: E42, E52, F33, F41

Keywords

monetary standard
monetary system
monetary policy
policy effects
currency

1. INTRODUCTION

How to achieve a stable, sustainable, and equitable growth path is the defining question in development economics. Unfortunately, a clear answer to this question has proved to be as elusive as the question is important, not the least because virtually any decision that policy makers make can be argued to have an effect on at least some of the main development dimensions. Monetary and exchange rate policies (MERP), the subject of this chapter, are no exception to this rule. Because they determine, to a great extent, the macroeconomic environment in which the economy operates, its relevance to development appears to be quite natural. But the analytical economic literature has not been supportive of this connection: nominal variables are often considered as irrelevant for the real economy in the long run. Moreover, even in the vast body of work that explores this link empirically, the multiplicity of country-specific channels that have been proposed—and usually examined separately—make the task of deriving usable policy conclusions rather arduous.

As we will argue below, the choice of MERP not only has a direct implication on the evolution of key nominal variables of the economy (prices, the exchange rate) and, as a result, on output volatility and the financial sector—which, in turn, may have an effect on policy objectives such as output growth and income distribution, but it may also affect many other variables that are only somewhat related to monetary issues. For example, stable exchange rates may foster trade, or feed into financial fragilities as it undermines the incentives of agents to hedge against currency risk. It is these indirect relations that explain, for example, why the adoption of the Euro was—at least officially—predicated on its potential trade gains rather than on the benefits of a monetary

anchor; or why the preference for nonpegged regimes is argued on the need to elude costly speculative currency attacks.

To tackle such a broad range of topics would be impossible without narrowing down the scope to a subset of issues that can be meaningfully covered within the space constraints imposed by this chapter. To meet this objective, we will focus on the aspects that we consider more relevant in the design of a development policy from the perspective of a policy maker in a small open economy. Specifically, our exploratory trip will be oriented to answer the following question: What exchange rate regime and monetary policy framework is more conducive to achieving development policy objectives in a particular country today, and why? Because the answer to this question cannot ignore the external and domestic scenario, or the structural characteristics of the country, our exploration will yield criteria for policy choices rather than one-size-fits-all recipes.

1.1 Do nominal variables matter for development?

Before getting into the crux of the matter, it is useful to revisit the broader methodological debate spurred by the overarching theme of the link between nominal and real variables in open economies. A good starting point is the so-called “classical dichotomy,” which argues that nominal variables have no lasting effect on the real economy, beyond, maybe, short-run output fluctuations. If so, it would be inconsequential whether countries choose fixed or floating regimes because price flexibility would make nominal variables irrelevant in the long run. At most, it would be argued in this context that monetary policy and the evolution of nominal aggregates will be related to the choice of the inflation rate, which in turn could have an effect on economic performance; any other channel would be obliterated in such a framework. Notice, however, that this argument is in stark contrast with the relevance assigned to MERP in policy discussions, where the choice of exchange rate regimes and monetary policy frameworks are considered critical—a reflection that the perfectly flexible and frictionless classical world is not a complete representation of reality.

Nevertheless, the skeptical classical view on the relevance of MERP has found some support in empirical work. In a classical reference, Baxter and Stockman (1989) looked at the time series properties of several macro variables and found that their change over time showed little relation with the choice of exchange rate regimes. They acknowledged that the real exchange rate appeared to move more under floating arrangements, but this did not affect the behavior of real variables. Backus and Kehoe (1992) also looked at the properties of output and prices over the whole of the twentieth century and found that the properties of business cycles have remained fairly constant regardless of the changing exchange rate regimes, and in spite of the fact that the evolution of price levels did exhibit significant differences, particularly before and after WWII. A similar result was found later by Flood and Rose (1995), who argued that there was little effect of exchange rate regimes on the volatility of output, stock markets, or even monetary

policy. This work, however, focused on industrialized countries and is thus of relative use for policy makers in developing economies as we will see below.

The view that nominal variables are irrelevant has been the premise underlying a large literature on real models that simply do not include MERP as a relevant dimension to understand open economies, but these models have had only minor empirical success in describing the business cycles in open economies (see Box 1). This is not surprising, since prices are generally less than perfectly flexible (particularly, when it comes to price cuts) and markets are generally less than perfectly efficient.

Box 1 Real Models

The classical dichotomy provides the justification for a large body of literature known as real models that include no reference whatsoever to monetary or exchange rate policies. In recent years “real models” have become a synonym of a large class of models representing the “real business cycle approach” pioneered by Kydland and Prescott (1986) and extended to open economies by Backus, Kehoe, and Kydland (1992). The innovation of these models is that they claimed to be able to replicate the patterns in output and main macroeconomic variables at business cycle frequencies, thus providing further support for the classical view.

Typically, these models postulate an economy with a representative intertemporal-utility-maximizing consumer that faces a labor-leisure choice. In the closed economy version, the consumer decides how much to save and invest, and the system is shocked by productivity disturbances that drive the dynamics. In open economies, however, the representative consumer can also trade goods and financial assets with other countries. The solution technique which consists of looking for the central planner solution is tantamount to assuming access to complete financial markets. The methodology consists in simulating an economy subject to shocks (which bear some resemblance with those in the real world), deriving the rational response of optimizing agents, and contrasting the statistical properties of the solution with those of the data. While highly elegant, and relatively successful in closed economies, the models have performed poorly in open economies. Investment tends to move dramatically in response to changes in productivity across countries, while the financial structure implies consumption levels that appear to be too correlated across countries. In addition, while output across countries typically is positively correlated, the models deliver a negative correlation. Mendoza (1991) provides a canonical application to small open economies, but again he needs to assume a large cost of adjustment for investment in order to obtain reasonable results.

While it is true that this framework has had little power to explain the overall macro data in open economies, this does not mean that the exercises are not useful. Sometimes these poor results are somewhat helpful in suggesting relevant deviations from the canonical basic structure. For example, Neumeyer and Perri (2004) show that in small open economies shocks to the risk premia help explain the volatility of outputs seen in a standard emerging economy, and Chari, Kehoe, and McGrattan (2005) show that in sudden capital account reversals, a real model suggests an output expansion, and not an observed contraction—a result that can be used to argue that it is not the capital reversal *per se* that accounts for the decline in output observed in these episodes but its combination with, for example, the use of tradables in the production function, or the typical balance-sheet borrowing constraint popularized in the third generation currency crisis models.

In fact, price rigidities are the key assumption behind the Mundellian view of the role of exchange rates. In a seminal contribution that kick started what has since been known as the theory of optimal currency areas, Mundell (1963) argued that exchange rate flexibility was useful both as a shock absorber and as an expenditure switching instrument to attain internal balance. Mundell argued that countries should weigh these benefits against the gains of stable exchange rates that reduce the costs of international trade (both due to reduced transaction costs and price uncertainty). In his setup, the gains from fixing are enhanced by openness (because it reduces the needed exchange rate adjustment: the larger the trade share, the easier it is to accommodate external shocks without major changes in relative prices) and labor mobility, or fiscal transfers (which make up for price rigidity, facilitating income smoothing within the currency area). Similarly, concentrated trade with a single partner increases the benefits of fixing vis-a-vis this partner's currency, because it maximizes the trade benefits of exchange rate stability. On the other extreme, volatile terms of trade call for greater exchange rate flexibility to facilitate adjustments to real shocks. Most of these predictions, as we will see below, are broadly validated by the data. Ultimately, in Mundell's world, MERP amounted to a tradeoff between output smoothing and trade gains.²

While highly popular among policy makers, this approach has been criticized in academic circles for its sometimes *ad hoc* assumptions and imprecise welfare implications. Overcoming these weaknesses has been the agenda of a large literature that has attempted to rescue the main intuitions of the paradigm, in models which provide explicit microfoundations for price rigidities in a world of optimizing rational agents. In a nutshell, the new Keynesian models in international finance typically consist of three equations: a dynamic IS curve, a Philips curve, and a policy reaction function. The IS curve is derived from the Euler equation of consumer maximization, where aggregate demand matters because the model assumes monopolistic competition, whereas the Philips curve is built on the assumption of price rigidities.³ Monetary policy, in turn, is usually represented by an interest rate rule. Because these models have well-defined objective functions they allow for precise statements on welfare, a key step to evaluate policy. With these models, the literature has come full circle, recovering the main tenets of the Mundellian approach, but now derived in coherent, fully specified general equilibrium models.⁴ More importantly, their emphasis on price rigidities and financial frictions sets the stage for a more realistic approach to the nominal-real link in the developing world.

To organize our presentation we need to distinguish between two aspects that have been at the center of the empirical literature, as it moved from industrial economies to a broader set of countries. First, the measurement of MERP, understood as the policy maker's reaction function, as opposed to the simple characterization of variables such as the interest rates or the exchange rate. Second, the precise identification and testing of

the (direct or indirect) channel through which MERP may exert their influence on the policy objectives.

In line with the implicit definition of the development problem proposed in the beginning of this chapter, we will focus on the following development policy objectives that have been recurrently discussed in the literature: stable output growth, price stability, and equitable distribution. How can MERP be characterized and how does it affect each of these policy dimensions? We approach these two questions in turn. We start in Section 2 with a conceptual characterization of exchange rate and monetary policies, including a critical survey of the many alternative classifications of exchange rate regimes that have appeared in recent years. In Section 3, we map the various channels linking MERP with development objectives, and survey the empirical evidence on each of them. Once MERP measurement and links to development objectives are properly discussed, in Section 4 we revisit the development policy question, bringing together theory and evidence to distill some criteria to help determine the optimal exchange rate/monetary policy mix.

2. WHAT DO WE TALK ABOUT WHEN WE TALK ABOUT MERP?

To summarize something that will become clear by the end of this chapter, the characterization of MERP in the real world is plagued by definitional and measurement problems that make any particular definition quite controversial. Hence, to analyze further, we need to be precise about what we understand by MERP. To that end, it is useful to start from the two-way scheme proposed by the International Monetary Fund.

The scheme for the latest period available online at the IMF website (mid-2006) is reproduced in Table 1. The rows indicate the *exchange rate regime*, which range from no national legal tender to fully floating exchange rates, spanning the standard three-way classification: pegs, intermediates, and floats. The columns characterize the *monetary policy framework* according to the target of choice for monetary policy (the exchange rate, the monetary aggregates, or the inflation rate). This reflects the fact that monetary policy has often been defined in terms of “nominal anchors,” namely, a nominal variable that the Central Bank chooses to be kept within a predetermined range as a means to anchoring expectations about the evolution of nominal variables, in general.⁵

Predictably, the table shows a strong correlation between the exchange rate regime and monetary policies, as reflected by the fact that countries tend to cluster along the diagonal. Countries that fix their exchange rate naturally choose the exchange rate as the nominal anchor. Conversely, countries that opt for a flexible exchange rate arrangement typically choose an alternative nominal anchor.

However, the correspondence between exchange rate and monetary policies is far from perfect. There are many different degrees of exchange rate commitments among those countries that use the exchange rate as nominal anchor, and there are alternative

Table 1 Exchange rate regimes and monetary frameworks

Exchange rate regime (number of countries)	Monetary policy framework					
	Exchange rate anchor		Monetary aggregate target	Inflation targeting framework	IMF- supported or other monetary program	Other
	Another currency as legal tender	ECCU				
CFA franc zone						Euro area
Another currency as legal tender	ECCU	WAEMU	CAEMC			
Exchange arrangements with no separate legal tender	Ecuador El Salvador Kiribati Marshall Islands Micronesia, Fed. States of Palau Panama San Marino, Timor-Leste, Dem. Rep. of	Antigua and Barbuda Dominica Grenada St. Kitts and Nevis St. Lucia St. Vincent and the Grenadines	Benin Burkina Faso Côte d'Ivoire Guinea-Bissau Mali Niger Senegal Togo	Cameroon Central African African Rep. Chad Congo, Rep. Equatorial Guinea Gabon	Austria Belgium Finland France Germany Greece	Ireland Italy Luxembourg Netherlands Portugal Spain
Currency board arrangements	Bosnia and Herzegovina Brunei Darussalam Bulgaria Hong Kong SAR		Djibouti Estonia Lithuania			

Continued

Table 1 Exchange rate regimes and monetary frameworks—Cont'd

Exchange rate regime (number of countries)	Monetary policy framework				
	Exchange rate anchor	Monetary aggregate target	Inflation targeting framework	IMF-supported or other monetary program	Other
Against a single currency					
Aruba	Cape Verde	Pakistan			
Ethiopia	Lesotho	Turkmenistan			
Bahamas,	China	Qatar			
The Guyana	Macedonia, FYR	Ukraine			
Bahrain,		Oman			
Kingdom of	Comoros	United Arab Emirates			
Honduras	Maldives	Rwanda	China		
Other conventional fixed peg arrangements	Barbados	Egypt	Venezuela, Rep.	Guyana	
	Iraq	Malta	Saudi Arabia	Sierra Leone	
	Belarus ⁷	Eritrea	Vietnam	Suriname	Pakistan
	Jordan	Mauritania	Seychelles		
	Belize	Namibia	Zimbabwe		
	Kuwait	Swaziland	Sierra Leone		
	Bhutan	Nepal	Solomon Islands		
	Latvia	Syrian Arab Rep.	Suriname		
	Bolivia	Netherlands			
	Lebanon	Antilles			
		Trinidad and Tobago			
Against a composite					
Fiji		Samoa			
Libyan Arab Jamahiriya		Vanuatu			
Morocco					

Pegged exchange rates within horizontal bands	Within a cooperative arrangement		Other band arrangements					
	Cyprus Denmark	Slovak Rep. Slovenia	Hungary Tonga		Hungary Slovak Rep.			
Crawling pegs	Azerbaijan Botswana Costa Rica		Iran, I.R. of Nicaragua	Iran, I.R. of				
Managed floating with no predetermined path for the exchange rate			Argentina Mauritius Bangladesh Moldova Cambodia Mongolia Gambia, The Sri Lanka Ghana Sudan	Haiti ⁷ Tajikistan Jamaica Tunisia Lao P.D.R. Uruguay Madagascar Yemen, R. Malawi Zambia	Colombia Czech Rep. Guatemala Peru Romania Serbia Rep. of Thailand	Afghanistan, I.R. Armenia Georgia Kenya Kyrgyz Rep. Papua N. Mozambique	Algeria Malaysia Angola Myanmar Burundi Nigeria Croatia Príncipe Liberia Guinea Singapore Dominican Rep.	Paraguay Guinea Russian Federation India S. Tomé and Príncipe Liberia Singapore Kazakhstan Uzbekistan
Independently floating			Albania Congo, Dem. Rep. of Indonesia Uganda	Australia Brazil Canada Chile Iceland Israel Korea Mexico New Zealand	Norway Philippines Poland South Africa Sweden Turkey United Kingdom Zealand	Japan Somalia Switzerland United States Tanzania		

anchors used by countries that favor more flexible exchange rate regimes. In some cases, the classification is unclear as the entry in the upper right quadrant indicates: the euro area could be characterized as a fixed regime vis-a-vis other union members, and a float vis-a-vis the rest of the world: Should that be included in the fix or float group? Classifying regimes has proved so challenging as to span a small literature on alternative methodologies and their empirical implications.

2.1 Classifying exchange rate policies

Few economists would contest the textbook definition of canonical exchange rate regimes: fixed regimes involve a commitment to keep the nominal exchange rate at a given level (typically, through central bank purchases and sale of foreign currency); floating regimes imply no market intervention by the monetary authorities and, therefore, an exchange rate that moves according to market forces to find its equilibrium (which could tautologically be defined as that induced by market forces in the absence of intervention).

Reality, as hinted in Table 1, is much more nuanced. Hardly any textbook float can be found among developing countries, and the empirical distinction between alternative nonpegged regimes is not always clear. Moreover, actual policies often tend to differ significantly from stated intentions. For example, it is not unusual that a country that officially announces a fixed exchange rate adjusts its parity if it finds the constraints imposed by the peg (on monetary policy or economic activity) too taxing. By the same token, there are situations in which a country that commits to a flexible exchange rate may choose to intervene in the foreign exchange market to dampen exchange rate fluctuations. There is vast anecdotal evidence on both behaviors. For example, realignments have been a pervasive feature of fixed exchange rates in emerging economies, and some countries which claimed to run a floating regime have exhibited very stable exchange rates (e.g., El Salvador prior to its unilateral dollarization).

Figure 1 shows the distribution of monthly changes in the exchange rate among fixers and floaters classified according to the IMF's *de jure* regime classification.⁶ As can be seen, many *de jure* pegs display large monthly exchange rate variations, while many floats exhibit very little exchange rate variability (Figure 1).⁷ The same can be said of the change in reserves. While one would expect this change to be smaller under floating regimes, the distribution of changes between floats and fixes (again, as defined by the IMF) is virtually indistinguishable (Figure 2).

This weak link between the variable that supposedly defines the regime and the official classification has led to the development of alternative classifications. In all cases, these attempts relied, to different degrees, on policies observed, and as such have been dubbed *de facto* classifications of exchange rate regimes. Table 2 succinctly describes the most widely used classifications.⁸

These new classifications go from the textbook three-way taxonomy (float-intermediate-fix) to more nuanced groupings that distinguish specific modalities (such

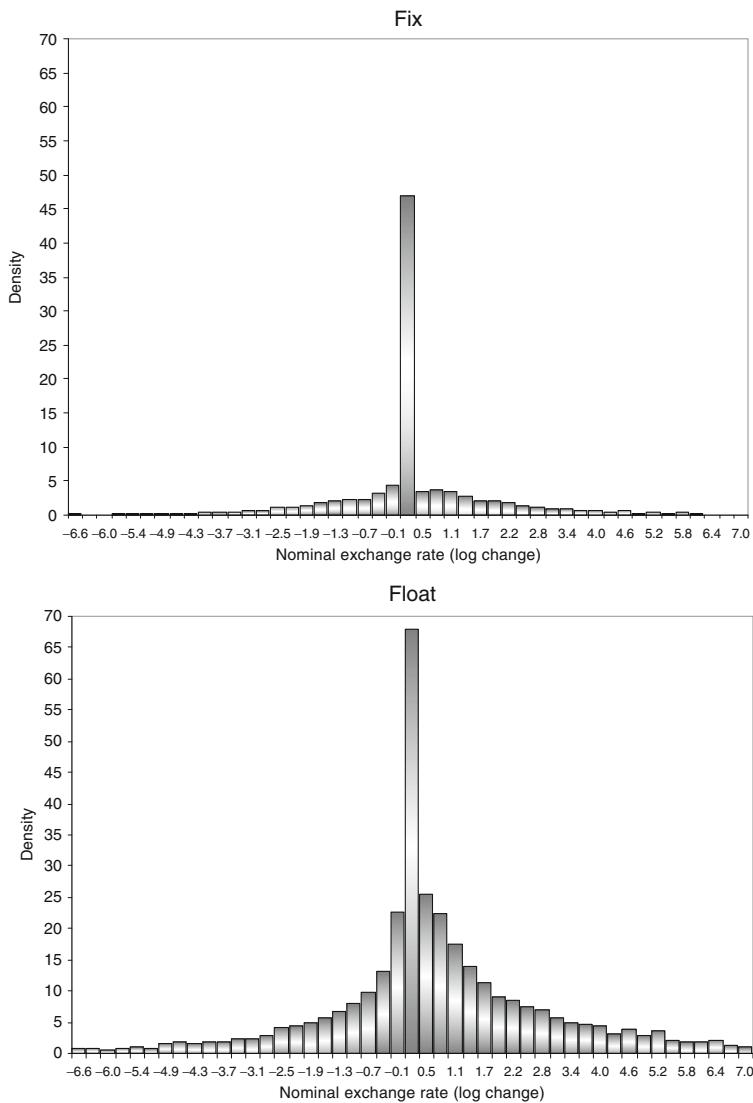


Figure 1 Exchange rate changes in fixers and floaters. Source: Author's calculations using IFS data.

as the Reinhart-Rogoff classification), and use to different degrees, a combination of statistical methods and reliance on the *de jure* classification. In their original paper, Ghosh, Gulde, Ostry, & Wolf (1997), for example, simply “corrected” the *de jure* classification by excluding from the peg group all countries that had more than one exchange rate realignment (in the parity or in the basket weights, depending on the case) during a calendar year. In Ghosh, Gulde, and Wolf (2003), they move further away from the original IMF coding to compute a *z*-score variable which combined

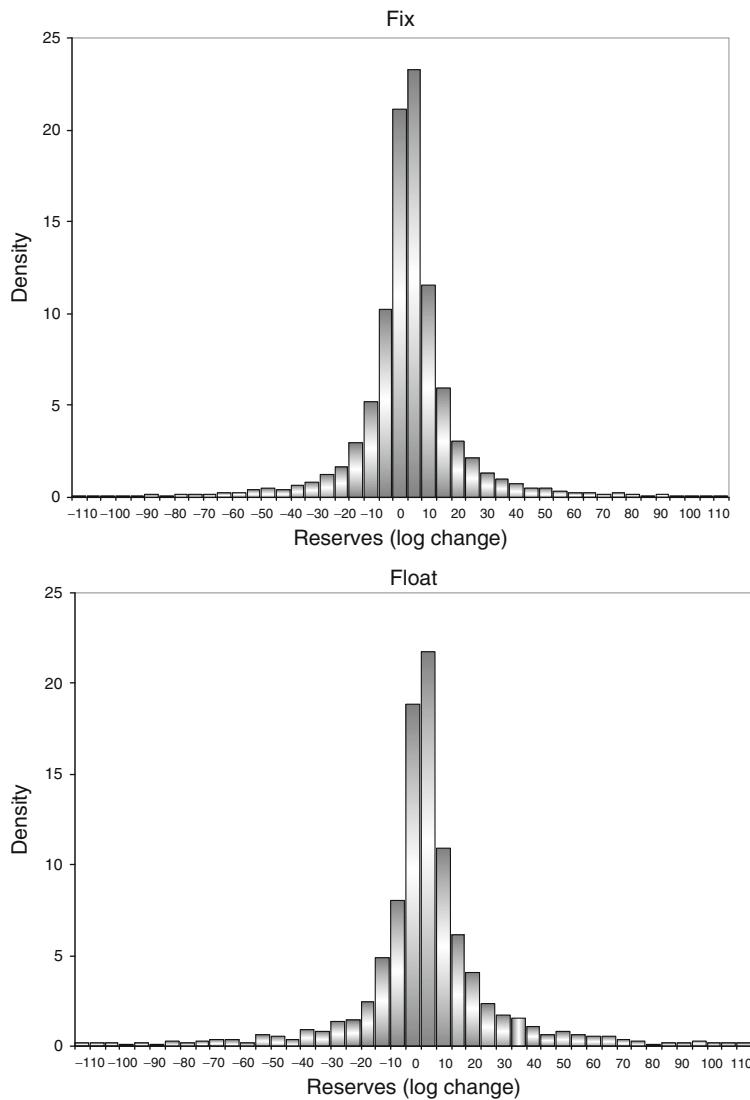


Figure 2 Reserve changes in fixers and floaters. Source: Author's calculations using IFS data.

the mean and variance of monthly depreciation rates and then mapped this continuous score into three *de facto* regimes (pegged, intermediate, and floats). Reinhart and Rogoff (2004) verified the compatibility of the *de jure* regime with the observed one; if this was found to be incompatible, they classified according to the volatility of the nominal exchange rate, identifying fixers with stable parities and floaters with more volatile parities. Others relied on purely statistical methods. Levy Yeyati and Sturzenegger (2001,

Table 2 *De facto* exchange rate regime classifications (in chronological order)

Study	Period	Frequency	Number of countries	Number of regime types	Approach
Ghosh et al. (1997), updated by Ghosh, Gulde, Wolf (2003)	1970–1999	Annual	150	3 coarse, 6 fine	Continuous classification based on z score, which is the square root of the square of both the mean and volatility of exchange-rate changes. Converted this measure into discrete classification using the relative-frequency distribution of <i>de jure</i> regimes. Retained those regimes for which the <i>de jure</i> and <i>de facto</i> methods coincided
Levy Yeyati and Sturzenegger (2001), extended backwards in Levy Yeyati and Sturzenegger (2005), updated in Levy Yeyati and Sturzenegger (2007)	1974–2004	Annual	179	3 coarse	Cluster analysis based on the behavior of exchange rate and reserves. Observations with very low exchange rate and reserve volatility excluded as inconclusive
IMF revised	1990–2003	Annual and monthly	190	3 coarse, 15 fine	<i>De jure</i> -based, revised according to the assessment of IMF desk economists, based on an analysis of exchange-rate and reserves

Continued

Table 2 *De facto* exchange rate regime classifications (in chronological order)—Cont'd

Study	Period	Frequency	Number of countries	Number of regime types	Approach
Bailliu, Lafrance, and Perrault (2003)	1973–98	Five-year average	60	3 coarse, 5 fine (pegged, intermediate with and without anchor, floating with and without anchor)	Used 2-step procedure: (1) regime classified as pegged if <i>de jure</i> peg or if exchange-rate volatility less than .45 percentage point in a given year, (2) remaining regimes classified on the basis of exchange rate volatility relative to average of country groups. Distinguished between regimes with and without anchors
Reinhart and Rogoff (2004)	1946–2001	Annual and monthly	153	5 coarse, 15 fine	Use the dual/parallel rate where it diverged from market rate. High inflations and crises grouped in the freely falling category (12-month rate of inflation above 40% or 6-months postcrisis period accompanied by a move from fix to float)
Shambaugh (2004)	1973–2000	Annual and monthly	155	2 coarse (pegs, nonpegs)	Used prespecified bands to determine if a regime is pegged or nonpegged. Tested only for degree of monetary autonomy
Dubas, Lee, and Mark (2005)	1971–2002	Annual	172	3 coarse 6 fine	Modeled <i>de jure</i> regimes as outcomes of a multinomial logit choice problem conditional on measures of volatility of (1) a country's effective exchange rate, (2) bilateral exchange rate against an anchor currency, and (3) international reserves. An “effective” <i>de facto</i> coding was obtained by assigning country-year observations to the regime with the highest predictive probability obtained from the multinomial logit

2005) computed the volatility of reserves and the nominal exchange rate, and then used cluster analysis to group countries: those with high exchange rate volatility and little reserves volatility go into the float cluster; those with high reserves volatility and little exchange rate volatility were assigned to the fix cluster, and those with moderate to high volatility in both dimensions were assigned to the intermediate cluster.

The key question when assessing alternative classifications is to what extent they capture appropriately the nature of exchange rate *policy* as opposed to the statistical behavior of exchange rates. As summarized in Table 2, many recent classifications have largely relied on the volatility of nominal exchange rates, paying no attention to the extent of policy intervention. As a result, countries with large movements in nominal exchange rates are typically classified as floats regardless of whether or not the authorities make efforts to reduce exchange rate volatility. Conversely, stable countries with little volatility are often classified as pegs in spite of little or no intervention. But, as illustrated in Figure 1, exchange rate volatility appears to be a poor indicator of exchange rate regimes.

This is not only a question of academic interest: assessing the positive implications of alternative exchange rate regimes will be critically influenced by how we classify them. In particular, mistaking flexibility with volatility may assign to the float category countries facing volatile external conditions or suffering strong market pressure, both situations that tend to coincide with subpar economic performance.

One way around this empirical conundrum may be to classify with attention to the intervention. Levy Yeyati and Sturzenegger (LYS) is an example of this approach: by comparing exchange rate volatility and changes in international reserves, they attempted to replicate the textbook analysis—according to which fixed regimes should exhibit little volatility in the nominal exchange rate coupled with larger movements in reserves—and sorted the data by similarity based on these two classifying dimensions. Along the same lines, Poirson (2001) used the ratio of the volatility of the nominal exchange rate to that of reserves.⁹

Even these broad ideas encounter significant difficulties when confronted with the specifics of each country. The many pending issues that plague existing classifications include the following:

- *Reference currency*: When testing for a regime, researchers need to define a currency (or a basket of currencies) that may be targeted by monetary authorities. This is relevant because to the extent that volatility in terms of this currency or currencies becomes a policy objective, it becomes a constraint on monetary policy. In most cases the reference currency is trivial to define, but in other cases it is not. For example, does the Swiss National Bank look to the euro or to the dollar when thinking of its monetary policy?
- *Monetary unions*: What is the exchange rate regime of countries that belong to a monetary union like the eurozone that floats relative to the rest of the world? Is it to be

considered a pegged or a floating regime? This is an unsettled issue, which casts doubts on the validity of the classifications and the related results for advanced countries after the launch of the euro.

- *Black markets and official exchange rates, which one should be used?* While the presence of parallel markets casts doubt on the relevance of the official rate, it may not be more informative than the latter for the purpose of a regime classification. For example, in the presence of massive intervention to contain the movement of the official exchange rate, one could argue that the peg imposes a binding (albeit insufficient) policy constraint and that the regime should not be classified as a float. At any rate, this issue has become a moot point with the declining importance of parallel markets in recent times.
- *Nontraditional forms of intervention:* Interventions in the exchange market by fiscal authorities, use of derivatives such as currency swaps or forwards, even verbal intervention, are all tools increasingly used in the past years that have mostly been ignored in available classifications due to data availability and comparability.

2.2 Exchange rate policy trends in the post-Bretton Woods era: A casual glance at the distribution of regime choices

Classifying exchange rate arrangements is relevant not only to be able to analyze whether and how different regimes have affected economic performance, but also to assess the trends in exchange rate choice and how they relate to global and country-specific contexts. The literature has identified or predicted two main trends in the way countries choose their exchange rate policies. First, the fairly established view that countries have systematically moved away from the US dollar pegs since the demise of Bretton Woods in the early 1970s, in favor of more flexible regimes. Second, the so-called “bipolar view,” in vogue during the mid-1990s, which suggested that intermediate regimes (including conventional pegs) would tend to disappear with financial development and integration, as large swings in cross-border flows and increasingly liquid domestic capital markets would make them vulnerable to speculative currency attacks.¹⁰ As a result, it was argued that countries would (or, at least, should) move either to more flexible regimes with no exchange rate precommitment (including dirty floats) or, when this was not an option, to superfixed regimes with no margin for monetary policy (the “hard pegs” which typically groups currency board agreements and regimes with no national legal tender). How have these hypothesis fared, based on the record of actual regime choices in the last 30 years?

Figure 3 illustrates the prevalence of specific regimes over time and around the world. The graph shows both the IMF *de jure* classification as well as the Levy Yeyati and Sturzenegger (2007) and the Reinhart and Rogoff (2004) *de facto* alternatives. The three show somewhat different paths. The IMF classification displays a clear trend toward floating regimes that accelerates in the 1990s and somewhat reverses in recent

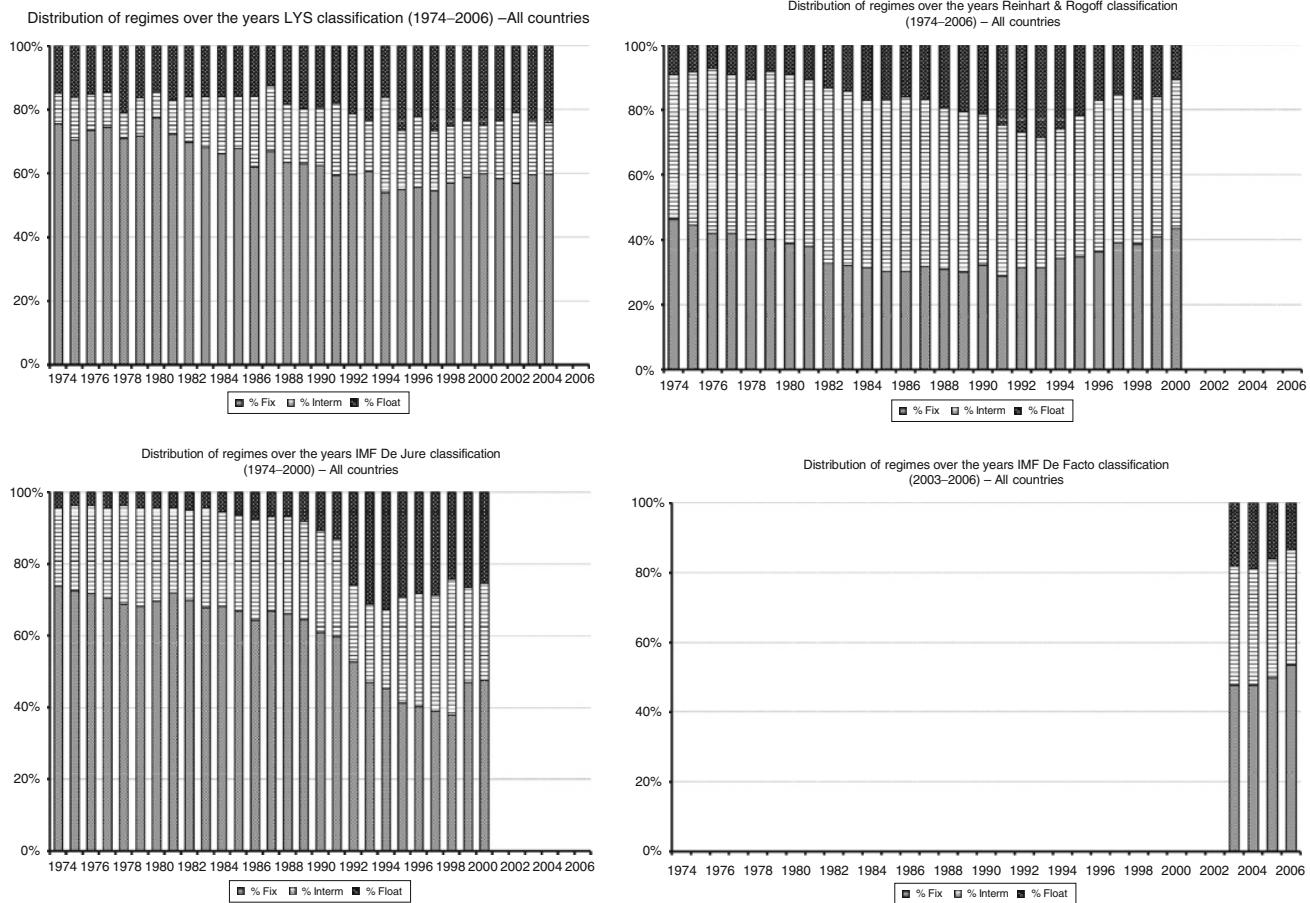


Figure 3 IMF, RR, LYS exchange rate trends. Source: Levy Yeyati and Sturzenegger (2008), IMF and Reinhart and Rogoff (2004).

years, a pattern that is replicated in the Reinhart and Rogoff classification. In the LYS classification, on the contrary, the trend is much less pronounced and the regime choices more stable. At any rate, while these trends were heralded as the triumph of floating regimes, in practice most countries still opt for fixed exchange rate arrangements, with the number of fixers oscillating between 40% and 60% depending on the classification. More striking is the recent reversal. Since 1998, the share of pegs increased for all classifications, while floating regimes declined from a 26% participation in 2000 to less than 10% in 2006 according to the IMF's *de facto* classification. Similarly, according to LYS, the share of nonfloats (intermediates, conventional, and hard pegs) represents 75% of the sample, exactly the same share as in 2000.

This broad distribution masks important differences across groups of countries. For example, according to LYS, Latin American countries seem to have embraced floating arrangements wholeheartedly (mostly, in combination with inflation targeting regimes), with the amount of *de facto* floats doubling between 2000 and 2004 at the expense of both intermediate and pegged regimes, whereas emerging Asia has preserved its bias toward more rigid arrangements. Interestingly, this evidence is *a priori* at odds with the bipolar view, since currency mismatches in Latin America have been large, and certainly larger than in Asia.¹¹

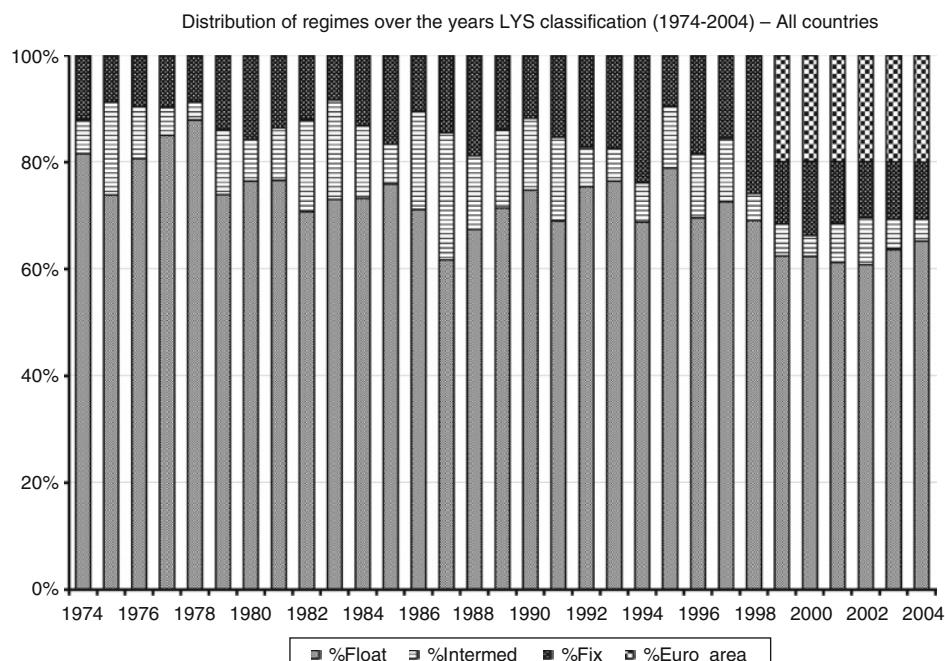


Figure 4 Classification weighted by market size. Source: Levy Yeyati and Sturzenegger (2008)

A somewhat different story is obtained when countries are weighted by their economic size. Because large economies tend to float, floating arrangements appear to prevail. Figure 4 shows that while most countries still choose to fix, most economic activity is conducted under floating regimes. The euro zone represents a peculiar case of a floating common currency, and is identified separately in the chart. If the euro zone is classified as a peg (as it usually is), the latter would display a slight jump in the new millennium; if it is classified as a float, the jump would favor the float group, which by 2004 would represent 80% of the world economy.

2.3 Inside the nonfloat group

As shown in Table 1, the nonfloat group in itself includes a wide array of alternatives. Nine countries that have chosen to adopt the currency of another state provide an extreme version of pegging, typically called “dollarization.”¹² Less extreme are monetary unions where several countries share their currencies. There are four such areas, the ECCU, WAEMU, CAEMC, and the EMU including a total of 42 countries. While the first three peg the common currency to another currency (the dollar in the case of the ECCU; the French franc and, later, the euro in the case of the WAEMU and CAEMC) and, as a result, can easily be classified as fixed regimes, the euro zone has, as noted, an ambiguous status, as it has established a currency union among its members (eliminating exchange rate flexibility at that level) that floats fully vis-a-vis the rest of the world.

A slightly weaker commitment to the peg—although still considered a “hard” peg is the currency board agreement that entails a legal obligation to keep (almost) full backing of monetary liabilities with liquid reserves, which in principle eliminates any margin for monetary policy.¹³ This regime, which was popular in the 1990s, when it received the blessing of the IMF (see, e.g., Enoch & Gulde, 1998, or Balino & Enoch, 1997) presently is in place in seven countries. Finally, we have the so-called “conventional” pegs to a currency (or currency basket) without additional legal constraints.

Traditionally, standard classifications have characterized exchange rate rigidities from a symmetric perspective, that is, focusing on exchange rate and reserve volatility without distinguishing between interventions to avoid a depreciation from those intended to avoid an appreciation. Underlying this focus is the Mundellian framework in which these rigidities amplify real shocks, both positive and negative. But the direction of intervention is not irrelevant, for at least two reasons. First, the price rigidities that introduce a role for exchange rate adjustments are generally asymmetric as well: prices tend to adjust upwards much more easily than downwards. Second, the motivation of intervention (and possibly its effects) differs with its direction: the prevention of

a depreciation may be geared to avoid financial distress or high inflation; the prevention of an appreciation may result from the target of an undervalued currency to gain competitiveness or reduce the odds of a traumatic depreciation in the future (as discussed in detail below).

This bias has been common in the literature. The RR or Shambaugh classifications classify countries on the basis of exchange rate volatility, while LYS looks at the *absolute* value of interventions, ignoring their sign. So do Calvo and Reinhart (2002), who examine three intervention variables: the *absolute value* of changes in exchange rate, in reserves, and in monetary aggregates.¹⁴ They find that emerging countries (as well as some industrial ones like Canada) intervene much more heavily than the prototypic float, and attribute this to *fear of floating*.

While academics have characterized regimes on the basis of symmetric measures, the motivation and actions of policy makers is clearly asymmetric. This leads to inconsistencies when justifying some of the empirical regularities. For example, Calvo and Reinhart (2002) argue that fear of floating responds to devaluation fears within economies with financial dollarization (FD) and a high pass-through (which tends to be associated with FD), in the context in which a realignment of the exchange rate may lead to massive balance sheet losses of currency mismatched debtors, and to high inflation. But this leads to a reaction function that is more responsive to depreciation pressures than to appreciation pressures—which according to this story entails no obvious policy concern.¹⁵ In other words, the motivation for exchange rate policy—unlike the methodology used to characterize it—is clearly one-sided.

On the other hand, recent years have witnessed the increasing popularity of an alternative motivation for intervention connected to the traditional “mercantilist” objective of preserving international competitiveness through a depreciated exchange rate or, more generally, protecting growing economies from the adverse effect of an appreciating currency.¹⁶ Again, the purpose here is clearly asymmetric: only changes in one direction are a source of concern.

Both the fear of floating and the mercantilist stories call for a regime grouping that unveils this asymmetric effects and policy responses. Based on Levy Yeyati and Sturzenegger’s (2007) extension of the LYS classification, Figure 5 shows that these asymmetries have evolved over the years: the share of intermediate regimes (alternatively, intermediates and pegs) that intervene purchasing reserves has changed dramatically (and predictably) over time. The debt crisis years found most developing countries selling foreign currency to defend their exchange rate anchors and to avoid sharp depreciations, whereas in recent years (with the unsurprising exception of 1998) countries have increasingly intervened in the opposite direction. The same story emerges when the interventions are detrended (to factor out the positive intervention that may be

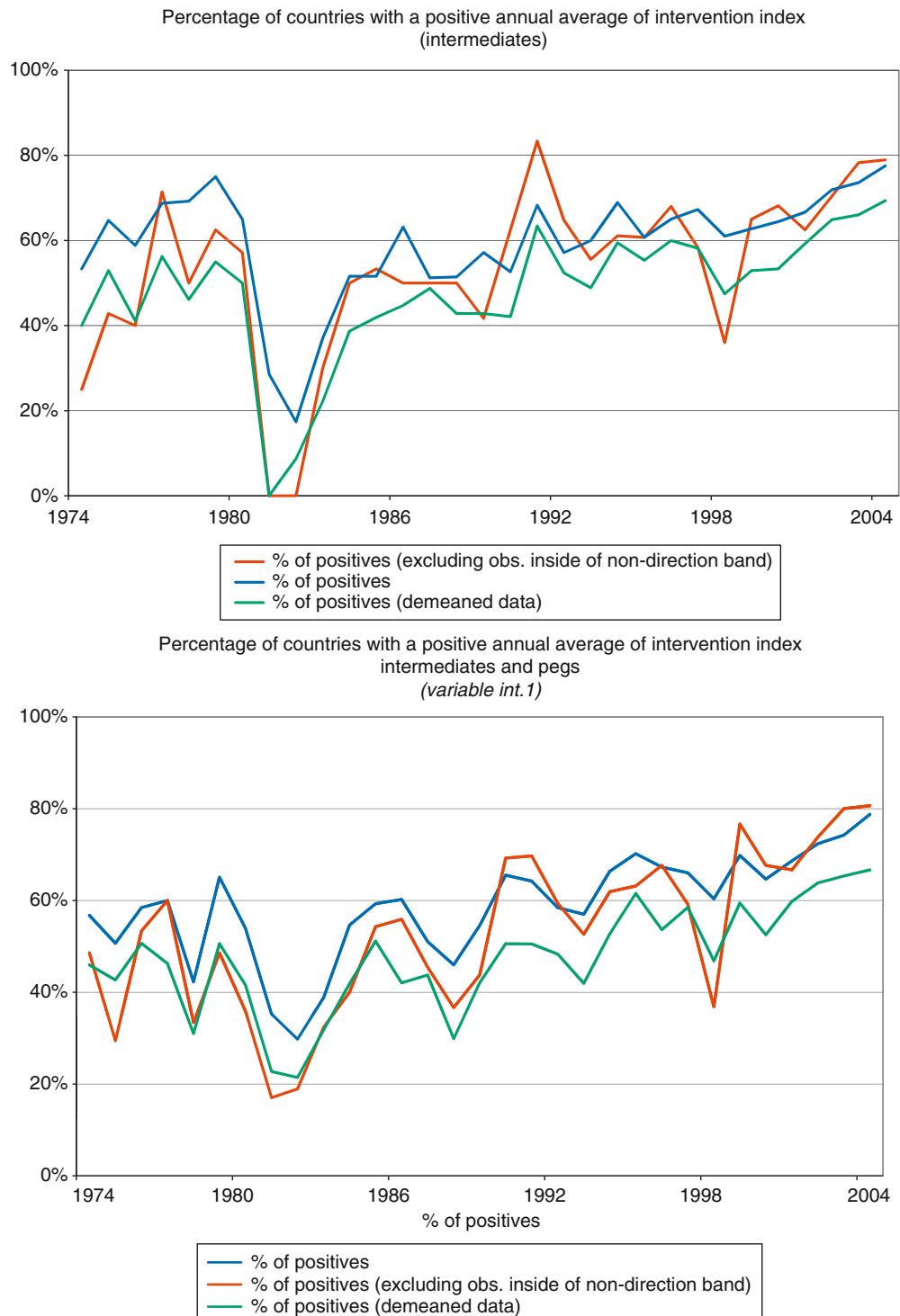


Figure 5 Direction of intervention. Source: Levy Yeyati and Sturzenegger (2007).

associated with the long-run growth of output and monetary aggregates) and when small interventions are filtered out (with the cutoff defined as 95% confidence interval of the distribution of interventions in benchmark floats Australia, Japan, and the US). As it turns out, conventional “defensive” pegs associated with *fear of floating* represented between 20% and 30% of the cases in 2004. This evidence suggests a comeback of proactive exchange rate policies, this time with a bias toward undervalued currencies (will be discussed below).

2.4 Monetary policy: From exchange rate anchors to inflation targeting

Since the early 1970s, when the rise of inflation led to increased skepticism on the role of monetary authorities, a significant body of literature has framed the debate over monetary policy in terms of the choice of the appropriate nominal anchor, motivated in part by the concepts of time inconsistency and inflation bias.¹⁷ Because it is widely accepted that there is no long-run tradeoff between inflation and output, the goal of the anchor is intended to reduce inflation expectations and, in turn, *ex post* incentives to validate them through monetary expansions, thus facilitating the conduct of monetary policy through the usual instruments (typically, monetary aggregates or the interest rate). As a result, monetary policy has been discussed as a tension between the credibility provided by an anchor, and the costs of the anchor in terms of a smaller degree of flexibility to respond to unanticipated shocks. Finding a credible anchor that imposes minimum constraints on the ability to react to shocks has been the dilemma at the core of the monetary policy debate to this day.¹⁸

It follows that the anchor of choice is a good starting point for a classification of monetary policies. Table 1 presents such a grouping, based on four mutually exclusive anchors: the exchange rate, the inflation rate, monetary aggregates, and other miscellaneous regimes. Sterne (1999) provides a more nuanced classification, to capture the overlaps usually observed in the choice of anchors: In Figure 6 we extend his exercise to show regimes during the 1990s and 2000s to illustrate how the weight of each anchor changed in the past decade. Casual inspection reveals a growing preference for explicit targets (in line with greater transparency in the conduct of monetary policy), and a declining incidence of monetary aggregates as the sole anchor (in line with an increasing preference for inflation targets).

Note, however, that the regimes depicted in Figure 6 correspond to *de jure* statements on the objective of monetary policy. Very much in parallel with our discussion on exchange rate policies, there remains the question about whether these policies are implemented in reality—or, alternatively, about the nominal anchor used *de facto* by the monetary authorities. While an exchange rate anchor is easy to verify, identification is a more complex problem with alternative anchors. Mishkin (2007) describes the problems with measuring monetary aggregates:

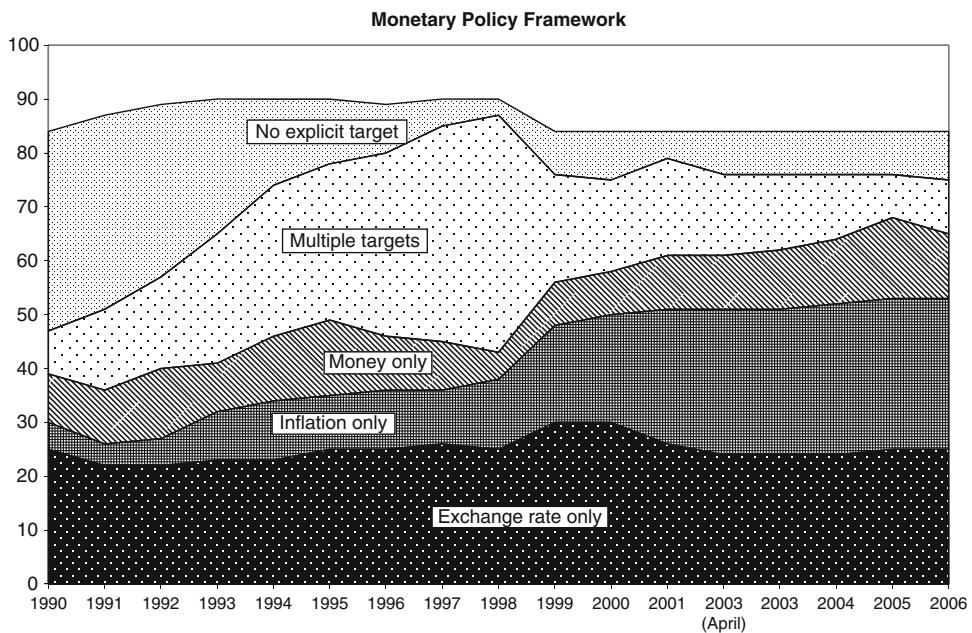


Figure 6 *De facto* monetary regimes. Source: Following Sterne (1999), author's calculations using IMF data.

Why did monetary targeting in the United States, Canada, and the United Kingdom during the late 1970s and the 1980s not prove successful in controlling inflation? There are two interpretations... One is that monetary targeting was not pursued seriously, so it never had a chance to succeed. The Federal Reserve, Bank of Canada, and particularly the Bank of England, engaged in substantial game playing in which they targeted multiple aggregates, allowed base drift (the initial starting point for the monetary target was allowed to shift up and down with realizations of the monetary aggregate), did not announce targets on a regular schedule, used artificial means to bring down the growth of a targeted aggregate, often overshoot their targets without reversing the overshoot later, and often obscured the reasons why deviations from the monetary targets occurred.

A similar obstacle appears with inflation targeting. As Mishkin and Schmidt Hebbel (2001) put it:

Classifying country cases into inflation targeting and other monetary regimes involves subjective choices for two reasons. First, there is lack of full agreement on the main conditions and features of inflation targeting and how they apply during transition to low inflation... Second, some countries have used simultaneously inflation targets and other nominal anchors (the exchange rate and/or a monetary aggregate), particularly at their early years of inflation targeting.

In addition, inflation targeters differ significantly among themselves on many other dimensions: target price index, target width, target horizon, escape clauses, accountability of target misses, goal independence, and overall transparency and accountability rules.¹⁹

More generally, monetary policy comprises so many dimensions to take into account that any characterization of monetary policy remains exceedingly difficult and always controversial.²⁰ The problem is compounded in the developing world by the fact that, in most countries, the exchange rate is bound to play an important role even in the absence of an exchange rate target, particularly in inflationary contexts associated with high exchange rate pass-through due to dollar pricing, or dollarized financial sectors where exchange rate fluctuations may be contractionary rather than expansionary.

Eventually, how can monetary policies in developing countries be characterized in an empirically useful way? Unlike exchange rate regimes, monetary policy cannot be easily identified by a few summary variables. If anything, *de facto* policies can be typified only by an analysis of the reaction function of the central bank. Hence, it is not surprising that no standard *de facto* classification has yet appeared (Box 2).

In spite of all the measurement drawbacks, Mishkin (2007) argues that there are six emerging consensus views in terms of monetary policy (i) that there is no long-run

Box 2 Estimating the Reaction Function

In their analysis of monetary rules, Bryant, Hooper, and Mann (1993) concluded that Central Bank's policy rules typically conformed to the "stated dual objective . . . to achieve sustainable growth in real activity while avoiding inflation" (p. 225). In recent years, there has been an active literature trying to estimate the policy reaction function of Central Banks, following Taylor's (1993) innovative description of a simple rule by which interest rates were adjusted in response to inflation changes and the output gap. Taylor suggested that the following simple equation represented US policy fairly well:²¹

$$i_t - \pi_t = r^* + 0.5(\pi_t - \pi^*) + 0.5(\ln Y_t - \ln \bar{Y}_t),$$

where $i(r^*)$ is the (real) interest rate, $\pi(\pi^*)$ is the inflation (target), and the last parenthesis represents the output gap. Orphanides (2001a, 2001b) criticized this rule on the basis that the information used is unavailable to policy makers at the time of decision-making and proposed a rule based on the available information set. Clarida, Gali, and Gertler (2000) argued that the Taylor rule has more to do with expectation of inflation and the output gap, and used an IV GMM procedure to estimate it, instrumenting future values of inflation and output on current and lag information. But do Taylor rules depend exclusively on the inflation rate and output as suggested by Taylor or do they take into consideration other variables? The exchange rate has become a usual argument in modern Taylor rules: Lubik and Shorfheide (2007) used Bayesian techniques to estimate the Taylor rules for four countries: the UK, Australia, NZ, and Canada, and found that the UK and Canadian monetary authorities do care about nominal exchange rates. Ball (1999) found that the inclusion of the exchange rate not only was relevant for small open economies but also improved the estimation for the US, while Taylor (2000) argued that exchange rates should be included in the estimation of monetary policy rules for emerging economies.²²

Far from being contradictory with the premise of exchange rate flexibility under inflation targeting, these findings highlight the incidence of exchange rate movements on output and inflation, and the need to implement countervailing policy action to keep inflation within target. They also illustrated the difficulties in producing a usable classification of monetary policy over time and across countries.

tradeoff between output (employment) and inflation; (ii) that expectations are critical to monetary policy outcomes; (iii) that inflation is costly; (iv) that monetary policy is subject to the time-inconsistency problem; (v) that central bank independence improves its efficacy; and (vi) that a strong nominal anchor is key to producing good monetary policy outcomes.

This consensus has been reflected in some visible trends, particularly in middle-income countries with more developed financial sectors. *Pari passu* with the decline in the preference for official commitments with exchange rate targets, recent years have witnessed a growing preference for targeting the inflation rate directly. It is only natural that, as many countries became increasingly dedollarized, financial stability considerations became less relevant and fluctuations in the exchange rate became less correlated with the inflation rate. The benefits of the exchange rate anchor declined accordingly, paving the way to what some observers regard as a new Floating and Inflation Targeting (FIT) paradigm. By 2006, 25 developed and middle-income developing countries officially ran inflation targeting regimes and claimed to sustain freely floating exchange rates.²³

However, the manifestation of FIT in the developing world is still far from the homogeneity implicit in the term paradigm. Varieties of inflation targeting in a developing economy may differ from that of an industrial country. In developing economies with important pass-through or balance sheet concerns, one would expect the central bank to react to exchange rate fluctuations (either through interest rates adjustments or outright intervention) even in the absence of an exchange rate target. Moreover, in some cases, two regimes may coexist: a FIT (or, more generally, a flexible regime with autonomous monetary policy) that tolerates moderate exchange rate movements, together with a *de facto* peg activated by substantial exchange rate realignments (see Box 3).²⁴ Even if the FIT paradigm ultimately prevails, a policy of benign neglect of the exchange rate may be difficult to conceive at the current stage; any characterization of monetary policy in the developing world should take this aspect into account.

Box 3 A Minimalist FIT Model for a Developing Economy

Consider the following reduced model of a small open economy under IT, based on the backward-looking framework in Ball (1999), which combines a dynamic IS curve with a Phillips curve:

$$\begin{aligned} \text{(IS)} \quad & \gamma_t = -\beta r_{t-1} + \delta e_{t-1} + \lambda \gamma_{t-1} + e_t, \\ \text{(PC)} \quad & \pi_t = \pi_{t-1} + \alpha \gamma_{t-1} + \gamma(e_{t-1} - e_{t-2}) + \mu_t, \end{aligned}$$

where r is the real interest rate, e the (log) real exchange rate, γ the (log) output gap, and π the inflation.

To solve this model, we update (PC) two periods and impose an inflation target (which, without loss of generality, we can assume equal to zero), to obtain

Continued

Box 3 A Minimalist FIT Model for a Developing Economy—cont'd

$$(IT) 0 = E_t \pi_{t+1} + \alpha E_t y_{t+1} + \gamma E_t (e_{t+1} - e_t) + \mu_t.$$

Next, we update (IS) and (PC) one period:

$$\begin{aligned} (IS1) E_t y_{t+1} &= -\beta r_t + \delta e_t + \lambda y_t, \\ (PC1) E_t \pi_{t+1} &= \pi_t + \alpha y_t + \gamma (e_t - e_{t-1}). \end{aligned}$$

Finally, substituting (IS1) and (PC1) into (IT) and rearranging, we have the following equation (where the left hand side is referred to as the Monetary Condition Indicator, or MCI):

$$(MCI) \alpha \beta r_t - (\gamma + \alpha \delta) e_t - \gamma E_t (e_{t+1} - e_t) = [\pi_t + \alpha(1 + \lambda) y_t - \gamma e_{t-1}] + \mu_t.$$

The first, trivial thing to note here is that a change in the nominal exchange rate e demands a compensating change in r_t . In other words, monetary policy under IT cannot neglect exchange rate fluctuations. The reaction function and the direction of the policy response, however, depends on a number of factors: the interest rate effect through domestic absorption ($\alpha \beta$), the pass-through of the exchange rate change to domestic prices γ , the effect of a depreciation on domestic demand, δ , and the link between the interest rate and the exchange rate, the equation needed to close the model.

For example, assuming uncovered interest rate parity, $E_t (e_{t+1} - e_t) = r_t r_t^f$ (where r^f is the international interest rate) implies that, in general, exchange rate changes would elicit a countervailing interest rate move in the opposite direction, as (IT) becomes:

$$(MCI1) r_t - \omega e_t = \frac{\pi_t + \alpha(1 + \lambda) y_t - \gamma e_{t-1} - \gamma r_t^f}{\alpha \beta - \gamma},$$

where $\omega = \gamma + \alpha \delta / (\alpha \beta - \gamma)$ which for very low pass-through $\omega \approx \delta / \beta$ would be roughly equal to the tradable share of GDP.

However, interest rate increases that raise the exchange rate may be “inflationary” if the pass-through coefficient is too large ($\gamma > \alpha \beta$). Similarly, contractionary devaluations ($\delta < 0$) that may arise, for example, due to balance sheet effects in financially dollarized economies, may call for lower interest rates if $\delta < -\gamma / \alpha$. Finally, when the foreign exchange market is under speculative pressure, lowering interest rates would reduce the cost of shorting the domestic currency and fuel a run. In these cases, the authorities may choose to intervene directly in the forex market.

3. WHY DO WE TALK ABOUT MERP?

Having characterized the set of MERP, we now turn to the implications of the different arrangements. To organize our discussion of the vast literature on the effects of MERP that will allow our policy discussion, it is useful to distinguish the multiple direct and indirect effects identified in the related literature, and their various interactions. At the risk of being excessively schematic, we can distinguish the following: (i) direct effects on some of the policy objectives mentioned in the introduction: For example, from exchange rate anchors to inflation rates, or through the effect of exchange rate flexibility on the output response to real shocks and, in turn, on output

volatility; and (ii) the links between MERP and “intermediate variables” that are not the policy objectives themselves but that have been portrayed in the literature as having an effect on some of them. For example, the link between exchange rate stability and trade, where the latter—has been argued—may affect growth.²⁵ Figure 7 summarizes the different nexus to be explored in this section.²⁶

On the left-hand side, we identify relevant exogenous shocks (real, such as changes in terms of trade (σ_{TOT}); or financial, such as changes in global liquidity (σ_{i^*}) or global risk aversion (σ_p) that drive cross-border flows and the country’s cost of capital). On the right-hand side, we have the four policy objectives: output growth (g_y), (low) output volatility (σ_y), (low) inflation (π) and equity. In the middle, we have the choice of MERP, which affects policy objectives directly (modifying the impact of exogenous shocks on policy objectives), and indirectly (affecting intermediate variables that may, in turn, have significant consequences for some of the policy objectives).²⁷ The rest of this section surveys the relevant contributions to the study of each of these channels.

3.1 The link between exchange rates and growth revisited

The first variable that comes to mind when talking about development and, more generally, macroeconomic performance is real per output growth. And there is, indeed, a body of work that has examined the direct link between MERP and growth from an empirical perspective. A point to clarify regarding this discussion is that, whereas historically it has

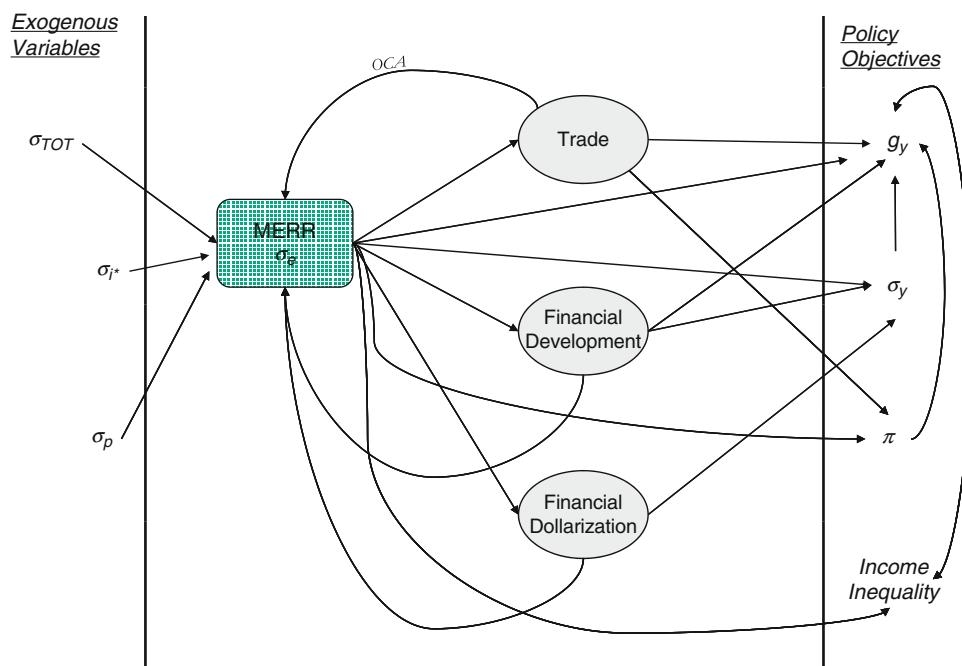


Figure 7 The big picture: direct and indirect links.

been the effect of the exchange rate level that has been at the center of the development policy debate, much of the recent empirical literature has dealt with implications for growth of exchange rate volatility (or policy regime). We review both in turn.

Several hypotheses have been presented on why the regime may be related to growth. Some channels have to do with global factors and others with domestic ones. From a global perspective, fixed exchange rates were viewed as one of the important drivers behind the development of international financial markets at the end of the nineteenth century. Johnson (1956) provides an early defense:

The advantages of a single currency within a nation's frontiers are, broadly, that it simplifies the profit-maximizing computations of producers and traders, facilitates competition among producers located in different parts of the country, and promotes the integration of the economy into a connected series of markets, these markets including both the markets for products and the markets for the factors of production (capital and labor). The argument for fixed exchange rates, by analogy, is that they will similarly encourage the integration of the national markets that compose the world economy into an international network of connected markets, with similarly beneficial effects on economic efficiency and growth.

Later on, the Mundellian paradigm shifted the attention to domestic factors by focusing on the shock absorber role of exchange rate, and the finding that fixed regimes tend to magnify real shocks. This, in turn, to the extent that volatility deters long run growth, implies that fixed regimes are likely to deliver a weaker economic performance. Gavin and Hausmann (1996), Ramey and Ramey (1995), Aizenman and Marion (1999), and Caballero (2000), among others, provide evidence on the link between higher volatility and lower growth.

Others have suggested that fixed exchange rates tend to create exchange rate misalignments that lead to speculative attacks and sharp crises resulting over the years in lower growth performance: here, the growth effect comes from a higher propensity to suffer an economically costly crisis event. Aizenman and Glick (2005) and Kuttner and Posen (2001) have both found that the harder and longer the peg, the larger are the depreciations upon exiting.

A somewhat related story is offered by Hausmann and Rigobon (2003), who argue that the volatility of exchange rates may induce an under-specialization in tradables that hurts growth performance. The argument is that volatile real exchange rates makes production in the tradable sector more risky relative to nontradables (because, if investment declines as a result of a negative shock, the price of nontradables increases, partially offsetting the effect in that sector). However, the argument hinges on the yet-to-be-tested assumption that growth opportunities are concentrated in the tradable sector, something that needs to be proven.

More generally, as can be viewed in a succinct review of the empirical literature in Table 3, while the exchange rate policies are often found not to be significant for

Table 3 Some results on the effects of exchange rate regimes (in chronological order)

Study	Estimation method	Sample period	Key results
Ghosh et al. (1997)	OLS panel data, two-stage instrumental variables	1960–1990	GDP growth was not affected by (<i>de jure</i> or <i>de facto</i>) regimes
Levy Yeyati and Sturzenegger (2001)	two-stage instrumental variables	1974–1999	No significant links for developed economies. For developing economies, pegs associated with slower growth.
Bailliu et al. (2003)	GMM panel data	1973–1998	Pegged regimes grow by about one-half percentage point faster than floats and about one percentage point faster than intermediate regimes. Regimes with anchors and pegged regimes grew faster than regimes (floats and intermediates) without anchors
Levy Yeyati and Sturzenegger (2003)	two-stage instrumental variables	1974–2000	For aggregate of all economies, growth for intermediate regimes and pegged regimes was about 1 percentage point and 0.8 percentage point, respectively, lower than under floating. Controlling for endogeneity, growth under pegs was about 2½ percentage points below floating, while for intermediate regimes there was little difference from floating. Application of separate regressions to industrial and developing countries showed little impact of regime for former group, while for developing countries less-flexible regimes were associated with slower growth

Continued

Table 3 Some results on the effects of exchange rate regimes (in chronological order)—Cont'd

Study	Estimation method	Sample period	Key results
Ghosh et al. (2003)	OLS panel data, fixed effects, two-stage instrumental variables	1970–1990	Compared with floats, growth was about 3.3 percentage points higher under intermediate regimes and 2.5 percentage points higher under pegs. Attributed this result to the fact that their coding tends to drop floats with stable exchange rates
Rogoff et al. (2005)	OLS panel data, fixed effects	1970–1999	For developing economies, real growth appears to decline with increased flexibility; for emerging markets, no evidence of a link between regimes and growth is found. For advanced economies, growth rose with increased flexibility
Dubas et al. (2005)	Random effects	1971–2002	For all countries, pegged regimes grew a bit more than one percentage point relative to floats. The difference between floats and intermediate regimes was not statistically significant. For industrial countries, regime dummies were not significant. For nonindustrialized economies, pegs grew 1.3 percentage points more than floats, but there was no statistically significant difference between floats and intermediates
De Grauwe and Schnabl (2005)	For inflation: GMM panel data for growth: GLS	1994–2004	For <i>de jure</i> regimes, no clear association with growth. Using <i>de facto</i> coding, pegged rates were associated with higher growth
Aghion et al. (2006)	GMM panel data	1970–1999	Pegs are associated with slower growth for not financially developed countries

industrial countries, there is no basic agreement in the case of developing economies. Levy Yeyati and Sturzenegger (2001, 2003) found that floating leads to higher growth, while Rogoff et al. (2005) found that this result applies only to advanced economies. Later works have found results supporting one or the other.

Why are these results so contradictory? One could think of several reasons linked with the regime classification procedure. First, regimes are endogenous: for example, peg failures are often recorded as intermediates or floats; more generally, most classifications do not control for crisis episodes in which the behavior of exchange rates and reserves cease to reflect a regime choice. Second, as noted, regime flexibility is usually measured as exchange rate volatility, which leads to an association with bad economic outcomes (rigid regimes under attack are often coded as floats; stable floats are often dropped or coded as intermediate or pegged regimes).²⁸ Third, information on intervention variables is seldom complete: Even in classifications that control for policy intervention, the focus on reserves fails to capture other intervention mechanisms such as interest rates, currency derivatives, or capital controls.

Rather than endorsing the skeptical view that no general conclusion can be drawn from existing studies, we believe that the emphasis needs to be placed on a critical methodological drawback faced by the exchange rate policy agenda: the limitations of reduced-form tests that conflate a variety of channels into one linear relation between MERP and economic performance. This contrasts with an analytical literature that reveals a complexity of specific, sometimes countervailing channels that renders the finding of a significant link between long-run growth and exchange rate policy an uncertain empirical endeavor.²⁹

As noted, the analytical economic literature has emphasized as much the link between regimes (exchange rate flexibility) and growth as the one between the latter and the *level* of the exchange rate. In fact, a recent body of work has recovered an old theme: the use of undervalued exchange rates to stimulate economic growth. Eichengreen (2006a) reviewed the argument and argued that the undervalued exchange rates implemented by the Bretton Woods agreement were a key driver of Europe's recovery in the postwar period. Ohkawa and Rosovksy (1973) and Eichengreen (2006a) made the point for Japan's post-WWII recovery.

This mercantilist view that exchange rate policy or, more precisely, a temporarily undervalued currency could be used to protect infant industries as a development strategy has recently enjoyed a minor revival.

Empirical evidence on the relation between the level of exchange rate and growth has been reported in a number of recent studies. Hausmann et al. (2005) found that depreciated real exchange rates (as well as trade growth) are important components of growth accelerations; conversely, Johnson, Ostry, and Subramanian (2006) showed that persistent overvaluations tend to be associated with poorer growth.³⁰

Moreover, under and overvaluation have been invoked to explain the "Dutch disease" effect of foreign aid (Rajan & Subramanian, 2005), the disappointing growth

dividends of financial integration (Prasad, Rajan, & Subramanian, 2006), or the positive correlation between intervention (reserve accumulation) and investment and growth (Levy Yeyati & Sturzenegger, 2007). However, these neo-mercantilist views supporting the growth effects of undervalued currencies have been saluted, at best, with skepticism, probably due to the disbelief in the relationship between nominal variables and growth mentioned in the introduction.³¹

3.2 ERR and output volatility

The relation between the exchange rate regime and output volatility is also a channel with a long tradition in international finance, and one of the key links underlying the debate on optimal currency areas. It involves understanding the role played by the exchange rate as shock absorbers: under floating exchange rates, the economy has a greater ability to adjust to “real” external shocks whereas fixed exchange rates have a larger ability to absorb “nominal” shocks (Box 4).³²

Box 4 Exchange Rates, Volatility and the Nature of the External Shocks

Calvo (1999) provides provided a minimal framework to understand the mechanics. Imagine a simple demand-determined output equation (this could be interpreted as the traditional “IS” curve)

$$y = \alpha e + u, \quad (1)$$

where y is the output, e is the exchange rate, α a parameter, and u a random shock; and a money demand equation (which could stand for the traditional “LM” curve)

$$m = y + v. \quad (2)$$

Here m is the stock of nominal money and v a liquidity shock. Consider two polar cases: fixed exchange rates where e is constant and y and m endogenous, and a floating regime where m is exogenous and e and y are endogenous. In the first case, output is determined by (1) and, in the second, by (2). If so, under fixed rates

$$\sigma_y^2 = \sigma_u^2 \quad (3)$$

and

$$\sigma_e^2 = 0,$$

whereas under float

$$\sigma_y^2 = \sigma_v^2 \quad (4)$$

and

$$\sigma_e^2 = \frac{1}{\alpha^2} (\sigma_u^2 + \sigma_v^2 + 2\rho\sigma_u\sigma_v).$$

Box 4 Exchange Rates, Volatility and the Nature of the External Shocks—cont'd

Clearly, from Eqs. (3) and (4), the regime that delivers the lowest volatility depends on the nature of shock. According to Eq. (3), fixed exchange rates deliver a larger volatility the larger real shocks, whereas according to Eq. (4) deliver a larger volatility of output the larger nominal shocks. This suggests that countries with large real shocks would be better off by choosing a float; countries with large nominal shocks would prefer to fix.

Testing the output response of output in the presence of fixed and flexible exchange rates with attendance to focus on the different types of shocks has received some attention in recent years. Empirically, the standard test examines whether a more flexible regime attenuates the output response to shocks: if nominal prices are (downward) inflexible, the output response to (negative) real shocks should be more muted under floating regimes.

Edwards and Levy Yeyati (2005) analyzed the effect of terms of trade shocks on economic performance under alternative exchange rate regimes. They estimated a two-equation model, one with the equilibrium growth rate and another explaining deviations from trend growth, and found evidence that terms of trade shocks get amplified in countries that have more rigid exchange rate regimes. They also confirmed that the response to terms of trade shocks is asymmetric: the output response is larger for negative than for positive shocks. Broda (2001) tackled the same question, using a VAR model to compute the way in which terms of trade shocks affect growth, and found that the effect of a 10% change in the terms of trade has a greater influence on growth under fixed than under flexible exchange rate arrangements. Ramcharan (2005) looked at the problem by exploiting the randomness of natural disasters. His evidence supports the idea that adverse natural shocks are associated with higher investment and FDI in countries with fixed regimes, but that the recoveries appear to be faster under floating regimes. His results combine two effects: the stability dividend of pegs in otherwise volatile countries and the benefits of greater exchange rate flexibility in the event of an adverse exogenous shock reported by Edwards-Levy Yeyati (2005) and Broda (2001, 2003).

3.3 MERP on price stability

The use (and benefits) of exchange rate anchors have typically been associated with what could be broadly referred to as a “deficit in monetary credibility,” which manifests in high inflation expectations, inflation inertia (backward indexation to past inflation), and a low impact of monetary policy announcements. Underlying this credibility story, there is a time inconsistency argument, by which high inflation expectations induce high inflation equilibria with steep nominal interest rates that, in turn, make it optimal for the government to dilute its debt burdens through inflation, generating an inflation bias.³³ In this case, the use of an exchange rate anchor may make dilution

more costly, to the extent that abandoning the anchor entails some (political or economic) reputation cost, playing the role of a partial commitment mechanism.^{34,35}

Exchange rate anchors present the additional advantage of coordinating expectations. In high inflation economies, it is not unusual to index prices partially to the exchange rate (typically, vis-a-vis the US dollar). Therefore, an exchange rate anchor could allow a quick transition from backward indexation to past inflation, to forward indexation to the announced exchange rate path. Canavan and Tommasi (1997) made this point. They explained the stronger link between an exchange rate anchor and expectations with a model that assumes that the public can monitor the nominal exchange rate more easily than it can the other variables. In their game of incomplete information with imperfect monitoring, they showed that serious stabilizers prefer more visible anchors, such as the nominal exchange rate even when fixed exchange rates have some costs, such as diminished capabilities to respond to external shocks.

From an empirical perspective, the literature has focused on the link between exchange rate regimes—and, in particular, varieties of exchange rate anchor such as (crawling) bands and pegs—and the inflation rate. Overall, there seems to be agreement on the fact that pegs are associated with lower inflation, even after controlling for money creation (e.g., by controlling for the presence of a peg in a standard monetary equation).³⁶ This suggests that the effect may work through the anchoring of expectations rather than through the imposition of monetary discipline. However, the direction of causality and, more importantly, the duration of the effect are more controversial.

Among the many qualifications raised by these studies, perhaps the most troubling is the well-known fact that failed pegs tend to collapse to floats, which in imperfectly specified tests may result in a spurious association between floating exchange rates and high inflation rates. Intuitively, in the long run, pegs not only may discipline monetary policy; but they are also endogenous to it, as they cannot be sustained in the face of persistently high inflation. This may explain why a closer inspection indicates that only long-lasting pegs are significantly linked to low inflation levels in the long run (Levy Yeyati & Sturzenegger, 2001). Ultimately, the effectiveness of an exchange rate anchor is always debatable, as it depends on the policy maker's ability to reign in the fiscal deficit and, if that is not fully achieved, his willingness and ability to refrain from monetary financing.³⁷

3.4 MERP on income distribution

There is not much of a debate on the relation between the monetary and exchange rate regimes and income distribution beyond perhaps the large literature on the adverse distributive consequences of inflation. To the extent that floating regimes are characterized by higher inflation rates, one could assign a regressive bias to more flexible regimes. However, this indirect connection seems a bit of a stretch and has seldom been made in the literature.

An alternative channel is the link between the *level* of the exchange rate and income distribution. The early reference is Kalecki's (1939) analysis of the effects of a devaluation in an open economy, according to which a depreciation would not necessarily increase aggregate demand because it would reduce the share of wages in output (and thus, the income of those with a larger propensity to consume). This point was later taken up by Díaz Alejandro (1965) who provided a careful analysis of the link of a depreciated real exchange rate with poverty and inequality. Diaz Alejandro's setup had in mind a country exporting food-biased commodities where landowners—the beneficiaries of the depreciated real exchange rate—had a large expenditure share of imported goods—in contrast to workers whose real wages fall with the real value of the local currency. In Diaz Alejandro's world, a depreciation redistributed income from workers to landowners, reducing aggregate demand and inducing a contractionary devaluation.

At a conceptual level, several caveats could be mentioned regarding Diaz Alejandro's original argument. First, it included a very restrictive class of beneficiaries—in modern societies the benefits of a real depreciation may be more widespread. Second, if an economy is subject to nominal wage rigidities, a devaluation may allow to soften this constraint leading to an expansion in employment with beneficial income distribution effects. Finally, in Diaz Alejandro's story resources transferred to the landowners/capitalists made their way to consumption abroad, rather than to the domestic financial sector that channels them toward investment activities at home—as assumed by Aghion, Bacchetta, Ranciere, and Rogoff (2006) where the extraordinary profits due to the depreciated currency increases the liquidity of financially constraint local firms, and enhances their access to finance, rendering expansionary devaluations.

At any rate, there has been relatively little empirical work testing this hypothesis. Edwards (1989) found that devaluations reduce the real wage with little impact on the labor share. Levy Yeyati and Sturzenegger (2007) revisited the issue and found that interventions aimed at depreciating the currency reduce the labor share of GDP, as well as unemployment, a channel, the authors argued, that would explain the benign growth impact.

3.5 Indirect links: The integration channel

There is vast body of work on MERPs (particularly, exchange rate regimes and exchange rate volatility) on economic integration (specifically, bilateral trade and, to a lesser extent, cross-border capital flows, including foreign direct investment). This literature is largely based on the well-known premise that the incidence on transaction costs of currency conversion (which includes not only the bid-ask spreads but also currency risk due to potential losses from exchange rate variations) plays the role of an implicit barrier for international transactions between countries using different currencies. The findings, mostly based on gravity models, point at a positive but small effect of exchange rate stability on trade.³⁸

Also, based on the gravity model is a stream of literature that became popular in the run up to the launch of the euro in the late 1990s, which focused on the trade effects of a monetary union, particularly since Rose (2000) argued that the average increase in bilateral trade due to the adoption of a common currency was of the order of 200%. These findings have since been greatly qualified for a number of reasons that included lack of representativeness (results were based on common currency pairs that include subnational entities with historical and political links with the issuer of the currency, and could hardly be extrapolated to real independent countries) and endogeneity (a common currency was more likely to be adopted in the presence of strong trade links).³⁹ At any rate, subsequent estimates by Rose himself placed the number between 100% (Glick & Rose, 2002), and 50% (the estimate for the trade effect of the euro reported by Rose and Van Wincoop (2001)).⁴⁰

Underlying these analyses was the Optimal Currency Area (OCA) precept that, in addition to removing transactions costs, a monetary treaty, by preventing competitive devaluations, fosters foreign direct investment and intraindustry trade. Interestingly, though, most of the related empirical literature is based on common currency pairs that do not belong to a monetary union, but rather would fit in the unilateral dollarization group. Moreover, more recent estimates using actual monetary union data (specifically, data from the European Monetary Union to measure the effect of the adoption of the euro) go as low as 5-10% (Micco, Stein, & Ordoñez, 2003) and 4% (De Nardis, De Santis, & Vicarelli, 2007), or even 0% (Berger & Nitsch, 2005).⁴¹

The influence of MERP on FDI flows is more scantily documented. There seems to be some effect from hard pegs (currency boards or the unilateral adoption of a common currency): For example, De Sousa and Lochard (2004), using data for EMU, find a link between the adoption of the euro and the increase in intra-EMU FDI flows.^{42,43}

In sum, while the order of magnitude of the integration channel is still under dispute, the existing evidence appear to support the view that, when measured in medium-sized countries, the effect is positive and small (for the trade channel, much smaller than originally argued), particularly in those cases in which integration has already been achieved by other means, such as trade agreements or investment treaties.

Covering all the implications of increased trade openness on developing policy variables would take us beyond the scope of this chapter. Let us just mention here the benign effect of trade on inflation, originally suggested by Rogoff (1985) and tested empirically by Romer (1993), who found that more open economies tend to have lower inflation rates due to the disciplining effect of international competition on domestic prices, particularly in the context of imperfectly competitive local markets.⁴⁴ Less conclusive results are offered by the literature documenting the impact of trade openness on output growth. A number of studies argue that trade has a positive

effect on growth (Ben-David, 1993; Frankel & Romer, 1999; Sachs & Warner, 1995) and productivity (Alcalá & Ciccone, 2004; Edwards, 1998). In turn, Frankel and Rose (2002) directly attempted to estimate the indirect growth effect through the trade channel: they combined the positive impact of a common currency on bilateral trade (which, in line with earlier gravity results, they estimate at a 100% increase in bilateral trade), with what they found to be a positive effect of trade openness on growth, to estimate an economically important impact of monetary integration on output growth.

However, the growth dividend from trade has been contested in recent work (Rodrik & Rodríguez, 2000; Rodrik et al., 2004). The main caveats raised by this literature are problems of mismeasurement and omitted variables. As Rodrik and Rodríguez (2000) put it, the indicators of openness used by researchers are poor measures of trade barriers or are highly correlated with other sources of bad economic performance.⁴⁵ In sum, whereas the conventional wisdom tends to view trade as beneficial for growth, the empirical evidence has been less supportive.⁴⁶

One area that has received increasing attention lately is the influence of financial and trade integration (often embedded in the broad term globalization) on output and consumption volatility. Here the profession appears to coincide: while industrial countries seem to benefit from financial integration, in nonindustrial countries the latter is associated with an increase in output and (particularly) consumption volatility (Kose, Prasad, & Terrones, 2003; O'Donnell, 2001).⁴⁷ The findings of these papers are more benign for the case of trade links, which are found to weaken the impact of macroeconomic volatility (Kose et al., 2003) and to increase the business cycle correlation between trade partners due to the propagation of demand shocks through external demand and intraindustry trade (Frankel & Rose, 1998).

In addition, recent studies by Calvo, Izquierdo, and Loo-Kung (2005) and Cavallo and Frenkel (2004) argue that trade openness reduces output volatility through lowering the probability of a financial crisis in financially dollarized economies, as the real exchange rate adjustments in the event of a capital account reversal is smaller, leading to a smaller balance sheet effect. This channel, it follows, should have weakened as FD declined in the 2000s.

3.6 Indirect links: The financial channel

MERP has been associated with financial development through two distinct channels. The first one is related to the consequences of exchange rate instability on cross-border flows, in turn, influenced by global trends in financial integration, which we briefly discussed above.

More recently, a second channel has come to the foreground in the context of financially integrated developing countries, namely, the implication of exchange rate regimes (most notably, pegs) on the degree of FD, where the latter is defined as the

use of a foreign currency to denominate financial assets and liabilities held by residents. In a nutshell, this literature points to four potential motives that may make a peg more conducive to the use of a foreign currency in financial transactions.

The first one, starting from the assumption that risk-averse resident investors choose their asset portfolio to optimize the real risk/return profile (in terms of the local consumption basket), argues that the dollar share of domestic savings and loans depend on the inflation risk of local currency assets (i.e., the volatility of the inflation rate) relative to the currency risk of dollar assets (i.e., the volatility of the real depreciation rate) (Ize & Levy Yeyati, 2003). If so, a mix of flexible exchange rates and low (and, as a result, less volatile) inflation should minimize the incentive to dollarize. By contrast, an exchange rate anchor that stabilizes the real exchange rate in a context of high and volatile inflation expectations would have the opposite effect. More generally, dollarization should mirror the exchange rate path-through coefficient; in the limit, full pass-through (alternatively, dollar indexation) virtually eliminates the currency risk of dollar assets, favoring dollarization.

A second motive is associated with the so-called “peso problem” (a large local currency interest rate premium due to persistent devaluation expectations), typically associated with imperfectly credible exchange rate anchors. In the presence of nonlinear liquidation costs, the currency composition of debt is optimally chosen to minimize the probability of default and liquidation. Thus, if the devaluation threat is large but unlikely, the borrower may opt for the less costly dollar funding: if peso interest rates are such that peso borrowers default in the absence of a devaluation that dilute their debts, the borrower may prefer to take his chances with currency risk (Jeanne, 2005).

A third explanation attributes the dollarization bias to the presence of externalities that generate the perception of implicit debtor guarantees (Burnside, Eichenbaum, & Rebelo, 2001). To the extent that the social cost of massive bankruptcies following the collapse of a peg makes a debtor bailout *ex post* optimal for the government, borrowers would anticipate this bailout, and underprice currency risk accordingly.⁴⁸

Finally, a fourth motive linking the two relates to financial regulation. A currency-blind regulation that fails to correct for the additional risk of dollar lending would underprice currency risk (at the expense of the regulator) and encourage the dollarization of the banking system. However, in the case of an exchange rate peg, a currency-specific regulation that flags the currency risk embedded in dollar intermediation (more precisely, the exchange rate-related credit risk of dollar lending to nondollar earners) would directly undermine the credibility of the peg. To the extent that pegs call for currency-blind regulation, they tend to facilitate (and implicitly subsidize) dollar intermediation.⁴⁹

The empirical evidence on regimes and FD is plagued by the lack of data (the currency composition of bank deposits is available for a broad group of countries only since the late 1990s; data on the currency composition of debt is even more scarce),

compounded by slow dynamics that limit the analysis beyond cross-country comparisons. With this caveat in mind, recent work has found support for the portfolio view (De Nicoló, Honohan, & Ize, 2005; Levy Yeyati, 2006), and for the presence of implicit guarantees in association with pegs, in the form of increased larger unhedged short currency positions at the firm level (Werner & Martinez, 2002).

A related literature documents the costs of nominal instability (moderate to high inflation, even if it is predictable) in terms of the demand for local assets and the deepening of local financial markets (Boyd, Levine, & Smith, 2001; Khan, Senhadji, & Smith, 2006). The same can be said about the incidence of exchange rate regimes (and exchange rate volatility) on the size of the financial sector. This has distinct implications, as financial development and stability have proved relevant for growth by improving access to enterprises (Levine, 2005) and fostering productivity gains rather than greater investment volume (Beck, Levine, & Loayza, 2000). Several channels through which this benign effect materializes have been documented: the promotion of start-ups and growth of firms (Klapper, Laeven, & Rajan, 2006; Laeven & Woodruff, 2007), greater innovation (Ayyagari, Demirgüç-Kunt & Maksimovic, 2006), a better use of investment opportunities (Love & Zicchino, 2006), and a more conducive environment for small firms (Aghion, Fally, & Scarpetta, 2006) that are often more financially constrained and, therefore, benefit proportionally more from a liquid financial market (Beck, Demirgüç-Kunt, Laeven, & Levine, 2005). Moreover, financial underdevelopment has been singled out as a factor behind the failure of low-income economies to converge, in income, to more advanced countries (Aghion, Howitt, & Mayer-Foulkes, 2005).

At the root of these arguments is the fact that financial constraints inhibit high-return entrepreneurial activity and limit firm growth, and that the factors underlying these constraints—insufficient loanable funds and high interest rates, lack of efficient collateral or costly and uncertain liquidation of the existing ones, burdensome bureaucracy or pervasive informality—are more prevalent in developing countries (Banerjee & Duflo, 2004). However, most of the evidence on finance and growth is based on aggregate cross-country studies and, given that finance tends to flourish with growth opportunities, the direction of causality of these links is not always unambiguous, particularly for low-income economies (Rioja & Valev 2004a, 2004b). Rajan and Zingales (1998) address this problem distinguishing firms with high need of financial markets from firms in sectors with lower need of financial markets. This categorization (they look at some financial ratios that signal strong use of financial markets in the US) is assumed to be exogenous. After ranking sectors by their degree of financial dependence, they show that financial development leads to faster growth of more financially dependent firms.

At any rate, from a conservative perspective, it would be safe to say that rather than finance being an independent growth driver, finance and growth feed into each other.

3.7 Summing up

This broad and often contradictory body of evidence, linking MERP with development objectives is made even more complex by the fact that the workings of some of these channels can change dramatically depending on specific country characteristics and global conditions. For example, *de facto* financial integration and capital mobility could foster exchange rate flexibility as they force the country to choose between a stable exchange rate and an autonomous monetary policy; however, if financial flows are denominated in the foreign currency, the concomitant increase in dollar liabilities may optimally inhibit exchange rate flexibility for fear of balance sheet losses in the event of a real depreciation. Similarly, whereas flexible exchange rates help buffer the economy against adverse external shocks, the same channel would be contractionary in heavily dollarized countries, which would be better off with more rigid exchange rate arrangements.

This multiplicity of effects has been present in the way countries actually choose their policies. The empirical literature has tested—often selectively—whether and how the arguments outlined above play a role in actual regime choices, and have found support for many of them. Table 4 summarizes the main findings in this body of work. The next section explores these policy choices in more depth.

4. THE MAKING OF A POLICY

All the channels previously described have been reflected in the policy debate (and in the actual implementation of exchange rate policy) over the years. As noted in the introduction, the policy question that we posed at the start incorporates both country- and time-specific aspects that tend to evolve (or, at least, vary) over time. Thus, as conditions in international financial markets and developing economies changed, the focus of the debate has shifted accordingly. Tracing the policy debate in the post-Bretton Woods clearly illustrates how the different intervening factors identified in the literature provide justification for different MERP. More importantly, it provides the broader perspective needed to go from the analytical arguments and the empirical results based on historical data, to policy decisions that need to factor in the current context and prognosis.

Keeping this in mind, a brief narrative of the debate, linking different conditions in international financial markets to different “trends” in the choice of regimes in the developing world, will be useful to set the stage to answer the more specific policy questions.

4.1 Exchange rate anchors in the 1980s

A casual review of the exchange rate debate in the late 1980s and early 1990s shows how the discussion hinged on the role of exchange rates and income policies as nominal anchors in a high inflation environment. A good reference point is (Bruno, Di

Table 4 The choice of exchange rate regimes

Choice of real exchange rate						
Papers	Number of countries	Methodology	Exchange rate classification	Ideas tested	Variables included	Results
Collins (1996)	26 (1978-1992) Latin America (LAC)	Probit	IMF	Exchange rate misalignment Political cost of debt Difficult to manage a flex Disciplinary effect of anchor	Past growth, inflation, size, openness, IMF program	Large and open economies, high inflation and larger current account deficit induce floating. Trend in favor of float
Edwards (1996)	63 (1980-1992)	Probit	IMF	Credibility vs. flexibility	Political instability variables, coefficient of variation of export growth, coefficient of variation of real exchange rate (1970-1982), interacted with openness, per capita GDP, inflation bias, past growth and reserves	Unstable countries, external volatility, inflation, high income countries and less reserve accumulation induce floating
Bayoumi and Eichengreen (1998)	21 (1963-1992) Industrial countries	Instrumental variable	Volatility of exchange rate and volatility intervention	Optimal Currency Area (OCA)	Variability of output, openness, size, dissimilarity of exports	Variables support theory

Continued

Table 4 The choice of exchange rate regimes—Cont'd

Choice of real exchange rate						
Rizzo (1998)	All developing countries w/data (1977-1995)	Probit binomial/ multinomial	IMF	OCA Fiscal pressure and inflation	Size, development level, diversification of trade, openness, debts, current account, deficit, reserves, inflation	OCA is confirmed except for openness, which is associated with floating. While other variables appear unstable, inflation is related to floating
Frieden, Ghezzi, and Stein (2000)	26 (1960-1994) LAC	Ordered logit	IMF	Macro variables Institutional factors Interest group factors Political factors	Inflation, hyperinflation, openness, reserves, terms of trade, volatility, capital controls, Central Bank independence, sectorial and political instability variables	Inflation is not significant, but hyperinflation leads to fixed regimes. More reserves, CB independence, political instability, less openness, volatility of terms of trade and weak governments lead to floating
Poirson (2001)	93 (1990-1996)	Ordered probit	IMF and $\text{FLT} = \sigma_e/\sigma_r$	Political institutions Currency mismatches OCA	Size, development level, export diversification, openness, vulnerability to external shocks, reserves, inflation, political instability variables, capital controls, concentration of trade, dollarization, ability to hedge	Large economies, capital mobility, political instability, exports diversification, external vulnerability, lower reserves, high ability to hedge all lead to floating. Both dollarization and temptation to inflate are related to <i>de jure</i> floats. Results of <i>de facto</i> regimes are weaker

Juhn and Mauro (2002)	All countries w/dta	Bivariate probit and multinomial logit	IMF Levy Yeyati and Sturzenegger (2001)	OCA Capital openness Macro variables Historical institutional variables	Openness size, concentration of trade, per capita GDP, volatility of terms of trade, capital controls, openness of the capital account, dummy for emerging countries, inflation, reserves, political instability variables.	Small and closed economies with no inflation exhibit floats, and other variables not robust.
Alesina and Wagner (2006)		Ordered logit	Reinhart and Rogoff (2004)	OCA Dollarization Institutional variables		Large and closed economies not dollarized, with weak institutions are more likely to float.
Von Hagen and Zhou (2006)	(1980-1987) developing countries	Dynamic random effect multinomial	IMF	OCA Stabilization consideration Currency crisis Political institutional factors	Openness, size, geographical concentration of trade, development level, financial development, inflation, real exchange rate volatility, monetary shocks, reserves, fiscal	Large and open economies with neither financial development nor volatility of real exchange rate are more likely to float. Political variables are unclear
Levy Yeyati and Sturzenegger (2007)	183 (1974-1999)	Pooled logit	IMF, LYS and RR	OCA Financial view Impossible trinity Balance sheet effects Credibility view	Size, openness, terms of trade shocks, capital openness, <i>de jure/de facto</i> financial development, dollarization, political institutions, inflation	OCA variables work. Impossible trinity applies to developed countries, but balance sheet issues appear relevant for developing countries. Political variables are relevant only for developing countries and suggest that weak governments float

Tella, Dornbusch, & Fischer, 1988), a book that brought together policy making experiences with inflation stabilization in emerging countries. The book index provides a summary view of the relevant issues at the time: a piece on Israel dealt with the modeling of the interaction of money, wages, prices; a chapter on Brazil addressed the effect of wage indexation and wage freezes; another one on Bolivia discussed the stabilizing role of the exchange rate in an economy with dollar pricing; a paper on Mexico asked whether an income policy-based program could control the ever increasing inflation in the country. Overall, the contents were an accurate reflection of the dominant role played by inflation concerns in the late 1980s.

The academic literature mirrored these concerns, assessing the merits of exchange rate-based stabilizations (ERBS) coupled with income policies, relative to the more traditional money-based stabilizations. Kiguel and Liviatan (1991, 1992) and Vegh (1992) documented that ERBS appeared to lead to an initial and temporary consumption boom that tended to end in a contraction, whereas money-based stabilizations often induced an initial recession followed by a boom. Calvo and Vegh (1993, 1994) provided a formalization: in their model, a one-shot credible stabilization tended to have the same result regardless of the anchor of choice, but transitory or not perfectly credible exchange ERBS lowered interest rates in the short run, fueling a consumption and output boom (and a trade deficit) in the short run that were reversed once the program collapsed. On the other hand, noncredible money-based stabilizations were expected to increase the demand for money jacking up interest rates in the short run, appreciating the exchange rate and causing a recession in the short term.⁵⁰ Calvo and Vegh's framework provided a fairly strong rationale for ERBS from the perspective of myopic politicians eager to obtain significant short-run effects.

4.2 Financial integration and financial crises in the 1990s

As inflation concerns subsided and financial integration increased in the second half of the 1990s, the exchange rate policy debate in developing economies shifted the focus to the interplay of two contrasting features of financial development. First, the fact that financial globalization led to a growing ineffectiveness of monetary policy or, more precisely, that capital controls were found to be decreasingly effective as economies became more sophisticated. As in the early years of the twentieth century, growing financial integration and sophistication in the developed world strengthened the restrictions imposed by the *impossible trinity*—previously circumvented due to the absence of *de facto* financial integration (Obstfeld & Taylor, 2004; Rose, 2006)—all of which made floating regimes more attractive.⁵¹

Second, the role of (domestic and external) FD, namely, the foreign currency denomination of residents' assets and liabilities that, to the extent that it introduced currency exposures that raised the risk associated with exchange rate jumps, made

pegged regimes look more attractive. Indeed, it was the risk of balance sheet losses to financially dollarized governments and firms in the event of a devaluation—stressed in third generation models of currency crises popularized in the context of the Asian crisis—that led to the definition of *fear of floating* (Calvo & Reinhart, 2002), namely, recurrent *de facto* exchange rate intervention in officially floating regimes. In turn, to the extent that FD detracted from the benefits of flexible regimes, exchange rate flexibility could become a source of volatility, and hard pegs could be viewed as a reasonable option.⁵²

The combination of these two factors led naturally to one of the dominant proposals in the late 1990s, the “bipolar” view (Fischer, 2000) that noted that pure flexible exchange rates or superfixed regimes (the so-called “hard” pegs, such as currency boards or unilateral dollarization) were the only viable alternative for financially integrated developing economies, at the expense of conventional pegs, inherently vulnerable due to monetary policy inconsistencies and self-fulfilling speculative attacks. Combined with the fear of floating view, this approach derived naturally into what could be called a “unipolar view” (Calvo 1999, 2000) according to which hard pegs were the only sensible option for financially dollarized economies: if devaluations in dollarized economies were contractionary due to balance sheet effects, exchange rate flexibility would only amplify the cycle, rather than smooth it out as predicated by the standard theory.⁵³ Thus, exchange rate anchors in the globalized world evolved from a signal to align expectations into legal constraints on the behavior of the Central Bank: ultimately, in the quest for credibility, the tyranny of the anchor eliminated the active pursuit of monetary policy altogether.

But while the debate suggested this one-way street, endorsed by multilateral organizations in the mid-1990s hand in hand with the successful example of the currency board in Argentina, policy was heading in the opposite direction. Paradoxically, by the turn of the century the failure of Argentina’s currency board to ensure fiscal and monetary discipline casted doubt on the premises underscoring the unipolar view. The market discipline that would impose a hard budget constraint on the government in the absence of monetary financing did not materialize: furthermore, procyclical capital markets lent to levels that proved unsustainable, and pulled off in bad times triggering a debt default.⁵⁴ On the other hand, the fact that, at the time of the currency run, the contraction of the monetary base caused by the unsterilized sale of reserves was neutralized by the issuance of fiat money by the national and subnational governments showed that not even monetary discipline was guaranteed by the currency board agreement. At any rate, for many observers, the hard pole of the bipolar view was restricted, at best, to the yet untested unilateral dollarization, a more extreme and less appealing choice.⁵⁵

Also, by the end of the decade the success in building central bank autonomy and monetary credibility, together with the resulting decline in inflation and exchange rate

pass-through, led to the growing popularity of the float pole of the bipolar view as the background for different inflation targeting arrangements that prioritized the inflation rate, rather than the exchange rate, as the key nominal anchor, an option that recovered the possibility of exercising monetary policy. Not surprisingly, among emerging countries, this trend started in economies with relatively low levels of FD (Chile, New Zealand, South Africa, Brazil), gradually extending to other countries *pari passu* with a reduction in their degree of dollarization. Ultimately, as mentioned above, the debate in the new millennium appears to have converged to an inverted unipolar view, where flexible regimes are seen as the only sensible (and durable) choice as economies grow financially integrated and sophisticated.⁵⁶

4.3 Float cum inflation targeting (FIT)

The declining degree of FD, combined with the improved quality of monetary institutions, explain the evolution of MERP in recent years. The recent changes in debt composition and policy quality in developing countries have led developing economies to use the inflation rate rather than the exchange rate as the main policy target, allowing greater flexibility for the latter. This has led some observers to salute FIT as a new, possibly more resilient MERP paradigm (Rose, 2006).

FIT is in practice a broad category that includes a large array of alternative varieties, going from soft numerical inflation target (in the form of a wide inflation band) to a more sophisticated system that includes, additionally: (i) a legal commitment to price stability as the primary goal of monetary policy, (ii) a dissemination strategy that allows agents to replicate and anticipate the policy decision context (if not the actual policy decision); (iii) direct accountability of the central bank management for attaining the targets.⁵⁷ From an operational point of view, an inflation targeting regime typically implies identifying an intervention variable, usually a reference interest rate for funds offered by the central bank. This rate is defined and discussed in regular meetings, the proceeds of which are made available to the public, sometimes with a lag.

Historically, middle income developing countries adopting IT gradually proceeded from the soft version that in the early years usually coexists with a dirty exchange rate regime (see Schmidt-Hebbel & Tapia, 2002 for Chile; Armas & Grippa, 2006 for Peru; Fraga, Goldfajn, & Minella, 2003 for Brazil, and Mishkin, 2004) to the more canonical version. At any rate, the interpretation of existing empirical studies trying to assess the real implications of IT should be qualified by the fact that they are likely to cover different IT varieties for each individual country. Moreover, the introduction of IT in developing countries often coincided with the transition from moderate two-digit to low one-digit inflation—and countries that choose IT exhibited higher initial inflation—so that a sacrifice ratio that captures this transition may overstate the net benefits of IT once inflation is brought under control.

4.3.1 The FIT paradigm and the real economy

The literature on the consequences of FIT on the real economy in the developing world suffers from two important shortcomings. The first one, as noted, is semantic: FIT adopts a number of varieties that are not always strictly comparable. This caveat is more generally related with a definition problem that plague inflation targeting as a distinct policy: if by inflation targeting one means an explicit commitment with low and stable inflation, then most central banks in mature economies (and most in high-middle income ones) are in fact inflation targeters. Thus, the empirical characterization of inflation targeting, in practice, hinges on the two other pillars mentioned above, namely, dissemination and accountability, and the boundaries of what constitutes IT and what does not are rather fussy.⁵⁸

The second drawback faced by the empirical literature owes to the fact that IT in developing countries has been adopted: (i) very recently (Chile and Israel lead the way in the mid-1990s, although they implemented a fully fledged IT framework only recently); (ii) in times of moderate (two-digit) inflation. In other words, whereas there is some evidence about the ability of IT to bring down inflation at a reasonable sacrifice ratio (in terms of slower and possibly more volatile growth),⁵⁹ much less can be said about its relative advantages for developing economies once inflation declines or new global shocks develop. Bearing this in mind, a number of recent empirical studies take stock of the IT experience in the developing world.

These studies (which often include both industrialized and developing economies) have yielded mixed results. On the one hand, there appears to be no conclusive evidence on its effect on the sacrifice ratio: inflation targeters enjoy sacrifice ratios and output volatility that is lower than before the adoption of IT, but comparable to those observed in noninflation targeting industrial countries (Cecchetti & Ehrmann, 1999; Corbo & Schmidt-Hebbel, 2001).⁶⁰

On the other hand, there is no convincing evidence that they perform better than comparable nontargeters in other respects. While IT advocates point out that the adoption of IT in developing countries help bring down inflation (Corbo & Schmidt-Hebbel, 2001) and align inflation expectations reducing pass-through coefficients (Corbo, Landerrretche, & Schmidt-Hebbel 2001), they also stress that developing countries have performed relatively worse than industrial targeters: deviations from targets are larger and more frequent (Fraga et al., 2003). Even for industrial economies the jury is still out on the IT advantage, either in terms of output volatility, interest rates, or even inflation level and variability (Ball & Sheridan, 2003). Indeed, it appears that inflation targeting countries tend to have a high initial inflation (which, not surprisingly, increases the propensity to adopt IT, Mishkin & Schmidt Hebbel, 2001) and, correspondingly, large short-run decreases. In short, IT has been instrumental in bringing inflation rates to one-digit levels, but once there, its benefits are more difficult to identify (Box 5).

FIT has had its most severe test to date during the 2007–2008 inflation rollercoaster. Supply shocks unrelated to domestic demand are usually transitory and, for this reason,

Box 5 Inflation Targeting and Demand and Supply Shocks

Consider a minimal world with demand and supply shocks. Assume output is determined by a stylized version of an IS curve:

$$y = d + s + \beta m,$$

where y stands for output, d for demand shocks, s for supply shocks and m is a monetary variable that affects output through the coefficient β . Inflation depends on the same three variables:

$$\pi = m - \omega s + \eta d.$$

Notice that while inflation is a monetary phenomenon, it increases with demand shocks and declines with supply shocks. Consider now a pure inflation targeting Central Bank. Such a Central Bank will choose monetary policy to minimize the volatility of inflation from its target. Thus, m will be chosen by

$$m = \pi^* + \omega s - \eta d,$$

where π^* is the target inflation rate. Replacing the solution for monetary policy into the output equation gives

$$y = \beta\pi^* + d(1 - \beta\eta) + s(1 + \beta\omega).$$

which results in

$$\sigma_\pi^2 = 0 \quad \text{and} \quad \sigma_y^2 = \sigma_d^2(1 - \beta\eta)^2 + \sigma_s^2(1 + \beta\omega)^2. \quad (1)$$

The key point of the model is to show that in the presence of supply shocks output volatility is magnified. Notice that the coefficient on the volatility of supply shocks σ_s^2 in Eq. (1), $(1 + \beta\omega)^2$, is greater than 1, indicating that inflation targeting tends to increase the volatility of output rather than reduce it. The reason is that supply shocks increase the inflation during downturns, forcing a procyclical contraction and vice versa. Thus, the inflation targeting paradigm will have a harder time when the supply shocks are common, explaining why some inflation targeters have abandoned an orthodox application of the paradigm during the 2008 financial crisis.

partially dismissed under the IT framework by targeting an adjusted (core) price index less sensitive to supply swings. In developing economies, the lack of institutional credibility led central banks to favor the more sensitive headline CPIs over opaque core measures more prone to be perceived as biased indicators of genuine inflation. The sharp increase in international food and energy prices through mid-2008, represented an unexpectedly large and long supply shock that ultimately reflected in the above-target inflation and, more importantly, inflation expectations deanchored from the inflation target, forcing central banks out of their benign neglect and into tightening mode—in some cases, even in a context of a cooling economy. Merely 3 months later, the deepening of the financial crisis after the bankruptcy of Lehman Brothers on September 15, and the ensuing collapse of commodity prices and downbeat growth

outlook, again led central banks to switch back to monetary easing, even before inflation came within the target band, in many cases intervening heavily to contain the currency. At the time writing this chapter, the outcome of the global crisis is uncertain, and it would be premature to draw conclusions about the fate of FIT. We can only speculate that the recent episode, coupled with the lack of convincing evidence on the superiority of FIT for low inflation countries, may ultimately lead to a reassessment of FIT in the developing world, its implementation (e.g., the inflation measure to target) and the weight it should give to growth considerations yet another twist in the history of MERP that we leave for a future survey to document.

4.4 The comeback of exchange rate regimes: Leaning against the appreciation wind

Unlike in the 1990s, where financially dollarized economies resisted depreciation because of the presence of currency mismatches and widespread dollar indexation, in recent years central bank intervention have been mostly leaning against the appreciation wind, a behavior that not only has distinct motivations and (presumably) economic consequences, but also is in stark contrast with the FIT paradigm predicated by many developing economies.

In recent years while about 25 middle-income developing countries officially subscribed to FIT, many countries (China, Malaysia, Thailand, Colombia, and Argentina, to name a few) were still pursuing active exchange rate policies, and three of them (Argentina in 2005, Thailand, and Colombia in 2006) introduced controls on capital inflows to countervail the appreciation of their currencies.

This comeback of exchange rate policies (which for simplicity could be labeled “fear of appreciation”) has been attributed to two main motives: a prudential motive aimed at countering the volatility of procyclical capital flows, and a mercantilist attempt to keep the currency undervalued as a means to foster exports and to protect the domestic industry from international competitors. We examine both motives in turn.

4.4.1 The prudential motive

The first interpretation of the current surge in international reserves in developing economies shown in Figure 8, had to do with prudential considerations, specifically, the fear of a shortage of liquid foreign assets of the type that caused the many emerging market financial crises in the second half of the 1990s. In this view, the less than perfectly flexible exchange rates that characterized many developing economies in the early 2000s were simply the result of the rapid accumulation of precautionary reserves in the aftermath of a crisis at home or in the neighborhood—a hypothesis partially supported by the data (Aizenmann & Lee, 2005; Aizenmann & Marion, 2004).⁶¹

Indeed, a similar motive could be conceived for a more explicit exchange rate objective. For example, a policy of leaning against the appreciation wind during expansions may be seen as the countercyclical prudential response to procyclical (and largely

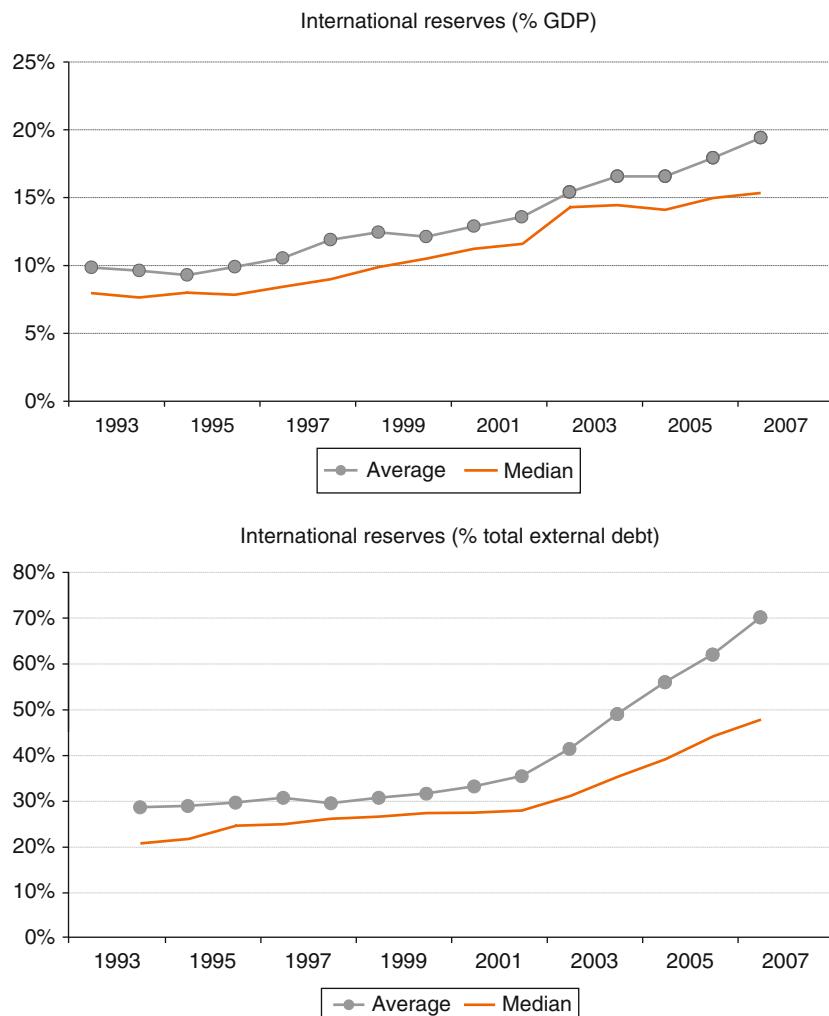


Figure 8 Reserve accumulation in recent years. Source: Author's calculations using data from Bank for International Settlements, IMF and World Bank.

exogenous) swings in capital flows and real exchange rates. Limiting the transitory (and possibly excessive) appreciation of the local currency through the accumulation of foreign reserves in this context would be a natural defensive strategy to limit the country's external vulnerability and minimize the real exchange rate adjustment and the associated balance sheet effects during the recessive phase.⁶²

But there are clear indications that this, if at all relevant, is only part of the story. On the one hand, many of these economies are not financially dollarized or have seen their external debt to GDP ratio fall dramatically in recent years, at the time reserve

accumulation was at its peak. If prudential concerns were at the root of the initial surge in intervention, it is difficult to attribute the still ongoing process to liquidity risk.

Prudential issues and currency mismatches certainly played an indirect role in the protectionist rationale for intervention: a declining degree of FD relaxed the balance sheet concerns behind the fear of floating, recovering the expansionary benefits of depreciations. Indeed, the main hypothesis of the mercantilist view (namely, the pro-growth consequences of an undervalued currency) depends critically on the absence of the currency mismatches usually found in financially dollarized economies. Thus, the intensity of the precautionary motive in the recent years is not independent of the decline in FD in the developing world.

4.4.2 The mercantilist motive

Perhaps the most intriguing new development in the MERP debate comes from an old unresolved question: Does a temporarily high real exchange rate have a persistent positive effect on economic activity? If so, does this effect come from an increase in external demand, a decline in the demand for imports (with a concomitant increase in the demand for domestic products), or is it related to income distribution and the dilution of producer costs?

A number of recent papers examine the issue and provide supportive (albeit contradictory) evidence. While they tend to agree with the fact that mercantilist interventions and undervalued currencies are associated with faster growth, they are far less clear about the specific channel in place. Some arguments are in line with traditional export-led dynamics (Prasad et al., 2006). Rajan and Subramanian (2005), for example, analyze the impact of foreign aid and show an adverse effect on sectors with a higher exported share, which they attribute to the effect of the real appreciation associated with the inflow of funds. Inverting this Dutch disease argument, a real depreciation would foster the growth of export-oriented firms.⁶³ Some offer an alternative argument: in a reversion of Diaz Alejandro's (1965) contractionary devaluation story, Levy Yeyati and Sturzenegger (2007) argue that because devaluations reduce labor costs in terms of producer prices, rather than fuelling capital flight, they increase firm profitability and real investment.⁶⁴

Many of these studies suffer from a potential endogeneity problem, to the extent that intervention (and, depending on how it is measured, undervaluation) may also be the result of good economic conditions (including faster growth). At any rate, even if the direction of causality implied by these findings were true, there would remain the question about the effectiveness of the policy; more precisely, its cost and the persistence of its benign effect.

How long and at what cost can this proactive exchange rate policy coexist with an autonomous monetary policy aimed at price stability? One important distinction needs to be made when computing the cost of intervention: whether or not the government holds debt in the foreign currency. If it does, the marginal cost of carrying reserves is

proportional to the marginal cost of the debt that implicitly funds them (alternatively, that could be cancelled with the reserves), net of the returns obtained on reserves—which typically amounts to the sovereign spread over the risk-free rate.⁶⁵ If it does not, the purchase of reserves to defer the appreciation of the local currency can be funded essentially in two ways: by issuing money, or by issuing local currency-denominated debt.⁶⁶ The first option introduces inflation pressures: the appreciation materializes—albeit over a longer time—through a change in domestic prices rather than in the nominal exchange rate. The second option pays the local currency interest rate (the central bank's quasi fiscal cost) or, belatedly, incurs valuation losses as the currency gradually appreciates.

Sterilization costs, however, are less straightforward than they sound in theory. Sterilized purchases of foreign exchange are seldom accompanied by higher interest rates—because appreciation expectations tend to depress borrowing costs in the local currency. Instead, to the extent that intervention simply delays the transition to an appreciated exchange rate (hence, the appreciation expectations), it should ultimately lead to a loss in the form of changes in the local currency value of international reserves, as the exchange rate appreciates toward the new equilibrium. It follows that, under the interest rate parity condition, the difference between the local currency interest rate and the expected appreciation rate should equal the dollar interest rate, so if expectations are unbiased, the difference in the cost should ultimately be, on average, similar to the case in which reserves are directly funded by dollar debt, except that the central bank bears the currency risk. Mercantilist reserves accumulation would be costly if appreciation pressures signal permanent changes in the country or the external environment.

Conversely, if appreciation pressures turned out to be a transitory phenomenon due, for example, to cyclical inflows or a transitory run on the currency, the reversion of the exchange rate to its earlier, more depreciated level may eliminate valuation losses and much of the intervention cost. The fact that equilibrium exchange rates are in practice so difficult to assess—and, as a result, often assumed to be random walks—makes the evaluation of long-term intervention costs rather difficult to pin down *ex ante*.

How was reserve accumulation financed in practice? A first answer to this question is shown in Figure 9. There, we picked fast-growing emerging economies that have been accumulating reserves in the period 2003–July 2008 (prior to the reversal of the appreciation phase), and compared the local currency equivalent of dollar purchases (adjusting for valuation changes using the monthly average exchange rate and assuming an average return on reserves equal to the 1-year Libor), with the contemporaneous expansion of the monetary base. As can be seen, the landscape is not homogenous: the share money creation (in turn, seigniorage and inflation tax) to reserves purchases ranges from over 100% in Philippines and Indonesia to less than 20% in Singapore, Korea and Brazil).

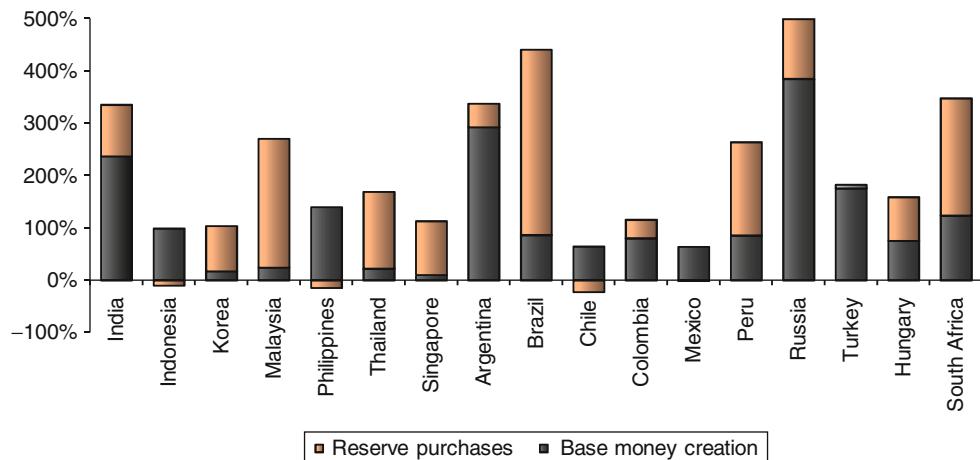


Figure 9 How are reserves funded? (2003-2007). Source: Author's calculations using data from Reserve Bank of India, Bank Indonesia, Bank of Korea, Bank Negara Malaysia, Central Bank of the Philippines, Monetary Authority of Singapore, Bank of Thailand, IMF, Banco Central de Chile, Banco de México, Central Bank of the Russian Federation, Central Bank of Turkey, National Bank of Hungary and South African Reserve Bank.

Advocates of reserve accumulation had their belated recognition in the midst of the financial crisis of 2007-2008, when the stock of reserves enabled financially integrated emerging economies to control the pace of the exchange rate adjustment needed to offset the rapid unwinding of foreign investment positions and the terms of trade shock—triggered by the global recession that in the 1990s may have caused a stream of balance of payment crises. As a result, the prudential motive—or, more precisely, the policy of smoothing out the cyclical pattern of exchange rates—looks, a fortiori, a plausible justification for reserve accumulation. Note, however, that the prudential and mercantilist motives are not at odds with each other: An eventual reversal of fortunes may have also been in the minds (and often in the words) of many policy makers in nonindustrial countries that targeted an undervalued currency as a development tool in the good years.

5. WHERE DO WE STAND?

A number of lessons can be drawn from the previous discussion. The first thing to note is that the MERP debate is far from closed. This is a natural consequence of the fact that the pros and cons of alternative MERP (and actual policy choices) evolve both with country characteristics and the global context. Exchange rate anchors that were popular in the developing world in the context of high, inertial inflation and partial dollar

indexation lost their edge when central banks won the inflation battle and pass-through coefficients declined—coincidentally, at a time when financial integration rendered pegged regimes more vulnerable to self-inflicted crises or self-fulfilling attacks. On the other hand, the recent process of external deleveraging and dedollarization in the developing world, by reducing currency imbalances, increased the scope to use flexible exchange rates as shock absorbers and, by eliminating the need to defend a parity in times of distress, enhanced the scope for countercyclical monetary policy.

The fact that most medium and large developing economies (and virtually all industrial ones) reveal a preference for exchange rate flexibility and nonexchange rate anchors simply reflects this evolution. However, pegs still represent more than half of the IMF reporting countries—particularly, small ones—indicating that exchange rate anchors are still favored by small open economies that give priority to the trade dividend of stable exchange rates and find the conduct of an autonomous monetary policy too costly, due to lack of human capital, scale, or an important non-tradable sector.⁶⁷

It would be misleading to draw the debate to a closure here. Do we need another paradigm? Our reading of the literature advises against it, as it highlights the importance of country characteristics and the correspondence between some schemes and the specific contexts for which they have been instrumental. Indeed, the two apparent contending paradigms of the new millennium look set for a reassessment after the current global crisis is over.

The neo-mercantilist's attempt to mitigate the appreciation of local currencies trend is not qualitatively different from the fixed exchange policies adopted in the wave of capital flows to emerging economies in the early 1990s. Does it constitute today a distinct MERP, one with the objective of a *persistently* undervalued local currency as a substitute to more specific tariff barriers? Or is it simply the countercyclical smoothing of high-frequency exchange rate variability—in turn, fueled by procyclical cross-border flows—that in financially integrated economies may lead to unwarranted swings in the real exchange rate? In the second case, one could invert the question to ask whether small open economies can “afford” full flexibility in such a context. The ongoing unwinding of cross-border positions and sharp exchange rate corrections—which elicited heavy central bank intervention in the opposite direction—suggests that, whatever the true objective was at the origin, the prudential motive for reserve accumulation may have been proven right by the global crisis. On the other hand, the margin for mercantilist policies has been severely reduced by the decline in commodity prices, so we would not be surprised that the policy debate—and, as a result, the literature—turns to this tension in the future.

Another natural candidate for renewed debate is the FIT paradigm, which appears to have failed the test imposed by a succession of extreme positive and negative real shocks for which it was only partially prepared. Will the IT toolkit remain intact after

the dust of the current crisis settles, or will it be augmented by a broader consideration to growth as in the US, or will it be dismissed altogether on its less than stellar relative performance in low inflation environments?

While it is still too early to judge IT or exchange rate intervention policies, a number of recent developments in emerging and developing economies highlighted by the present study has enhanced their ability to conduct independent monetary and exchange rate policies as countercyclical tools. This progress, which owes as much to economic theory as it does to economic history, underlies the current state of the debate on MERP in the developing world, and its growing role in the context of economic development.

End Notes

1. Levy Yeyati: Head of Emerging Markets Strategy, Barclays Capital; Sturzenegger: President, Banco Ciudad de Buenos Aires, respectively. This chapter was largely written when the authors were with the World Bank and the Kennedy School of Government, Harvard University, respectively. The authors are grateful to participants in the Handbook of Developments Economics Launch Conference at Harvard University and Mauro Roca and Rodrigo Valdés for helpful comments, and to Pablo Gluzmann and Andrea Kiguel for outstanding research assistance. The views expressed here do not necessarily reflect those of the institutions the authors are affiliated with.
2. Later developments, including Dornbusch's famous overshooting model (Dornbusch, 1976), broadened the applications of the model to the workings of foreign exchange and financial markets.
3. A popular choice to model this price rigidity is Calvo's (1983) price staggering mechanism. In Calvo's model firms are allowed to change prices randomly, but once they do so, they do so rationally anticipating the conditions of the economy during the period in which they think the price will be relevant. Because change opportunities appear stochastically and independently across firms, it means that a constant fraction of firms adjust their prices in each period, making the price level a smooth variable that changes only over time. A simpler structure (assuming that prices have to be set one period in advance) are used in Obstfeld and Rogoff (1995), the first fully fledged general equilibrium model with price rigidities applied to the international framework. See also Obstfeld and Rogoff (2000), Bacchetta and van Wincoop (2000), Betts and Devereaux (2000), and Corsetti and Pesenti (2001).
4. These neo-keynesian models have come a long way from its closed economy versions (Woodford, 2003). Following the initial lead of Calvo (1983) and his work on stabilization (to which we come back below), to applications to macro model building by Kollmann (2001) and Chari, Kehoe and McGrattan (2002). Other relevant references include McCallum and Nelson (2000), Corsetti and Pesenti (2005), Clarida et al. (2001, 2002), Schmitt-Grohe and Uribe (2001), Kollman (2002), Parraido and Velasco (2002), and Benigno and Benigno (2003) among others.
5. The classification reflects the subjective assessment of IMF country economists of the *de facto* policies conducted in the country. We describe and compare this and other MERP classifications later in this chapter.
6. The IMF's *de jure* classification, sourced from Ghosh et al. (2003), reproduced the regimes officially informed by the countries' monetary authorities and was discontinued in 2000.
7. Arguably, while a mobile exchange rate is in direct contradiction with a peg, limited variability does not necessarily contradict a float, since exchange rate stability may simply reflect a stable environment. We come back to this identification problem below.

8. Taylas et al. (2006) provide a comprehensive survey.
9. See also Bénassy-Quéré and Coeuré (2001).
10. See Eichengreen (1994) and Fischer (2000), among others. We revisit these views in Section 4, when we discuss the connection between MERP and country characteristics.
11. Following the convention in the literature, an economy is denoted as emerging when it is included in the Emerging Market Bond Index Global Portfolio compiled by J. P. Morgan, which requires that the country has issued frequently traded sovereign bonds in international markets.
12. While typically it is the US dollar that is used, the term is just a convention for the use of any foreign currency as national legal tender.
13. See Sturzenegger (2008) for description and further references.
14. Calvo and Reinhart (2002) define the typical intervention of a floater as the changes in these variables of some uncontroversial floats: Australia, Japan, and the US, and then compare the intervention variables in specific emerging countries with those in the model floats.
15. Unless policy makers believe that an appreciation increases the risk of a posterior devaluation.
16. Aizenmann and Lee (2005) who test whether this motive explains the buildup or reserves, associate this view with the Bretton Woods II approach of Dooley, Folkers-Landau, and Garber (2003). But, as we discuss below, the mercantilist approach has a large tradition in development economics; see Eichengreen (2006) for a survey.
17. The seminal contribution on this front was Kydland and Prescott (1977). Calvo (1978) provided an alternative model, focusing on the time inconsistency problem of domestically denominated debt. The setup achieved textbook status with Barro and Gordon (1983). In later years, the problem of time inconsistency led to an explosion of work focused on how policies should be framed to deal with it. See Rogoff (1985) on conservative central bankers, Backus and Driffil (1985) or Cukierman and Meltzer (1986) on reputation models; and Alesina (1988), Alesina and Summers (1993), Grilli, Masciandaro, and Tabellini (1991), and Cukierman, Webb, and Neyapti (1992) on the independence of the Central Bank. The time inconsistency problem has been and is still a key feature of monetary policy debates, all the way through the current discussion on inflation targeting.
18. To reduce this constraint and allow some margin to accommodate unexpected shocks, the anchor is sometimes stipulated as a range (a band) to be met in the medium term.
19. The options include, for the target price index: Core CPI, Headline CPI, Core excluding food, energy and indirect taxes, Core excluding regulated prices, and other similar variants. For target width the range is typically 2–3% and most countries have a positive floor for inflation except New Zealand and Thailand whose range starts at zero. The target horizon is typical one year but in many cases it is indefinite. Most countries do not have escape clauses but some do under fairly undefined circumstances. Accountability also differs. It is quite famous in New Zealand the provision that the Minister of Finance may ask for resignation of the Governor. Less drastic is the open letter to the Minister of Finance explaining the target breach in the UK.
20. Consider for example that description that the IMF does of the Chinese monetary regime, an alleged monetary aggregate targeter: On July 21, 2005, China announced a 2.1% revaluation of the renminbi-US dollar exchange rate and a change in its exchange rate arrangement to allow the value of the renminbi to fluctuate based on market supply and demand with reference to an undisclosed basket of currencies. To permit a greater role for market forces in determining the renminbi exchange rate, steps have been taken since July 2005 to liberalize and develop China's foreign exchange markets, including the establishment of an over-the-counter spot foreign exchange market and markets for currency swaps and futures. From end-July 2005 to end-July 2006, the renminbi exchange rate was more flexible, but the fluctuation in the renminbi-US dollar exchange rate was less than the 2% range (for a 3-month period) used in the IMF's *de facto* exchange rate classification system as an indicator for a conventional fixed peg exchange rate arrangement.

21. Svensson (1997) is the classical reference where IT rules are derived from an optimal program for the Central Bank.
22. Sturzenegger and Talvi (2008) estimate the reaction function of central bankers for a group of Latin American countries following the methodology of Lubik and Shorfheide. They find that reaction functions, with exceptions, have turned out pretty stable with low and declining weights on the exchange rate motive. Typically, the inflation coefficient is above 1, signaling a countercyclical monetary policy, with two exceptions—Argentina and Ecuador.
23. This does not include the economies of the euro zone, which target inflation jointly but are typically excluded from the float group.
24. In recent years, the decline in financial dollarization in economies like Peru or Turkey has made financial stability speed up the convergence to fully fledged inflation.
25. Note the similarity with Kose et al. (2006), where a similar distinction is used to characterize another nominal-real connection: financial integration and economic performance.
26. In the figure, i^* represents international risk-free interest rates, ρ stands for the sovereign risk premium, e denotes the exchange rate, g_y and σ_y are output growth and volatility, and π is the inflation rate.
27. These objectives may interact among themselves: Output volatility may be associated to (lower) output growth, (high) inflation may be associated with (lower) growth, and equity and growth may affect each other (an old debate dating back to Kuznets's inverted U curve hypothesis). This survey, however, will not deal specifically with these complex interactions.
28. These two arguments suggest a potential bias of classifications based on exchange rate variability to find flexibility associated with bad outcomes—and an opposite bias for codings where flexibility is associated with no policy intervention.
29. The growth literature, in general, has been usually criticized along these lines.
30. Assuming that growth opportunities are concentrated in the tradable sector, Hausmann and Rodrik (2003) argue in favor of a depreciated exchange rate to foster innovation. A similar reasoning leads Rodrik (2006a) to argue that a competitive exchange rate may be an efficient development tool.
31. Neo-mercantilism as a deliberate policy decision has also been under dispute. For example, Aizenman and Lee (2005) argue that the evidence on reserve accumulation favors prudential over mercantilist motives. There is a literature on overvaluation-misalignments and growth: Razin and Collins (1997), Aguirre and Calderón (2005), Aizenman and Lee (2005), Dollar (1992), Sachs (1985). We come back to this issue below.
32. A view that goes back to Meade (1951) and Friedman (1953). See also Dornbusch (2001) and Kenen (2002).
33. The time inconsistency version of the inflation bias builds on work for closed economies: Kydland and Prescott (1977), Calvo (1978), Barro and Gordon (1983), Rogoff (1985), Walsh (1995), Persson and Tabellini, (1993), and Svensson (1995). For open economies, see Giavazzi and Pagano (1988) and Obstfeld (1996).
34. Some authors have suggested that the channel may work in the opposite direction: flexible rates provide more credibility. That is the argument, for example, in Tornell and Velasco (2000), on the basis that fiscal mismanagement implies costs in the long run under fixed regimes but is immediately apparent when exchange rates are flexible, which then provides the strongest incentives for consistent fiscal behavior.
35. As we will discuss below, hard pegs represent the extreme example of this line of reasoning, increasing exit cost in a number of ways (attaching a legal framework to the peg, fostering the use of the peg currency, and even eliminating the national currency in the case of unilateral dollarization).
36. See Ghosh et al. (1997, 2003), Levy Yeyati and Sturzenegger (2001), Rogoff et al. (2004), and De Grauwe and Schnabl (2005). This suggests that countries with higher pass-through coefficients will tend to benefit the most from the immediate impact of the anchor on inflation expectations—and explains why they were its most active promoters.

37. There seems to be some evidence on the benign effect of hard pegs on fiscal discipline. Ghosh, Gulde, and Wolf (1998) and Culp, Hanke, and Miller (1999) argue that countries on currency boards tend to run tighter fiscal policies, whereas Fatas and Rose (2000) find that currency boards are associated with fiscal restraint (although, somewhat surprisingly, this restraint does not carry on to unilaterally dollarized economies or to members of a monetary union).
38. Among many others, see Thursby and Thursby (1987), De Grauwe (1988), Brada and Mendez (1988), and Parsley and Wei (2001a, 2001b).
39. See Barro and Tenreyro (2007) and Persson (2001) among others. Frankel (2005) provides a defense of Rose's results in the face of these criticisms.
40. Rose's (2004) meta-analysis of 34 previous studies yield an estimated increase of between 30% and 90%.
41. See also Klein (2002), who in related work contends that the trade effect of unilateral dollarization does not differ significantly from that of a conventional peg.
42. See also Wei and Choi (2002). It has to be noted, however, that their result is subject (albeit to a lesser extent) to a sample problem similar to that plaguing the early literature on common currencies and trade: apart from a few currency board countries, the rest of the hard peg group comprises very small economies and subnational entities.
43. There is, in addition, a literature documenting the complementary link between trade and FDI (Clausing, 2000; Lipsey & Weiss 1981; Svensson, 1996).
44. Terra (1998) provides a critical view of this result, suggesting that it confounds the effect of openness with that of the debt crisis.
45. While Lee, Ricci, and Rigobon (2004), in response to Rodrik and Rodríguez (2000), apply identification through heteroskedasticity—a methodology developed by Rigobon (2000)—to find that openness has a positive—albeit small—effect, Rigobon and Rodrik (2004), using the same methodology, found that openness has a negative impact on income levels.
46. Broda and Weinstein (2006) have argued that increased product variety resulting from trade leads to significant undermeasurement of import price indices, suggesting that the welfare gains may be larger than anticipated. Based on an analysis of the share of income spent on food, and an estimate of Engel curves, Chamon and Irineu de Carvalho (2006) also argue that the gains from trade are larger than typically measured. They apply this methodology to Brazil and find that as a result of trade liberalization real income growth was closer to 4.5% per year rather than the official 1.5%.
47. See also the Kose et al. chapter in this volume for a more detailed survey.
48. The argument goes beyond the case of bailouts: any implicit debtor insurance, to the extent that defaults are correlated with the real exchange rate, would favor dollarization. For example, the accumulation of international reserves may fuel the dollarization of the banking sector, if they are perceived by commercial banks as increasing the probability that the central bank provides dollar liquidity in the event of a dollar shortage (Broda & Levy Yeyati, 2003).
49. See Broda and Levy Yeyati (2003, 2006) for the case of banking regulation; Chamón (2001) for sovereign debt contracts, and De la Torre, Levy Yeyati, and Schmukler (2002) for a discussion in the context of the Argentine currency board.
50. De Gregorio, Guidotti, and Vegh (1998) suggest that the boom in ERBS comes from the effect of the interest rate collapse on the purchase of durable goods.
51. The impossible trinity refers to the inability to sustain simultaneously three policy objectives: an independent monetary policy, open capital markets, and fixed exchange rates: If monetary policy and open capital markets are priorities, exchange rates need to float. If exchange rate and capital markets are priorities countries cannot have an independent monetary policy. If monetary policy and exchange rates are priorities capital markets need to be shut down.
52. See, for example, Barro, (1999), Hausmann, Gavin, Pages, and Stein (1999), Haussmann, Pianzza, and Stein, (2001), Ghosh et al. (1997), and Dornbusch (2001).

53. See Frankel (2005) on balance sheet effects and contractionary devaluations.
54. It is possible that the shift from bank to (typically atomistic) bond financing as a result of the creation of the emerging market bond class with the Brady plan deepened this uncoordinated procyclical behavior displayed by international capital markets *vis-a-vis* developing economies.
55. For a discussion of the 2001 Argentine crisis along these lines, see De la Torre et al. (2002) and Hausmann and Velasco (2002). See Levy Yeyati and Sturzenegger (2006a) for a review of the dollarization debate.
56. See Levy Yeyati (2005) and references therein.
57. Truman (2003) provides a comprehensive and general discussion of IT. Price stability need not be the only mandate; IT may assign a role for output stability (e.g., the Reserve Bank of Australia). The same is true for financial stability in financially dollarized economies like Perú or Uruguay, although in those cases the application to the IT club is still under consideration.
58. The European Central Bank (ECB), for example, has a numerical inflation target, but is not considered an inflation targeter due to a general lack of transparency in the communication of the policy-making process (Svensson, 2000).
59. The standard measure of the sacrifice ratio computes the output loss associated with a unit percentage change in inflation.
60. Interestingly, Cecchetti and Ehrmann (1999) find similar results for noninflation targeting European Union (EU) countries as they focus on inflation in the run up to the monetary union.
61. Caballero and Cowan (2006) argue that while there are arguments for the government to purchase insurance, the latter should be done not through reserve accumulation but rather through the use of derivative markets. Summers (2006) considers that reserves are larger than justifiable from a prudential motive—and should, therefore, be managed as long-term savings. Rodrik (2006b) also argues reserves are too large for a prudential motive.
62. See Levy Yeyati (2005) and Caballero and Lorenzoni (2006).
63. However, due to the way in which they measure sector growth, the effect may capture the higher price level perceived by the exporter as a result of the devaluation, rather than actual growth.
64. Levy Yeyati and Sturzenegger (2007) show that depreciations work not so much through the trade channel but through an increase in savings and investment associated with the regressive income distribution effects of devaluations. See also Aghion et al. (2006) for a model along these lines.
65. To the extent that, for a given net debt stock, a larger stock of liquid foreign currency assets may tighten the sovereign spread, the resulting gain in rollover costs should be net out from the spread in computing the marginal cost of reserves (Levy Yeyati, 2006). For alternative takes on the cost of precautionary reserves, see also Rodrik (2005) and Jeanne and Ranciere (2006).
66. Since intervention is geared to offset the demand for the local currency, the issuance dollar debt would not do the trick in this case.
67. Pegs account for more than 50% of classified countries, both under *de jure* and all three *de facto* classifications described in Section 1.

References

- Aghion, P., Bacchetta, P., Ranciere, R., & Rogoff, K. (2006). *Exchange rate volatility and productivity growth: The role of financial development*. NBER Working Papers 12117.
- Aghion, P., Fally, T., & Scarpetta, S. (2006). *Credit constraints as a barrier to the entry and post-entry growth of firms: Lessons from firm-level cross country panel data*. Mimeo.
- Aghion, P., Howitt, P., & Mayer-Foulkes, D. (2005). The effect of financial development on convergence: Theory and evidence. *Quarterly Journal of Economics*, 120(1), 173–222.
- Aguirre, A., & Calderón, C. (2005). *Real exchange rate misalignments and economic performance*. Documento de Trabajo No. 315. Banco Central de Chile.

- Aizenman, J., & Glick, R. (2005). *Pegged exchange rate regimes a trap?* NBER Working Papers 11652. National Bureau of Economic Research, Inc.
- Aizenman, J., & Marion, N. P. (1999). *Uncertainty and the disappearance of international credit.* NBER Working Papers 7389. National Bureau of Economic Research, Inc.
- Aizenmann, J., & Lee, J. (2005). *International reserves: Precautionary vs. mercantilist views, theory and evidence.* IMF Working Paper WP/05/198.
- Aizenmann, J., & Marion, N. (2004). International reserve holdings with sovereign risk and costly tax collection. *The Economic Journal*, 114(497), 569–591.
- Alcalá, F., & Ciccone, A. (2004). Trade and productivity. *Quarterly Journal of Economics*, 119(2), 613–646.
- Alesina, A. (1988). Credibility and policy convergence in a two-party system with rational voters. *American Economic Review, American Economic Association*, 78(4), 496–805, September and Finance, 46(2), 190–210.
- Alesina, A., & Summers, L. H. (1993). Central bank independence and macroeconomic performance: Some comparative evidence. *Journal of Money, Credit and Banking*, 25(2), 151–162.
- Alesina, A., & Wagner, A. (2006). Choosing (and reneging on) exchange rate regimes. *Journal of the European Economic Association*, 4(4), 770–799.
- Armas, A., & Grippa, F. (2006). Targeting inflation in a dollarized economy economy: The Peruvian experience. In A. Armas, A. Ize & E. Levy Yeyati (Eds.), *Financial dollarization: The policy agenda*. New York: Palgrave Macmillan.
- Ayyagari, M., Demirgüç-Kunt, A., & Maksimovic, V. (2006). *How important are financing constraints? The role of finance in the business environment.* Policy Research Working Paper No. 3820. Washington, DC: World Bank.
- Bacchetta, P., & van Wincoop, E. (2000). Does exchange-rate stability increase trade and welfare? *American Economic Review*, 90(5), 1093–1109.
- Backus, D., & Driffill, J. (1985). *Credibility and commitment in economic policy.* CEPR Discussion Papers 63. C.E.P.R. Discussion Papers (other versions).
- Backus, D., & Kehoe, P. (1992). International evidence on the historical properties of business cycles. *The American Economic Review*, 82(4), 864–888.
- Backus, D., Kehoe, P. J., & Kydland, F. E. (1992). International real business cycles. *Journal of Political Economy*, 100(4), 745–775.
- Bailliu, J., Lafrance, R., & Perrault, J. (2003). Does exchange rate policy matter for growth? *International Finance*, 6, 381–414.
- Balino, T., & Enoch, C. (1997). *Currency board arrangements: Issues and experiences.* IMF Occasional Paper No. 151.
- Ball, L. M. (1999). Efficient rules for monetary policy. *International Finance*, 2(1), 63–83.
- Ball, L. M., & Sheridan, N. (2003). *Does inflation targeting matter?* IMF Working Papers 03/129, IMF.
- Banerjee, A. V., & Duflo, E. (2004). *Do firms want to borrow more? Testing credit constraints using a directed lending program.* CEPR Discussion Paper No. 4681.
- Barro, R. (1999, March 8). Let the dollar reign from Seattle to Santiago. *Wall Street Journal*.
- Barro, R., & Gordon, D. (1983). A positive theory of monetary policy in a natural rate model. *Journal of Political Economy*, 91(3), 589–610.
- Barro, R., & Tenreyro, S. (2007). Economic effects of currency unions. *Economic Inquiry, Western Economic Association International*, 45(1), 1–23.
- Baxter, M., & Stockman, A. (1989, May). Business cycles and the exchange-rate regime: Some international evidence. *Journal of Monetary Economics*, 23(3), 377–400.
- Bayoumi, T., & Barry, E. (1998). Exchange rate volatility and intervention: Implications of the theory of optimum currency areas. *Journal of International Economics*, 45(2), 191–209.
- Beck, T., Demirgüç-Kunt, A., Laeven, L., & Levine, R. (2005). *Finance, firm size and growth.* World Bank Policy Research Working Paper No. 3485. Washington, DC: World Bank.
- Beck, T., Levine, R., & Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics*, 58, 261–300.
- Bénassy-Quéré, A., & Coeuré, B. (2001). *On the identification of de facto currency pegs.* CEPII, Mimeo.
- Ben-David, D. (1993). Equalizing exchange: Trade liberalization and income convergence. *Quarterly Journal of Economics*, 108(3), 653–679.

- Benigno, G., & Benigno, P. (2003). Price stability in open economies. *Review of Economic Studies*, 70(4), 743–764.
- Berger, H., & Nitsch, V. (2005). *Zooming out: The trade effect of the Euro in historical perspective*. Cesifo Working Paper No. 1435.
- Betts, C., & Devereaux, M. B. (2000). Exchange rate dynamics in a model of pricing-to-market. *Journal of International Economics*, 50, 215–244.
- Boyd, J. H., Levine, R., & Smith, B. D. (2001). The impact of inflation on financial sector performance. *Journal of Monetary Economics*, 47(2), 221–248.
- Brada, J., & Mendez, J. (1988). Exchange rate risk, exchange rate regimes, and the level of international trade. *Kyklos*, 41(2), 198.
- Broda, C. (2001, May). Coping with terms of trade shocks: Pegs vs floats. *American Economic Review (Papers and Proceedings)*, 91(2), 376–380.
- Broda, C. (2003). Terms of trade and exchange rate regimes in developing countries. *Journal of International Economics*. Forthcoming. (See also Federal Reserve Bank of New York Staff Reports, No. 148.).
- Broda, C., & Levy Yeyati, E. (2003). *Dollarization and the lender of last resort, in dollarization: Debates and policy alternatives*. MIT Press.
- Broda, C., & Weinstein, D. (2006, May). Globalization and the gains from variety. *Quarterly Journal of Economics*, 121(No. 2).
- Broda, C., & Levy Yeyati, E. (2006). Endogenous deposit dollarization. *Journal of Money, Credit and Banking*, 38(4), 963–988.
- Bruno, M., Di Tella, G., Dornbusch, R., & Fischer, S. (1988). *Inflation stabilization: The experience of Israel, Argentina, Brazil, Bolivia, and Mexico*. MIT Press.
- Bryant, R., Hooper, P., & Mann, C. (1993). *Evaluating policy regimes: New research in empirical macroeconomics*. Washington, DC: The Brookings Institution.
- Burnside, C., Eichenbaum, M., & Rebelo, S. (2001). Hedging and financial fragility in fixed exchange rate regimes. *European Economic Review*, 45(7), 1151–1193.
- Caballero, R., & Cowan, K. (2006). *Financial integration without the volatility*. Working Papers Central Bank of Chile 387. Central Bank of Chile.
- Caballero Ricardo, J. (2000). *Macroeconomic volatility in Latin America: A view and three case studies*. NBER Working Papers 7782. National Bureau of Economic Research, Inc.
- Caballero, R., & Lorenzoni, G. (2006). *Persistent appreciations and overshooting: A normative analysis*. MIT and NBER, Mimeo.
- Calvo, G. (1978). On the time consistency of optimal policy in a monetary economy. *Econometrica*, 46(6), 1411–1428.
- Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12, 383–398.
- Calvo, G. (1999). *Fixed versus flexible exchange rates: Preliminaries of a turn-of-millennium rematch*. University of Maryland, Mimeo.
- Calvo, G. (2000). *The case for hard pegs in the brave new world of global finance*. University of Maryland, Mimeo.
- Calvo, G. A., & Végh, C. A. (1993). *Exchange rate-based stabilization under imperfect credibility in open-economy macroeconomics*, by Helmut Frisch and Andreas Worgötter (pp. 3–28). London: MacMillan Press.
- Calvo, G., Izquierdo, A., & Loo-Kung, R. (2005). Relative price volatility under sudden stops: The relevance of balance-sheet effects. *Journal of International Economics*, June 2006, also NBER Working Paper No. 11492.
- Calvo, G., & Reinhart, C. (2002). Fear of floating. *The Quarterly Journal of Economics*, 117, 379–408.
- Calvo, G., & Végh, C. (1994). Inflation stabilization and nominal anchors. *Contemporary Economic Policy*, 12(2), 35–45.
- Canavan, C., & Tommasi, M. (1997). On the credibility of alternative exchange rate regimes. *Journal of Development Economics*, 54(1), 101–122.
- Cavallo, E., & Frankel, J. (2004). *Does openness to trade make countries less vulnerable to sudden stops? Using gravity to establish causality*. NBER Working Paper No., 10957.
- Cecchetti, S. G., & Ehrmann, M. (1999). *Does inflation targeting increase output volatility? An international comparison of policymakers: Preferences and outcomes*. NBER Working Paper No. 7426.

- Chamón, M. (2001). *Foreign currency denomination of foreign debt: Has the ‘original sin’ been forgiven but not forgotten?*. Harvard University, Mimeo.
- Chamon, M., & Irineu de Carvalho, F. (2006). *The myth of post-reform income stagnation in Brazil*. IMF Working Papers 06/275. International Monetary Fund.
- Chari, V. V., Kehoe, P., & McGrattan, E. (2002). Monetary shocks and real exchange rate in sticky price models of international business cycles. *Review of Economic Studies*, 69, 533–563.
- Chari, V. V., Kehoe, P., & McGrattan, E. R. (2005). *Sudden stops and output drops*. NBER Working Papers 11133. National Bureau of Economic Research, Inc.
- Clarida, R., Galí, J., & Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics*, 115(1), 147–180.
- Clarida, R., Galí, J., & Gertler, M. (2001). Optimal monetary policy in open vs. closed economies: An integrated approach. *American Economic Review*, 91(2), 248–252.
- Clarida, R., Galí, J., & Gertler, M. (2002). A simple framework for international monetary policy analysis. *Journal of Monetary Economics*, 49(5), 879–904.
- Clausing, K. (2000, April). Does multinational activity displace trade? *Economic Inquiry*, 38(2), 190–205 (16). Oxford University Press.
- Collins, S. (1996). On becoming more flexible: Exchange rate regimes in Latin America and the Caribbean. *Journal of Development Economics*, 51, 117–138.
- Corbo, V., Landerretche, O., & Schmidt-Hebbel, K. (2001, October). Assessing inflation targeting after a decade of world experience. *International Journal of Finance & Economics*, 6(4), 343–368.
- Corbo, V., & Schmidt-Hebbel, K. (2001, September). *Inflation targeting in Latin America*. Working Paper No. 105. Central Bank of Chile.
- Corsetti, G., & Pesenti, P. (2001). Welfare and macroeconomic interdependence. *Quarterly Journal of Economics*, CXVI(2), 421–446.
- Corsetti, G., & Pesenti, P. (2005). International dimensions of optimal monetary policy. *Journal of Monetary Economics*, 52(2), 281–305.
- Cukierman, A., & Meltzer, A. H. (1986). A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica, Econometric Society*, 54(5), 1099–1128.
- Cukierman, A., Webb, S. B., & Neyapti, B. (1992). Measuring the independence of central banks and its effect on policy outcomes. *World Bank Economic Review*, 6(3), 353–398.
- Culp, C., Hanke, S., & Miller, M. (1999). The case for an Indonesian Currency Board. *Journal of Applied Corporate Finance*, 11(4), 57–65.
- De Grauwe, P. (1988). Exchange rate variability and the slowdown in growth of international trade. *IMF Staff Papers*, 35, 63–84.
- De Grauwe, P., & Schnabl, G. (2005). Nominal versus real convergence—EMU entry scenarios for the new member states. *Kyklos*, 58(4), 537–555.
- De Gregorio, J., Guidotti, P., & Végh, C. (1998). Inflation stabilization and the consumption of durables. *Economic Journal*, 108(446), 105–131.
- De la Torre, A., Levy Yeyati, E., & Schmukler, S. (2002). *Argentina’s financial crisis: Floating money, sinking banking* (pp. 1–20). World Bank, Mimeo.
- De Nardis, S., De Santis, R., & Vicarelli, C. (2007). *The Euro’s effect on trade on a dynamic setting*. ISAE Working Papers 80. Rome, Italy: ISAE Institute for Studies and Economic Analyses.
- De Nicolo, G., Honohan, P., & Ize, A. (2005). Dollarization of bank deposits: Causes and consequences. *Journal of Banking & Finance*, 29(7), 1697–1727.
- De Sousa, J., & Lochard, J. (2004). *The currency union effect on trade and the FDI channel*. University of Paris and University of Rennes, Mimeo.
- Díaz Alejandro, C. (1965). *Exchange rate devaluation in a semi-industrialized country: The experience of Argentina 1955–61*. MIT Press.
- Dollar, D. (1992). Outward-oriented developing economies really do grow more rapidly: Evidence from 95 LDCs, 1976–85. *Economic Development and Cultural Change*, 40(No. 3), 523–544.
- Dooley, P. M., Folkerts-Landau, D., & Garber, P. (2003). *An essay on the revived Bretton Woods system*. NBER Working Papers 9971. National Bureau of Economic Research, Inc.
- Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy*, 84, 1161–1176.

- Dornbusch, R. (2001). Fewer monies better monies, discussion on exchange rates and the choice of monetary-policy regimes. *The American Economic Review*, 91(No. 2), 238–242.
- Dubas, J., Lee, B., & Mark, N. (2005). *Effective exchange rate classifications and growth*. NBER Working Paper Series No. 11272.
- Edwards, S. (1989). *Real exchange rates, devaluation, and adjustment: Exchange rate policies in developing countries*. Cambridge, MA: MIT Press.
- Edwards, S. (1996, September). The determinants of the choice between fixed and flexible exchange-rate regimes. *NBER*.
- Edwards, S. (1998). Openness, productivity and growth: What do we really know? *Economic Journal*, 108 (447), 383–398.
- Edwards, S., & Levy Yeyati, E. (2005). Flexible exchange rates as shock absorbers. *European Economic Review*, 49(8), 2079–2105.
- Eichengreen, B. (1994). *International monetary arrangements for the 21st century*. Washington, DC: Brookings Institution.
- Eichengreen, B. (2006a). *Global imbalances and the lessons of Bretton Woods*. MIT Press.
- Eichengreen, B. (2006b). *Insurance underwriter or financial development fund: What role for reserve pooling in Latin America?* NBER Working Papers 12451. National Bureau of Economic Research, Inc.
- Enoch, C., & Gulde, A. (1998). Are currency boards a cure for all monetary problems? *Finance and Development*, 35(No. 4), 40–43.
- Fatas, A., & Rose, A. (2000). *Do monetary handcuffs restrain leviathan? Fiscal policy in extreme exchange rate regimes*. Discussion Paper CEPR No. 2692.
- Fischer, S. (2000). Exchange rate regimes: Is the bipolar view correct? *Journal of Economic Perspectives*, 15, 3–24.
- Flood, R. P., & Rose, A. K. (1995). Fixing exchange rates: A virtual quest for fundamentals. *Journal of Monetary Economics*, 36(1), 3–37.
- Fraga, A., Goldfajin, I., & Minella, A. (2003). *Inflation targeting in emerging market economies*. NBER Working Paper No. 10019.
- Frankel, J. (2005). *Comments on Richard Baldwin's The Euro's trade effects*. Kennedy School of Government, Mimeo.
- Frankel, J. A., & Rose, A. K. (1998). The endogeneity of the optimum currency area criteria. *Economic Journal*, 108(449), 1009–1025.
- Frankel, J., & Romer, D. (1999). Does trade cause growth? *American Economic Review*, 89(3), 379–399.
- Frankel, J., & Rose, A. (2002). An estimate of the effect of currency unions on trade and output. *Quarterly Journal of Economics*, CXVII(2), 437–466.
- Frieden, J., Ghezzi, P., & Stein, E. (2000). *Politics and exchange rates: A cross-country approach to Latin America inter-American development bank*. Working Paper No. R-421.
- Friedman, M. (1953). *The case for flexible exchange rates, in essays in positive economics*. University of Chicago Press.
- Gavin, M., & Hausmann, R. (1996). *Securing stability and growth in a shock prone region: The policy challenge for Latin America inter-American development bank*. Working Paper No. 315.
- Ghosh, A., Gulde, A. M., Ostry, J., & Wolf, H. (1997). *Does the nominal exchange rate matter?* NBER Working Paper No. 5874.
- Ghosh, A., Gulde, A., & Wolf, H. (1998). Currency boards: The ultimate fix? *International Monetary Fund*. Working Paper No. 98/8.
- Ghosh, A., Gulde, A., & Wolf, H. (2003, June). *Exchange rate regime: Choices and consequences*. Cambridge and London: MIT Press.
- Giavazzi, F., & Pagano, M. (1988). The advantage of tying one's hands: EMS discipline and central bank credibility. *European Economic Review*, 32(5), 1055–1075.
- Glick, R., & Rose, A. K. (2002). Does a currency union affect trade? The time series evidence. *European Economic Review*, 46, 1125–1151.
- Grilli, V., Masciandaro, D., & Tabellini, G. (1991). Political and monetary institutions and public financial policies in the industrial countries. Vol 6, In *Economic Policy*, (13), Oxford: Reino Unido, Blackwell Publishing, 342–392.

- Hausmann, R., Gavin, M., Pages, C., & Stein, E. (1999). *Financial turmoil and the choice of exchange rate regime*. Working Paper No. 400. Inter-American Development Bank, Research Department.
- Hausmann, R., Pritchett, L., & Rodrik, D. (2005). Growth accelerations. *Journal of Economic Growth*, 10(4), 303–329.
- Hausmann, R., & Rigobon, R. (2003). *IDA in UF: On the benefits of changing the currency denomination of concessional lending to low-income countries*. The World Bank (unpublished manuscript).
- Hausmann, R., & Rodrik, D. (2003). Economic development as self-discovery. *Journal of Development Economics*, 72(2), 603–633.
- Haussmann, R., Panizza, U., & Stein, E. (2001). Why do countries float the way they float? *Journal of Development Economics*, 66, 387–414.
- Haussmann, R., & Velasco, A. (2002). *Hard money's soft underbelly: Understanding the Argentine crisis* (p. 64). Harvard University/KSG.
- Ize, A., & Levy Yeyati, E. (2003). Financial dollarization. *Journal of International Economics*, 59, 323–347.
- Jeanne, O. (2005). Why do emerging economies borrow in foreign currency? In B. Eichengreen & R. Hausmann (Eds.), *Other people's money: Debt denomination and financial instability in emerging market economies*. University of Chicago Press, 190–217.
- Jeanne, O., & Ranciere, X. (2006). *The optimal level of international reserves for emerging market countries: Formulas and applications*. IMF Working Paper No. 06/229.
- Johnson, H. (1956). The transfer problem and exchange stability. *The Journal of Political Economy*, 64(3), 212–225.
- Johnson, S., Ostry, J., & Subramanian, A. (2006, March). Levers for growth: Policy lessons from earlier bouts of growth in developing countries. *Finance and Development*. Washington, DC: International Monetary Fund.
- Juhn, G., & Mauro, P. (2002). *Long-run determinants of exchange rate regimes: A simple sensitivity analysis*. Working Paper, IMF WP-02-104.
- Kalecki, M. (1939). Money and real wages. In J. Osiatinsky (Ed.), *Collected works of Michal Kalecki*. (Reprinted ed., Vol. 2, p. 1991). Oxford: Clarendon Press.
- Kenen, P. (2002). Currencies, crises, and crashes. *Eastern Economic Journal*, 28(1), 1–12, Winter.
- Khan, M. S., Senhadji, A. S., & Smith, B. D. (2006). Inflation and financial depth. *Macroeconomic Dynamics*, 10(02), 165–182.
- Kiguel, M., & Leviatan, N. (1991). The inflation-stabilization cycles in Argentina and Brazil. In M. Bruno, et al. (Ed.), *Lessons of economic stabilization and its aftermath* (pp. 91–232). Cambridge, MA: MIT Press.
- Kiguel, M., & Leviatan, N. (1992). The business cycle associated with exchange rate-based stabilizations. *The World Bank Economic Review*, 6(2), 279–305.
- Klapper, L., Laeven, L., & Rajan, R. (2006). Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics*, 82, 591–629.
- Klein, M. (2002). *Dollarization and trade*. Tufts University, Mimeo.
- Kollmann, R. (2001). The exchange rate in a dynamic optimizing current account model with nominal rigidities: A quantitative investigation. *Journal of International Economics*, 55, 243–262.
- Kollmann, R. (2002). Monetary policy rules in the open economy: Effects on welfare and business cycles. *Journal of Monetary Economics*, 49, 989–1015.
- Kose, A., Prasad, E., Rogoff, K., & Wei, S. J. (2006). *Financial globalization: A reappraisal*. IMF Working Paper No. 06/189.
- Kose, M. A., Prasad, E. S., & Terrones, M. (2003). Financial integration and macroeconomic volatility. *IMF Staff Papers*, 50(special issue), 119–142.
- Kuttner, K., & Posen, A. (2001). Beyond bipolar: A three-dimensional assessment of monetary frameworks. *International Journal of Finance & Economics*, 6(4), 369–387.
- Kydland, F., & Prescott, E. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85, 473–490.
- Kydland, F., & Prescott, E. (1986, Spring). Business cycles: Real facts and a monetary myth. *Quarterly Review*, 14(2), 3–18 (Federal Reserve Bank of Minneapolis).

- Laeven, L., & Woodruff, C. (2007). The quality of the legal system, firm ownership, and firm size. *The Review of Economics and Statistics*, 89(4), 601–614.
- Lee, H. Y., Ricci, L. A., & Rigobon, R. (2004, December). Once again, is openness good for growth? *Journal of Development Economics*, 75(2), 451–472.
- Levine, R. (2005). Finance and growth: Theory and evidence. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth*, Elsevier, (Vol. 1, Chapter 12, (pp. 865–934).
- Levy Yeyati, E. (2005). *Exchange rate regimes in the 2000s: A Latin American perspective*. A New Economic Development Agenda for Latin America, Fundación Cidob.
- Levy Yeyati, E. (2006). Financial dollarization: Evaluating the consequences. *Economic Policy*, 21(45), 61–118.
- Levy Yeyati, E., & Sturzenegger, F. (2001). Exchange rate regimes and economic performance. *IMF Staff Papers*, 47, 62–98.
- Levy Yeyati, E., & Sturzenegger, F. (2003). To float or to fix: Evidence on the impact of exchange rate regimes on growth. *American Economic Review*, 93(4), 1173–1193.
- Levy Yeyati, E., & Sturzenegger, F. (2005). Classifying exchange rate regimes: Deeds vs. words. *European Economic Review*, 49, 1603–1635.
- Levy Yeyati, E., & Sturzenegger, F. (2006a). *Dollarization*. MIT Press.
- Levy Yeyati, E., & Sturzenegger, F. (2006b). *On the endogeneity of exchange rate regimes*. Kennedy School of Government Working 06-047. Harvard University.
- Levy Yeyati, E., & Sturzenegger, F. (2007). *Fear of appreciation*. Kennedy School of Government Working Paper 07-047. Harvard University.
- Lipsey, R. E., & Weiss, M. Y. (1981). Foreign production and exports in manufacturing industries. *The Review of Economics and Statistics*, 63(4), 488–494.
- Love, I., & Zicchino, L. (2006). Financial development and dynamic investment behavior: Evidence from panel vector autoregression. *Quarterly Review of Economics and Finance*, 46(2), 190–210.
- Lubik, T. A., & Schorfheide, F. (2007). Do Central Banks respond to exchange rate movements? A structural investigation. *Journal of Monetary Economics*, 54(4), 1069–1087.
- McCallum, B. T., & Nelson, E. (2000). Monetary policy for an open economy: An alternative framework with optimizing agents and sticky prices. *Oxford Revue of Economic Policy*, 16(4), 74–91.
- Meade, J. E. (1951). *The Theory of International Economic Policy*. Vol. 2: *Trade and Welfare*, London: Oxford University Press.
- Mendoza, E. G. (1991). Real business cycles in a small open economy. *American Economic Review*, 81, 797–818.
- Micco, A., Stein, E., & Ordoñez, G. (2003). The currency union effect on trade: Early evidence from EMU. *Economic Policy*, 37, 315–356.
- Mishkin, F. (2004). *Can inflation targeting work in emerging market countries*. Working Paper No. 10646. National Bureau of Economic Research.
- Mishkin, F. S. (2006). *Monetary policy strategy: How did we get there?* NBER Working Paper No. 12515.
- Mishkin, F. S. (2007). *Inflation dynamics*. NBER Working Paper No. 13147.
- Mishkin, F., & Schmidt-Hebbel, K. (2001). *One decade of inflation targeting in the world: What do we know and what do we need to know?* NBER Working Paper No. 8397.
- Mundell, R. (1963). Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economics and Political Science*, 29, 475–485.
- Neumeyer, P. A., & Perri, F. (2004). Business cycles in emerging economies: The role of interest rates. *Journal of Monetary Economics*, 52, 345–380.
- O'Donnell, B. (2001, unpublished). *Financial openness and economic performance*. Dublin: Trinity College.
- Obstfeld, M. (1996). Models of currency crises with self-fulfilling features. *European Economic Review*, 40, 1037–1047.
- Obstfeld, M., & Kenneth, R. (1995, July). *The mirage of fixed exchange rates*. Working Paper No. 5191. Cambridge, MA: NBER.
- Obstfeld, M., & Rogoff, K. (2000). *New directions for stochastic open economy models*. International Finance, 0004002, EconWPA.
- Obstfeld, M., & Taylor, A. (2004). *Global capital markets: Integration, crisis and growth* (pp. 4–41). Cambridge University Press.

- Ohkawa, K., & Rosovsky, H. (1973). *Japanese economic growth*. Stanford: Stanford University Press.
- Orphanides, A. (2001a, September). Monetary policy rules based on real-time data. *American Economic Review*, 91(4), 964–985.
- Orphanides, A. (2001b, December). Monetary policy rules, macroeconomic stability and inflation: A view from the trenches. Board of Governors of the Federal Reserve System.
- Parrado, E., & Velasco, A. (2002). *Optimal interest rate policy in small open economy*. NBER Working Paper No. W8721.
- Parsley, D. C., & Wei, S. J. (2001a). Explaining the border effect: The role of exchange rates, shipping costs, and geography. *Journal of International Economics*, 55(1), 87–105.
- Parsley, D. C., & Wei, S. J. (2001b, September). *Limiting currency volatility to stimulate goods market integration: A price based approach*. NBER Working Paper No. 8468.
- Persson, T. (2001). *Currency unions and trade: How large is the treatment effect?* Institute for International Economic Studies.
- Persson, T., & Tabellini, G. (1993). Designing institutions for monetary stability. *Carnegie-Rochester Conference Series on Public Policy*, 39, 53–84.
- Poirson, H. (2001, April). *How do countries choose their exchange rate regime?* IMF Working Paper No. 01/46.
- Prasad, E., Rajan, R. G., & Arvind, S. (2006). *Patterns of international capital flows and their implications for economic development*. Proceedings (pp. 119–158). Federal Reserve Bank of Kansas City.
- Rajan, R., & Arvind, S. (2005). *What undermines aid's impact on growth?* IMF Working Papers 05/126. International Monetary Fund.
- Rajan, R., & Zingales, L. (1998). Financial dependence and growth. *American Economic Review*, 88, 559–587.
- Ramcharan, R. (2005). *Cataclysms and currencies: Does the exchange rate regime matter for real shocks?* IMF Working Paper No. 05/85.
- Ramey, G., & Ramey, V. (1995). Cross-country evidence on the link between volatility and growth. *American Economic Review*, 85(5), 1138–1151.
- Razin, O., & Collins, S. (1997). *Real exchange rate misalignments and growth*. NBER Working Paper No. 6174.
- Reinhart, C. M., & Rogoff, K. (2004). The modern history of exchange rate arrangements: A reinterpretation. *Quarterly Journal of Economics*, 119, 409–436.
- Rigobon, R. (2000, September). *A simple test for stability of linear models under heteroskedasticity, omitted variable, and endogenous variable problems*. MIT, Mimeo.
- Rigobon, R., & Rodrik, D. (2004, September). *Rule of law, democracy, openness, and income: Estimating the interrelationships*. NBER Working Paper No. 10750.
- Rioja, F., & Valev, N. (2004a). Does one size fit all? A reexamination of the finance and growth relationship. *Journal of Development Economics*, 74(2), 429–447.
- Rioja, F., & Valev, N. (2004b). Finance and the sources of growth at various stages of economic development. *Economic Inquiry*, 42(1), 127–140.
- Rizzo, J. (1998, June). The economic determinants of the choice of an exchange rate regime: A probit analysis. *Economics Letters*, 59(3), 283–287.
- Rodrik, D. (2006a). *Industrial development: Stylized facts and policy*. Kennedy School, Mimeo.
- Rodrik, D. (2006b). *The social cost of foreign exchange reserves*. NBER Working Paper No. 11952.
- Rodrik, D., & Rodríguez, F. (2000). Trade policy and economic growth: A skeptics guide to the cross-national evidence. In B. Bernanke & K. Rogoff (Eds.), *NBER macroeconomics annual* (Vol. 15). Cambridge, MA: MIT Press.
- Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9, 131–165.
- Rogoff, K. (1985). The optimal degree of commitment to an intermediate monetary target. *Quarterly Journal of Economics*, 100, 1169–1190.
- Romer, D. (1993). Openness and inflation: Theory and evidence. *The Quarterly Journal of Economics*, CVIII, 869–903.
- Rose, A. (2000). One money, one market? The effects of common currencies on international trade. *Economic Policy*, 15(30), 7–46.

- Rose, A. (2004). Do we really know that the WTO increases trade? *American Economic Review*, 94, 98–114.
- Rose, A. (2006). *A stable international monetary system emerges: Bretton Woods*. NBER Working Paper No. 12711.
- Rose, A., & Van Wincoop, E. (2001). National money as a barrier to international trade: The real case for currency union. *American Economic Review, Papers and Proceedings*, 91(2), 386–390.
- Sachs, J. (1985). External debt and macroeconomic performance in Latin America and East Asia. *Brookings Papers on Economic Activity*, 2, 523–573.
- Sachs, J., & Warner, A. (1995). Economic reform and the process of global integration. *Brookings Papers on Economic Activity*, 1, 1–118.
- Schmidt-Hebbel, K., & Tapia, M. (2002). Inflation targeting in Chile. *The North American Journal of Economics and Finance*, 13(2), 125–146.
- Schmitt-Grohe, S., & Uribe, M. (2001). Stabilization policy and the costs of dollarization. *Journal of Money, Credit and Banking*, 33(2), 482–509.
- Shambaugh, J. (2004). The effects of fixed exchange rates on monetary policy. *Quarterly Journal of Economics*, 119, 301–352.
- Sterne, G. (1999, August). The use of explicit targets for monetary policy: Practical experiences of 91 economies in the 1990s. *The Bank of England Quarterly Bulletin*, 272–281.
- Sturzenegger, F. (2008). “Currency boards”. The New Palgrave Dictionary of Economics, 2nd edition, Eds. Steven N. Durlauf and Lawrence N. Blume, Palgrave Macmillan, 2008.
- Sturzenegger, F., & Talvi, E. (2008). *Unveiling monetary policy in Latin America*. Universidad Torcuato Di Tella, Mimeo.
- Summers, L. (2006, March). *Reflections on global account imbalances and emerging markets reserve accumulation. L.K. Jha Memorial Lecture, Reserve Bank of India, Mumbai*.
- Svensson, L. (1996). pp. 40.
- Svensson, L. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European Economic Review*, 41, 1111–1146.
- Svensson, L. (2000). The first year of the Eurosystem: Inflation targeting or not? *American Economic Review*, 90(2), 95–99.
- Svensson, L. E. (1995). *Optimal inflation targets, ‘conservative’ Central Banks, and linear inflation contracts*. CEPR Discussion Papers 1249.
- Tavlas, G., Dellas, H., & Stockman, A. (2006). *The classification and performance of alternate exchange rate systems*. Bank of Greece, Mimeo.
- Taylor, J. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Series on Public Policy*, 39, 195–214.
- Taylor, J. (2000). Alternative views of the monetary transmission mechanism: What difference do they make for monetary policy? *Oxford Review of Economic Policy*, 16(4), 60–73.
- Terra, C. T. (1998). Openness and inflation: A new assessment. *Quarterly Journal of Economics*, 113, 641–648.
- Thursby, J. G., & Thursby, M. C. (1987). Bilateral trade flows, the linder hypothesis, and exchange risk. *Review of Economics and Statistics*, 69, 488–495.
- Tornell, A., & Velasco, A. (2000). Fixed versus flexible exchange rates: Which provides more fiscal discipline? *Journal of Monetary Economics*, 45(2), 399–436.
- Truman, E. (2003). *Inflation targeting and the international financial system: Challenges and opportunities*. Washington: Institute for International Economics.
- Vegh, C. (1992, September). Stopping high inflation: An analytical overview. *IMF Staff Papers*, 39, 626–695.
- von Hagen, J., & Zhou, J. (2006). *Fear of floating and fear of pegging: An empirical analysis of de facto exchange rate regimes in developing countries*. CEPR Discussion Papers 5530.
- Walsh, C. (1995). Optimal contracts for central bankers. *American Economic Review*, 85(1), 150–167.
- Wei, S. J., & Choi, C. (2002). *Currency blocs and cross-border investment*. Mimeo.
- Werner, A., & Martinez, L. (2002). The exchange rate regime and the currency composition of corporate debt. The Mexican experience. *Journal of Development Economics*, 69(2), 315–334.
- Woodford, M. (2003). *Interest and prices: Foundations of theory of monetary policy*. Princeton, NJ: Princeton University Press.

This page intentionally left blank

Financial Globalization and Economic Policies*

M. Ayhan Kose

IMF, Washington, DC, USA

Eswar Prasad

Cornell University, Ithaca, NY, USA

Kenneth Rogoff

Harvard University, Cambridge, MA, USA

Shang-Jin Wei

Columbia University, New York, NY, USA

Contents

1. Introduction	4284
2. Financial Globalization: Measurement and Trends	4288
2.1 How to measure financial integration?	4288
2.2 Evolution of financial globalization: Some basic stylized facts	4289
2.3 Factors driving financial globalization	4292
3. Macroeconomic Implications of Financial Globalization	4292
3.1 Economic growth	4294
3.2 Macroeconomic volatility and patterns of international risk-sharing	4303
4. Economic Policies and Growth Outcomes	4306
5. Structural Characteristics and Economic Policies	4310
5.1 Financial sector development	4310
5.2 Institutional quality	4316
5.3 Trade openness	4323
6. Macroeconomic policies	4330
6.1 Theory	4330
6.2 Evidence	4330
7. Direction of Capital Flows and Economic Growth	4332
7.1 Empirical evidence	4332
7.2 How to interpret these findings?	4334
7.3 Are developing countries savings-constrained?	4336
7.4 Policy implications	4337
8. Capital controls as a policy tool	4337
8.1 Implementation and Measurement Issues	4337
8.2 Macroeconomic implications of capital controls	4338
8.3 Microeconomic Implications of capital controls	4339

9. How to Approach Financial Integration?	4341
10. Conclusion	4343
End Notes	4345
References	4350
Openness, Structural Features and Policies	4360
Policies Regulating Capital Account	4361
End Notes	4361
References	4362

Abstract

We review the large literature on various economic policies that could help developing economies effectively manage the process of financial globalization. Our central findings indicate that policies promoting financial sector development, institutional quality, and trade openness appear to help developing countries derive the benefits of globalization. Similarly, sound macroeconomic policies are an important prerequisite for ensuring that financial integration is beneficial. However, our analysis also suggests that the relationship between financial integration and economic policies is a complex one and that there are unavoidable inherent tensions in evaluating the risks and benefits associated with financial globalization. In the light of these tensions, structural and macroeconomic policies often need to be tailored to take into account country-specific circumstances to improve the risk-benefit tradeoffs of financial integration. Ultimately, it is essential to see financial integration not just as an isolated policy goal but also as part of a broader package of reforms and supportive macroeconomic policies.

JEL Classifications: F02, F21, F36, F4

Keywords

capital account liberalization,
 financial integration,
 growth and volatility,
 risk sharing,
 capital inflows,
 financial crises,
 developing countries

1. INTRODUCTION

Financial globalization has been one of the most intensely debated topics of our times. Some academic economists view increasing capital account liberalization and unfettered capital flows as a serious impediment to global financial stability (e.g., Bhagwati, 1998; Rodrik, 1998; Stiglitz, 2002), leading to calls for capital controls and the imposition of frictions such as “Tobin taxes” on international asset trade. In contrast, others argue that increased openness to capital flows has, by and large, proven essential for

countries aiming to upgrade from lower to middle income status, while significantly enhancing stability among industrialized countries (e.g., Fischer, 1998; Summers, 2000).

Financial globalization is clearly a matter of considerable policy relevance, especially with major economies like China and India recently taking steps to open up their capital accounts. A number of developing countries are still in the early stages of financial globalization, facing numerous ongoing policy decisions about the timing and pace of further integration. The stakes for such policy decisions are high because financial globalization is often blamed for the string of damaging economic crises that rocked a number of emerging markets in the late 1980s in Latin America and in the 1990s in Mexico, and a handful of Asian countries. The market turmoil and resulting bankruptcies prompted a rash of finger-pointing by those who suggested that developing countries had dismantled capital controls too hastily—leaving themselves vulnerable to the harsh dictates of rapid capital movements and market herd effects.

Moreover, financial globalization is also a fascinating topic to study for researchers of development economics not only because of its compelling policy relevance but also because of the enormous variation of approaches and experiences across countries. Differences in speed and approach to financial globalization have often been driven as much by philosophy, regional fads, and political circumstances as by economic factors. Hence, cross-country studies of the effects of financial integration can potentially exploit a wide array of natural variation in experiences.

There has been an explosion of research in this area over the past two decades. Most of the work is of relatively recent vintage, since the latest wave of financial globalization started in earnest only in the mid-1980s. However, the research program on financial globalization has proceeded along a number of disparate paths, with the results from some of these strands seeming at odds with each other.¹ The inconclusive nature of the debate on the merits of financial globalization reflects on the design of economic policies aiming to manage the process of financial integration. While consensus on the outcomes of financial globalization and the complex policy issues surrounding them may be too much to hope for, some clarity on what theory and data reveal—and what they do not—is important for the ongoing debate.

The objective of this chapter is to review the large literature focusing on various economic policies that could help developing economies effectively manage the process of financial globalization. In particular, we try to identify structural and macroeconomic policies that can improve the growth and stability benefits of financial globalization for developing countries.

In Section 2, we present some basic facts about the temporal evolution of financial flows. Studying policy issues surrounding the financial globalization necessarily requires an analysis of the associated measurement issues, and this section starts with a brief

summary of those. We then analyze how the volume and composition of financial flows have changed over time. The volume of flows has risen substantially during the past two decades. Not only has there been a much greater volume of flows among advanced countries over this period but there has also been a surge in flows between advanced and developing countries. There are important differences across country groups in the relative importance of different types of inflows, although there has been a broad shift away from debt financing toward FDI and equity flows in all groups.

In Section 3, we review the theoretical arguments and empirical evidence of the macroeconomic outcomes associated with financial globalization. This section largely relies on the framework put forward by Kose, Prasad, Rogoff, and Wei (2006). We focus on the implications of financial integration for the dynamics of growth, volatility, and risk-sharing patterns. Although our overall take is that the literature is still inconclusive, we argue that newer approaches that attempt to focus more on the indirect effects of financial globalization hold considerable promise. At the same time, we find that there is scant empirical support to underpin the more polemic claims of those who argue that capital account liberalizations (as opposed to, say, inappropriately rigid exchange rate regimes) are the root problem behind the financial crises faced by most developing countries in the past two decades (Bhagwati, 1998; Stiglitz, 2002).

The survey of the rapidly evolving literature on the merits of financial globalization also reveals that newer approaches depart from the standard neoclassical framework that largely guided the earlier studies. In particular, the earlier literature viewed the key benefit of financial globalization as arising from long-term net flows of capital from advanced to developing economies. As the former group of countries is capital-rich while the latter is relatively capital-poor, this should generate higher growth in developing economies and welfare gains for both groups. Perhaps not surprisingly, in light of the corresponding literature on growth in closed economies, this literature often found conflicting results.

The fundamental conceptual point that guides our interpretation of the newer literature is that the main benefits to successful financial globalization are probably catalytic and indirect. The benefits are not simply, or even primarily, the result of an enhanced access to financing for domestic investment. We document that there is modest but increasing evidence that financial openness can, in many circumstances, promote development of the domestic financial sector, impose discipline on macroeconomic policies, generate efficiency gains among domestic firms by exposing them to competition from foreign entrants, and unleash forces that result in better public and corporate governance. That is, it can generate significant indirect or “collateral” benefits which, in quantitative terms, are likely to be the most important sources of enhanced growth and stability for a country engaged in financial globalization.

The notion that financial globalization mainly influences growth through indirect channels has potentially important implications for the design of economic policies

toward financial globalization. In particular, if one can identify which reform priorities are the key ones for a particular country, then one can design an approach to liberalization that could generate specific benefits while minimizing the associated risks. This also provides a broader analytical framework within which one can incorporate country-specific features and initial conditions into the design of appropriate capital account liberalization programs.

There is also a growing literature studying a range of supporting conditions associated with structural and policy-related factors (thresholds) that appear to play an important role in the relationship between growth and financial openness. This literature argues that economic policies designed to foster these necessary supporting conditions are key in deriving better outcomes from financial globalization. Sections 4–6 provide an overview of this literature and attempt to draw some policy messages. In particular, we focus on an economy's structural features—the extent of financial sector development, institutional quality, and trade integration—and its macroeconomic policy framework. For each of these factors, we review the underlying theoretical arguments and survey the relevant empirical evidence.

Our findings suggest that economic policies promoting financial sector development, institutional quality, and trade openness are important not only in their own right, but also in helping developing countries derive the benefits of globalization. Similarly, sound macroeconomic policies appear to be an important prerequisite for ensuring that financial integration is beneficial for these countries. We also find that excessive reliance on fixed exchange rate regimes has probably been one of the major contributing factors to financial crises in emerging market countries over the past 15 years. Moving to more flexible exchange rate regimes is, therefore, likely to considerably alleviate some of the risks countries must endure as they become more financially integrated (for countries that are not financially integrated, fixed exchange rate regimes may be a perfect choice). In addition, countries that consistently face problems associated with government debt are more likely to benefit from financial globalization if their governments simultaneously take policy measures to avoid an excessive buildup of debt.

Capital has recently been flowing “uphill” from poor to rich countries. More interestingly, among nonindustrial countries there appears to be a positive correlation between a country's current account surplus and its growth rate. Section 7 studies the implications of these somewhat perverse empirical observations for economic policies in light of some recent studies. We argue that these findings are broadly consistent with the policy implications stemming from our framework of collateral benefits and threshold factors.

The next section analyzes the macroeconomic implications of capital controls. Since sudden stops and reversals of inflows of foreign capital have precipitated costly crises in some emerging market countries, capital controls have regained some of their

luster, among certain academics and policymakers, as effective policy tools to dampen the potentially adverse effects of financial integration. The evidence on the macroeconomic implications of capital controls is at best mixed while some recent studies indicate that controls appear to lead to various costs at the micro level.

In Section 9, we consider some potential approaches to financial globalization in light of the findings of some recent studies. These studies reflect the notion that financial globalization carries a short-run cost—one that must inevitably be paid if a developing country, which typically has weak institutions and a fragile financial sector, wants to move on to a high-growth path. Given that the collateral benefits perspective argues financial globalization as being a potentially useful catalyst for improving domestic institutions and financial sector, it appears that developing countries face a very complex policy problem with respect to financial integration. We argue that the collateral benefits perspective could be helpful in resolving this problem.

The final section of this chapter provides some concluding remarks and outlines a number of key questions for future research.

2. FINANCIAL GLOBALIZATION: MEASUREMENT AND TRENDS

Defining the concept of financial globalization requires us to confront a multitude of measurement problems.² Resolution of these problems is key to analyzing the implications of financial globalization as well as in designing effective policy measures to utilize its gains. After providing a brief discussion of these measurement issues, this section documents the evolution of the degree of financial globalization using a couple of well-known metrics and then summarizes the factors driving the process of financial globalization.

2.1 How to measure financial integration?

2.1.1 Capital controls (*de jure* measures)

Most of the earlier empirical studies use measures of legal restrictions (capital controls) on cross-border capital flows to assess the degree of financial openness. Such capital controls come in many varieties—controls on inflows versus those on outflows, quantity versus price controls, restrictions on foreign equity holdings, etc. Based on information from the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER), the early literature on capital account liberalization often employed a binary (0/1) measure of capital account openness. Some researchers have used a “share” measure, reflecting the fraction of years in the sample in which a country's capital account was open. Other authors have taken the detailed information in the AREAER publications to construct finer measures of capital account restrictions.³

Although there has been substantial progress in developing finer and more sophisticated measures of capital controls, all of these measures suffer from a variety of similar shortcomings. First, they do not accurately reflect the degree of openness of the capital

account because they are partially based on various restrictions associated with foreign exchange transactions that may not necessarily impede capital flows. Second, they do not capture the degree of enforcement of capital controls (or the effectiveness of this enforcement), which can change over time even if the legal restrictions remain unchanged.⁴ Third, and most importantly, these measures do not always reflect the actual degree of integration of an economy into international capital markets, as we have already noted. As another example, China, despite its extensive regime of capital controls, has not been able to stop inflows of speculative capital in recent years (see Prasad & Wei, 2007).⁵

2.1.2 Financial flows/stocks (*de facto* measures)

Quantity-based measures of integration based on actual flows appear to be the best available measure of a country's *de facto* integration with global financial markets.⁶ Should one measure integration using gross flows (the sum of total inflows and total outflows) or net flows (the difference between inflows and outflows)? Although the choice depends on the precise question one is interested in, gross flows in general provide a less volatile and more sensible picture of integration as it has the advantage of capturing two-way flows. However, annual gross flows tend to be volatile and prone to measurement error. To mitigate these problems, it is preferable to use the sum of gross stocks of foreign assets and liabilities as a ratio to GDP. This preserves the spirit of measuring *de facto* integration and obviates many of the problems associated with flow data. Moreover, for some purposes—particularly, the analyses of risk sharing—stock measures are more appropriate.

De facto measures of financial integration based on gross flows/stocks also have some drawbacks. For example, Collins (2007) argues that *de facto* indicators are likely to be endogenous in growth regressions, making it difficult to pin down causal effects. As we discuss later, *de jure* measures also have a strong element of endogeneity to them, in addition to other deficiencies. While there is important information in both the *de jure* and *de facto* measures of financial integration, *de facto* measures provide a better picture of the extent of a country's integration into global financial markets and, for many empirical applications, this measure is more suitable.

2.2 Evolution of financial globalization: Some basic stylized facts

Figure 1 displays the absolute level of integration of different country groups into global financial markets, calculated as the sum of gross international financial assets and liabilities.⁷ There has been an obvious surge in financial globalization, especially since the mid-1980s.⁸ While the level of integration is clearly the highest for advanced economies, emerging market countries have accounted for the bulk of integration experienced by developing countries. The gross stocks of assets and liabilities of this group has risen by more than fivefold and has been on average an order of magnitude larger than that of other developing countries during the past two decades.

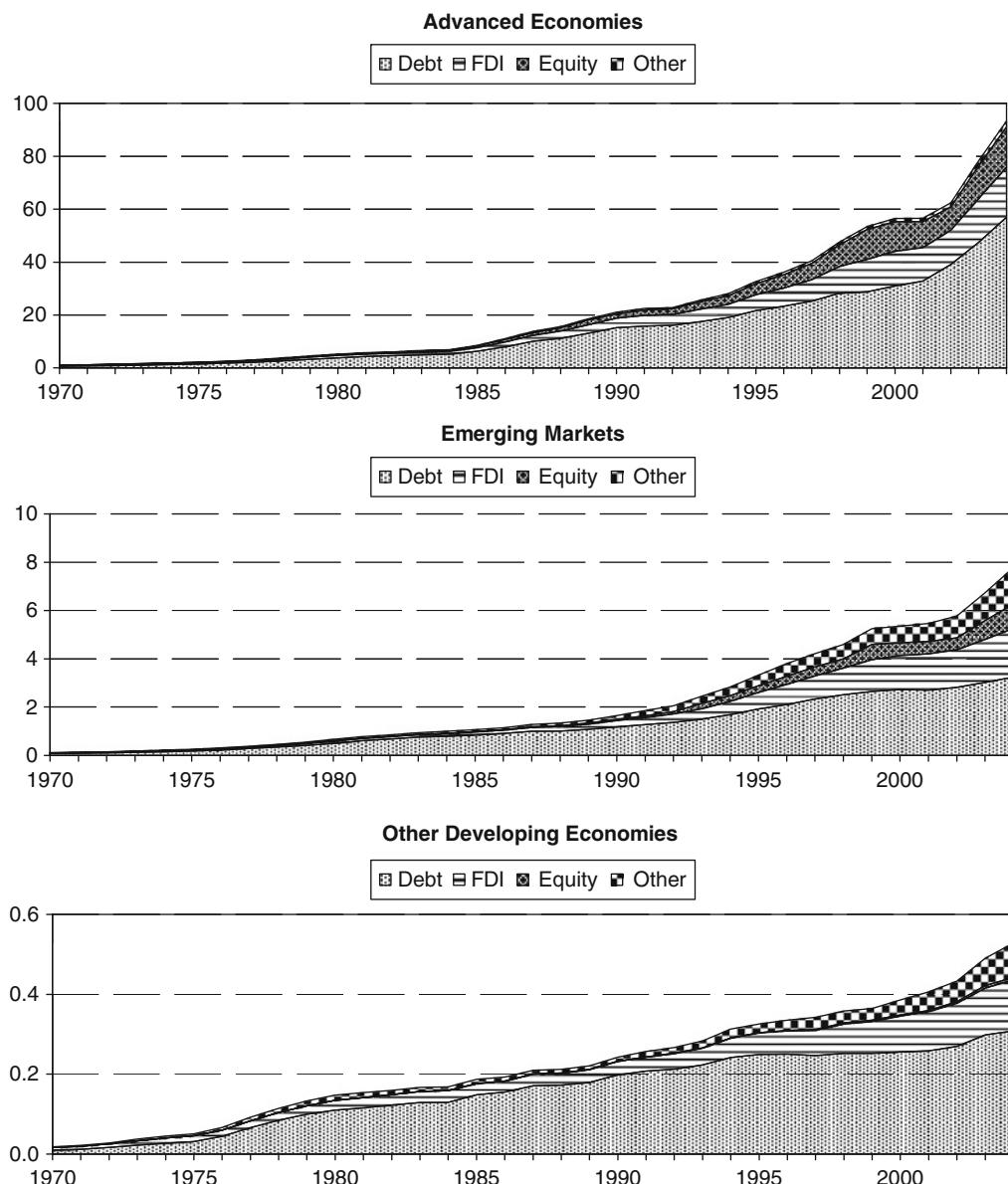


Figure 1 Gross international financial assets and liabilities (trillions of US dollars). Notes: The financial integration data are based on a dataset constructed by Lane and Milesi-Ferretti (2006). The time period of analysis is 1970-2004. The charts show how the components add up to the total integration measure in each period. Debt includes both official and unofficial debt. The category "Other" includes financial derivatives and total reserves minus gold.

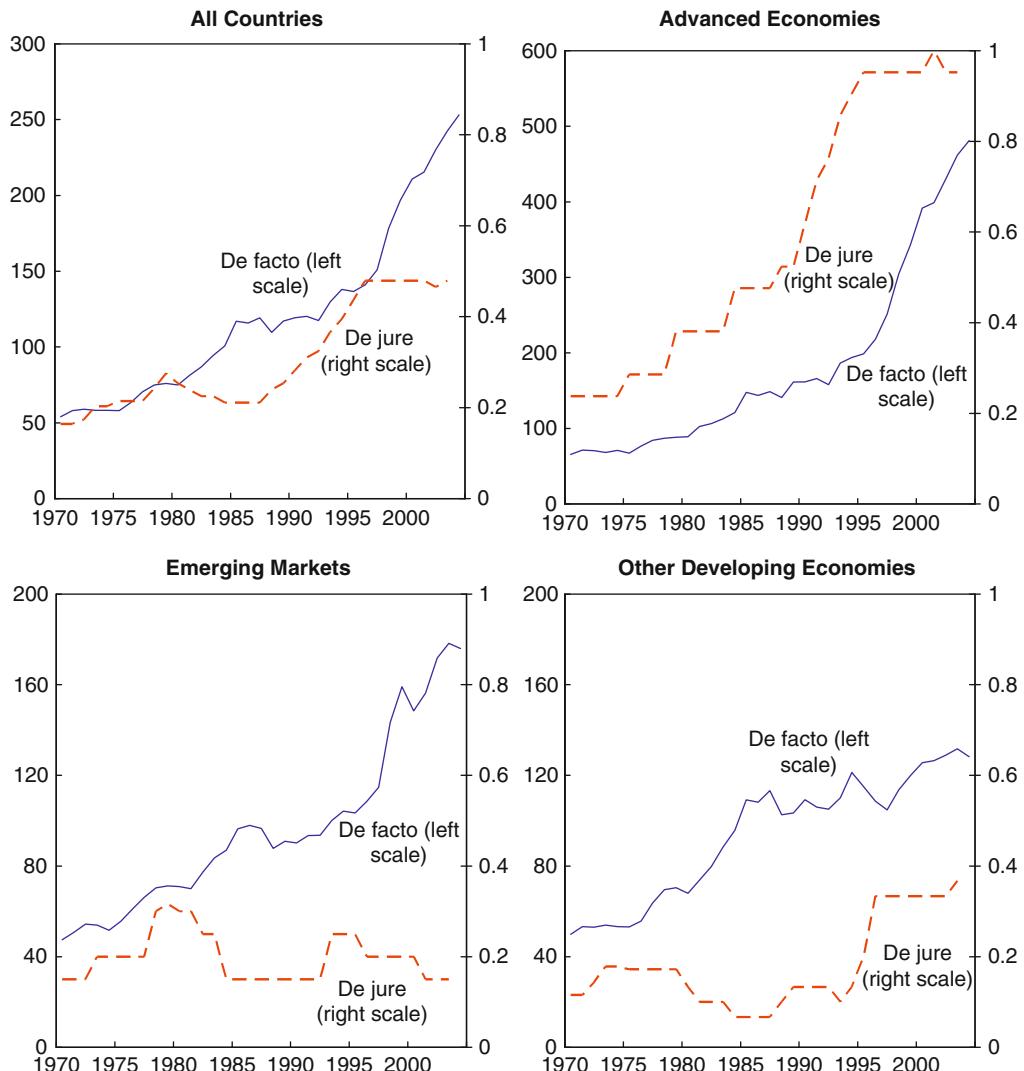


Figure 2 Evolution of international financial integration. *Notes:* This figure shows unweighted cross-country averages, within each group, of two measures of capital account openness. The time period of analysis is 1970-2004. The de jure measure is based on the IMF 0-1 capital account restrictiveness classification, with 1 representing countries that have open capital accounts. The de facto measure is based on the ratio of gross stocks of foreign assets and liabilities to GDP (in percent), with the raw data taken from Lane and Milesi-Ferretti (2006).

Figure 2 compares the evolution of *de jure* integration based on the IMF's binary capital account restrictiveness measure, averaged across all countries in each group, and corresponding group averages of the *de facto* financial openness measure (stock of international financial assets and liabilities expressed as a ratio to GDP). By both measures, advanced economies have become substantially integrated into global

financial markets. For emerging market economies, the average *de jure* openness has not changed much based on the IMF measure, but the *de facto* integration has increased sharply over the last two decades. For other developing economies, the *de jure* openness on average rose sharply over the last decade, to a level higher than that for emerging market economies, but the *de facto* measure has stayed flat over this period. This figure highlights the different informational content in the two types of integration measures and the importance of taking these differences into account in analyses of the effects of financial globalization.

Figure 3 presents the evolution of the composition of total foreign assets and liabilities for different groups of countries. Among the advanced economies, the biggest increase has been in the share of portfolio equity during the past two decades. The share of debt in gross stocks of foreign assets and liabilities of emerging market economies has declined from 75% to 50% during the same period while the share of FDI and portfolio equity has risen from a total of 13–40%. The share of portfolio equity has been rather small in the total stocks of other developing countries. Accumulation of official international reserves has recently accounted for a significant portion of the increase in gross foreign assets of developing economies. In general, these findings suggest that there has been a broad shift away from debt financing toward FDI and equity flows in all groups and some of these patterns are stronger when one looks at gross private inflows (see Kose et al., 2006).⁹

2.3 Factors driving financial globalization

The surge in financial flows to developing countries, as well as the shifts in the composition of these flows, can be broken down into “pull” and “push” factors (Calvo, Leiderman, & Reinhart, 1994). These are related to, respectively (i) policies and other developments in developing countries and (ii) changes in global financial markets. The first category includes factors such as policies with respect to capital and trade accounts, institutional quality, and governance practices and policies toward privatization of state-owned companies. For example, there has been a substantial increase in the fraction of countries with liberalized capital and trade accounts since the mid-1980s (Figure 4). Moreover, more financially integrated economies are those that have registered the largest increase in the degree of trade openness over the same period (Figure 5).¹⁰ As discussed later in the chapter, some economic policies associated with these pull factors can affect the macroeconomic outcomes of financial globalization through their impact on the volume and composition of financial flows. The second category includes the growing importance of depositary receipts and cross-listings, and the emergence of institutional investors as key players driving international capital flows to emerging markets (Prasad, Rogoff, Wei, & Kose, 2003).

3. MACROECONOMIC IMPLICATIONS OF FINANCIAL GLOBALIZATION

We begin with a brief introduction to the theoretical and empirical links between financial globalization and macroeconomic outcomes in this section. In particular, we focus

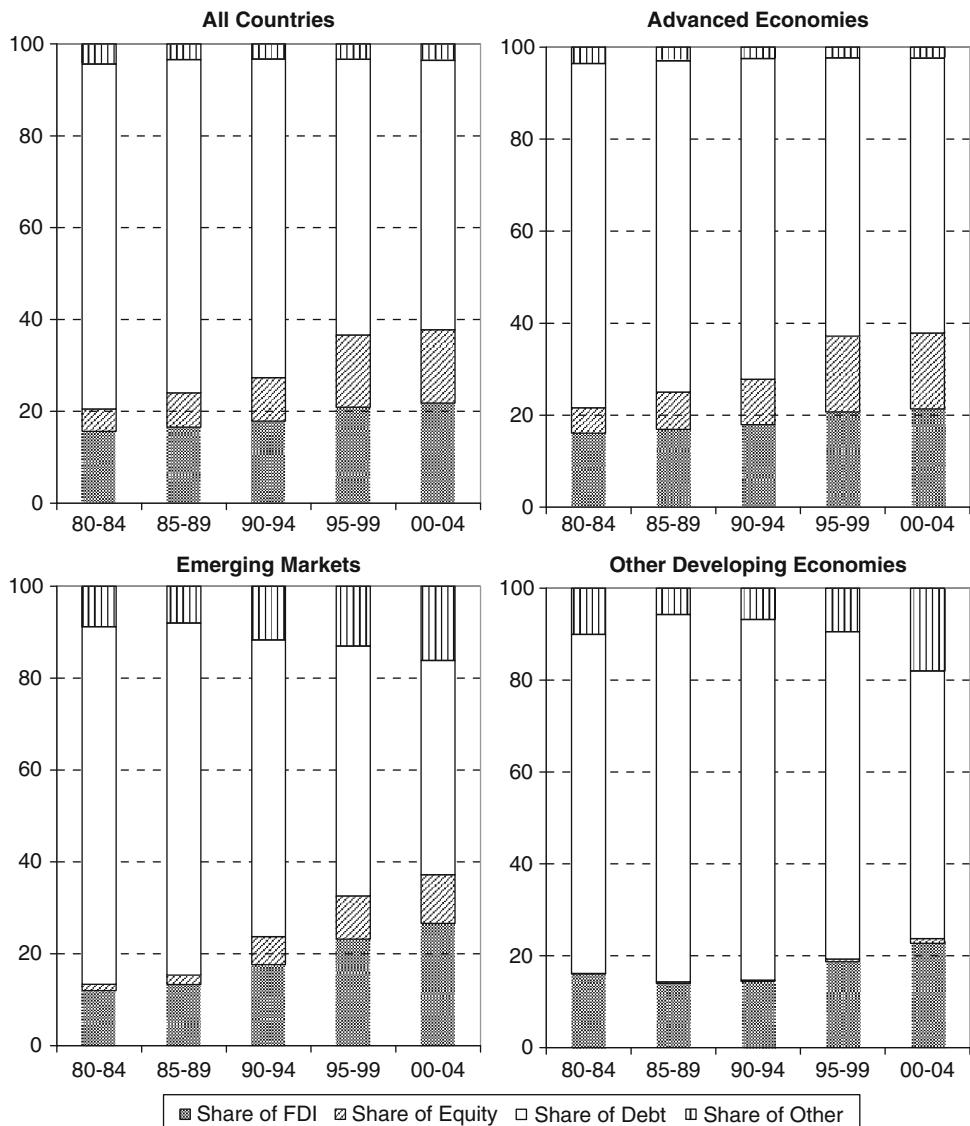


Figure 3 Composition of gross stocks of foreign assets and liabilities (in %). Notes: Data shown in this figure are based on cross-country averages of annual data over the relevant five-year period for each group of countries. The sample comprises 21 industrial, 20 emerging market and 30 other developing countries. The category "Other" includes financial derivatives and total reserves minus gold. Shares are in percentage of total. The raw data are based on a dataset constructed by Lane and Milesi-Ferretti (2006).

on the effects of financial globalization on growth, volatility, and patterns of risk-sharing. Since financial globalization has often been associated with the recent emerging market financial crises, we also analyze its impact on crises, which can be considered as special cases of volatility.

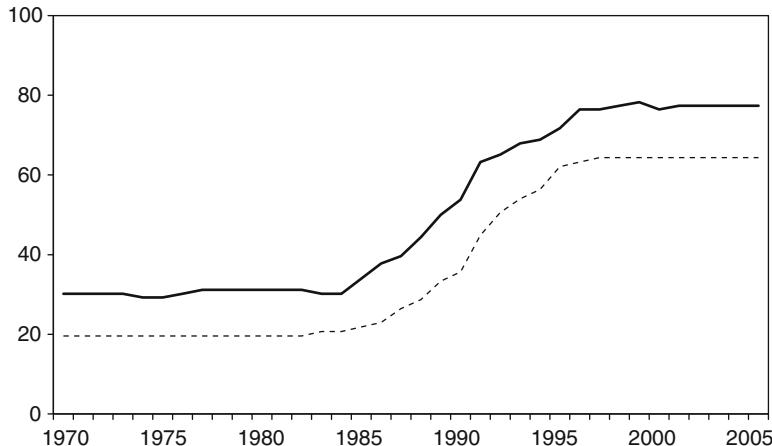


Figure 4 Evolution of trade and financial liberalization (percentage of open countries). Solid line = Trade Liberalization; Dashed line = Financial Liberalization. Notes: Trade liberalization measure indicates the fraction of countries with a fully liberalized trade regime. The dates of trade liberalization are determined on the basis of the Sachs and Warner (1995) and Wacziarg and Welch (2003). Financial liberalization measure indicates the fraction of countries with a liberalized financial system. The dates of official liberalization are determined on the basis of stock market liberalization and removal of restrictions on foreign investment based on the Bekaert, Harvey, and Lundblad (2005).

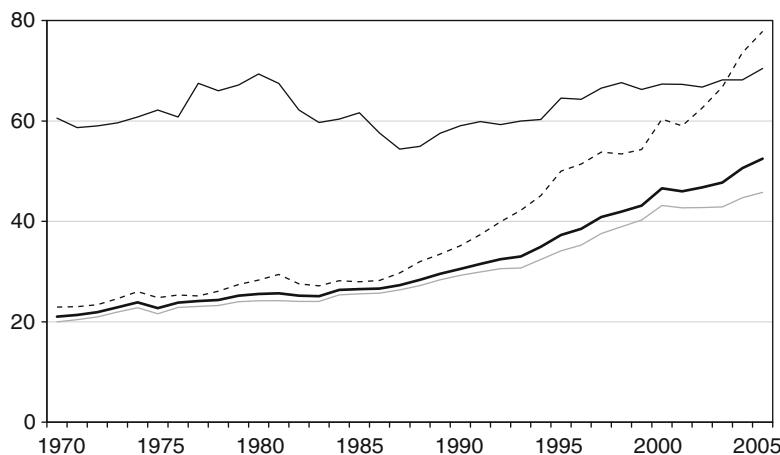


Figure 5 Trade openness (exports + imports/GDP, in %). Thick line = World; Gray line: Advanced; Dashed line = Emerging Markets; Thin Line= Other Developing.

3.1 Economic growth

3.1.1 Theory

Based on the standard one-sector neoclassical growth model, the traditional (direct) theoretical channel through which financial globalization affects economic growth is the

augmentation of capital. In other words, the standard theory predicts that financial globalization should lead to flows of capital from capital-rich economies to capital-poor economies since, in the latter, the returns to capital should be higher. In theory, these financial flows should complement limited domestic saving in capital-poor economies and, by reducing the cost of capital, allow for increased investment.¹¹ Certain types of financial flows could also generate technology spillovers and serve as a conduit for imbibing managerial and other forms of organizational expertise from more advanced economies.

Newer analyses emphasize the importance of indirect channels arguing that it is not just the direct financial flows, but the collateral benefits of these flows that drive the growth benefits of financial globalization (see Kose et al., 2006). These indirect channels include development of the domestic financial sector, improvements in institutions (defined broadly to include governance, the rule of law, etc.), and better macroeconomic policies.

These indirect theoretical channels are the subject of recent work. For example, Levine (2005) and Mishkin (2006, 2008) discuss the impact of financial integration on financial sector development. Stulz (2005) focuses on institutional quality and concludes that globalization weakens certain agency problems by reducing the cost of outside finance, thereby creating incentives for firms that use more external finance to improve their governance. Gourinchas and Jeanne (2005) show that financial integration can impose discipline on macroeconomic policies by improving the benefits of good policies and catalyzing political support for reforms while Bartolini and Drazen (1997) argue that, in exposing itself to such costs through increased financial openness, a country may signal its commitment to better macroeconomic policies.

We could continue at considerable length about how financial globalization matters in theory, and will indeed keep introducing further ideas throughout this chapter. However, what makes the debate on financial globalization fascinating is that several prominent economists question whether, in practice, the effects are positive at all. Most of these economists base their arguments on the theory of the second best and the potential presence of other distortions stemming from the trade policy regime, macroeconomic policies, labor markets, and information asymmetries. For example, if certain industries are protected by trade barriers, international capital could flow into these sectors to exploit the benefits of protection in domestic markets and result in welfare losses and sub-optimal growth (Brecher & Diaz-Alejandro, 1977). Information asymmetries stemming from a lack of transparency in financial institutions could lead to inefficient allocation of financial flows, generate maturity mismatches, and result in costly crises (Stiglitz, 2004).

The concern that financial globalization can sometimes spin off negative side effects in highly-distorted developing economies is a legitimate one. Indeed, as we shall see, in light of the ambiguity of theoretical findings, the critical question regarding policy in this entire literature is whether empirical evidence can guide us on why financial globalization seems to have clearly positive effects in some cases, whereas it appears to be counterproductive in others.

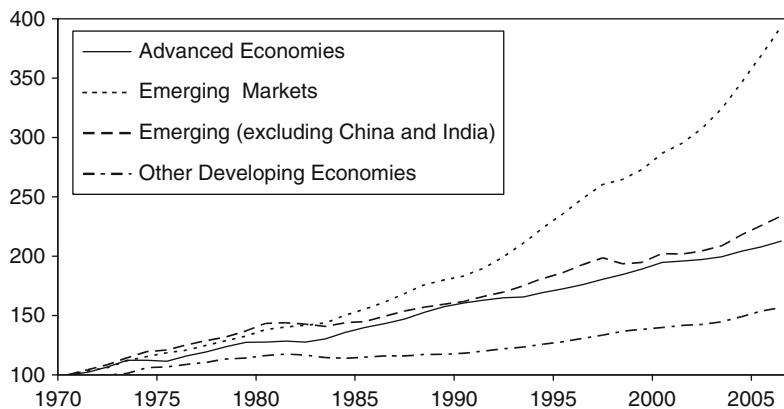


Figure 6 Evolution of GDP (per capita, PPP weighted). Notes: This plot shows cumulative changes in indexes of per capita GDP for each group of countries, computed using growth rates of real GDP for each country and weighting these by a PPP adjustment factor. The indices are set to 100 in the base period. The time period of analysis is 1970-2006.

3.1.2 Empirical evidence

3.1.2.1 Evidence on direct channels Apparently, there seems to be a positive association between embracing financial globalization and economic growth. For example, emerging market economies have, as a group, experienced far higher cumulative growth since 1970 than other developing countries or even industrial countries (Figure 6). Excluding China and India from the list of emerging markets makes the performance of this group look less spectacular, although it is still better than that of the group of other developing countries.¹²

To further illustrate the relationship between economic growth and financial openness, Figure 7 presents a scatter plot of the average growth rate of real per capita GDP against the average level of *de facto* financial openness over the past two decades. There is no systematic relationship between these variables. There is a weak positive association between average GDP growth and the *change* in the financial openness measure, consistent with the notion that economies that integrated into global financial markets grew faster. But once other growth determinants are controlled, even this relationship vanishes.

The interpretation drawn from these figures is consistent with the inconclusive findings from the large part of the literature on the benefits of financial globalization based on cross-country growth regressions. While some of these studies conclude that there are growth benefits associated with international financial integration, a majority of them tend to find no effect or a mixed effect for developing countries.¹³ This once again confirms that if financial integration has a positive effect on growth, it is apparently not robust, especially once the usual determinants of growth are controlled.

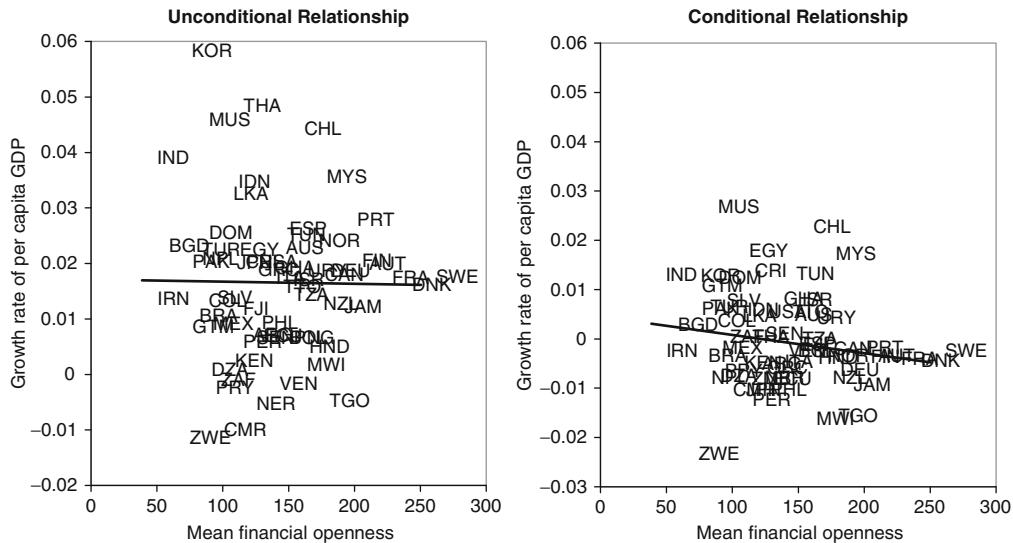


Figure 7 Level of financial openness and GDP growth. Notes: Growth refers to the average real per capita GDP growth. The time period of analysis is 1985–2004. Financial openness is defined as the ratio of gross stocks of foreign assets and liabilities to GDP and is based on a dataset constructed by Lane and Milesi-Ferretti (2006). The second panel uses residuals from a cross-section regression of growth on initial income, population growth, human capital and the investment rate.

Why do different studies focusing on direct channels reach such diverse conclusions about the importance of financial integration in affecting long-run economic performance? Empirical studies using finer *de jure* measures of capital account openness appear to achieve more positive results on the impact of financial integration on economic growth.¹⁴ In studies that use both *de jure* and *de facto* measures, specifications where capital account openness is measured using *de facto* measures tend to lend more support for the potential growth-enhancing effects of financial integration than those employing *de jure* measures. There are other reasons why the results differ markedly across studies—the sample period, country coverage, and choice of empirical methodology all make a big difference.

Moreover, depending on the type of financial flows considered, existing studies report vastly different results for the growth benefits of financial integration. Flows that have equity-like features—that is, FDI and portfolio equity flows—are not only presumed to be more stable and less prone to reversals, but are also believed to bring with them some additional theoretical benefits of financial globalization, such as transfers of managerial and technological expertise. In contrast, the procyclic and highly volatile nature of debt flows, especially short-term bank loans, can magnify the adverse impact of negative shocks on economic growth.

Although the aggregate growth benefits of FDI flows are hard to document, a reassessment of micro channels for technological spillovers from FDI inflows has begun to turn up more positive evidence of such spillovers. For example, direct evidence on the role of horizontal spillovers—productivity spillovers from foreign firms to domestic firms in the same sector—in transmitting the productivity benefits of FDI remains inconclusive. However, foreign firms have incentives to transfer knowledge to their local suppliers and customers, implying that productivity spillovers from FDI may occur through “vertical” linkages (see Javorcik, 2004).

The rising importance of portfolio equity flows to emerging markets has spurred a rapidly expanding literature that examines the growth effects of equity market liberalizations. Equity market liberalizations appear to improve economic performance, with an across the board increase in the growth rates of all major macro aggregates (Figure 8). Most papers in this literature report significant positive effects of equity market liberalizations on growth.¹⁵ However, whether these estimated growth effects (in macroeconomic data) could be picking up the effects of other factors—especially, other reforms that tend to accompany these liberalizations—remains an open question.

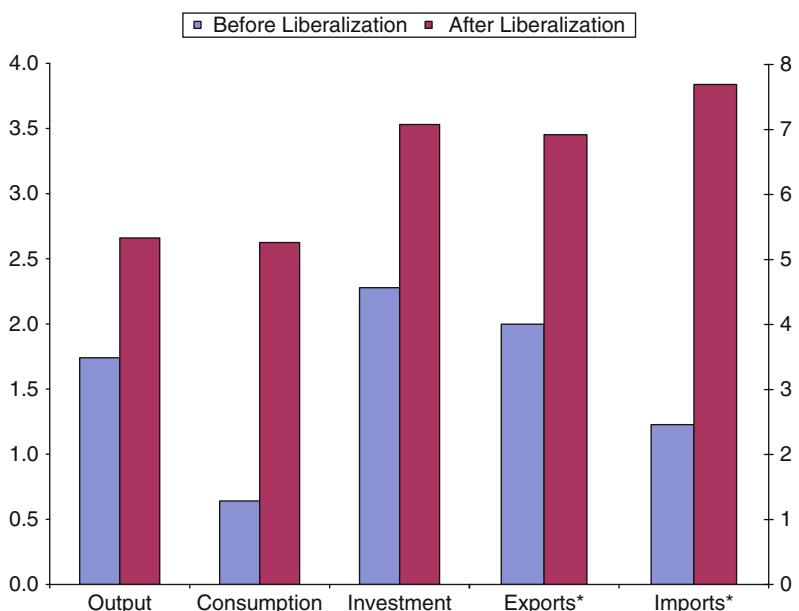


Figure 8 Growth before and after equity market liberalization (medians). *Notes:* The time period of analysis is 1980 to 2000. Countries with less than five observations, either before or after liberalization have been dropped from the sample. *Indicate categories plotted against right axis. The dates of official liberalization are determined on the basis of stock market liberalization and removal of restrictions on foreign investment based on the Bekaert, Harvey, and Lundblad (2005).

On the other hand, the body of microeconomic evidence (using industry- and firm-level data) supporting the macro evidence of the benefits of equity liberalizations is growing. Some of these papers also document the empirical relevance of various theoretical channels that link equity market liberalization to economic growth—including through increases in investment and total factor productivity growth (see Chari & Henry 2004, 2008; Mitton, 2006).¹⁶

The empirical literature is fairly decisive about debt flows worsening the benefit-risk tradeoff related to inflows. In particular, there is a systematic empirical link between exposure to short-term debt and the likelihood (and severity) of financial crises. One reason could be that countries with unfavorable conditions are forced to rely more on short-term external debt denominated in foreign currencies as their main source of foreign capital (Eichengreen, Hausmann, & Panizza, 2006).¹⁷

The summary so far suggests that there is no robust empirical evidence indicating that financial integration results in growth benefits through direct channels emphasized by the standard theory. Another empirical challenge facing the standard theory is about the direction (and actual volumes) of flows from capital-rich to capital-poor economies. In theory, capital should flow from relatively capital-rich countries (typically, the advanced economies) to relatively capital-poor economies (typically, the emerging markets and other developing economies) in order to equate marginal products of capital across countries. In fact, in recent years, capital has been flowing “uphill”—from the developing economies to the advanced economies. We will discuss the implications of this observation in detail later in the chapter. We now turn to the summary of the empirical literature focusing on indirect channels and collateral benefits stemming from these channels.

3.1.2.2 Indirect channels The indirect channels studied in empirical studies include financial sector development, institutional quality, and macroeconomic policies. Figure 9A–C presents some simple unconditional correlations, indicating that there could be some link between these channels and the extent of financial integration. During the recent period of financial globalization (1985–2004), financial openness is positively correlated with measures of financial development and institutional quality, and negatively correlated with log inflation. Its correlation with the government budget deficit is, however, essentially zero.

Recent empirical research provides some preliminary evidence that financial openness can promote development of the domestic financial sector, catalyze forces that result in better public and corporate governance, and impose discipline on macroeconomic policies. For example, work based on a variety of techniques supports the notion that increased foreign bank presence raises competition and leads to a decline in both bank overhead costs and profits.¹⁸ As for equity markets, the overwhelming theoretical presumption is that foreign entry increases efficiency yet the evidence is more mixed. While event studies suggest that stock markets become larger and more liquid after equity market liberalizations in a number of

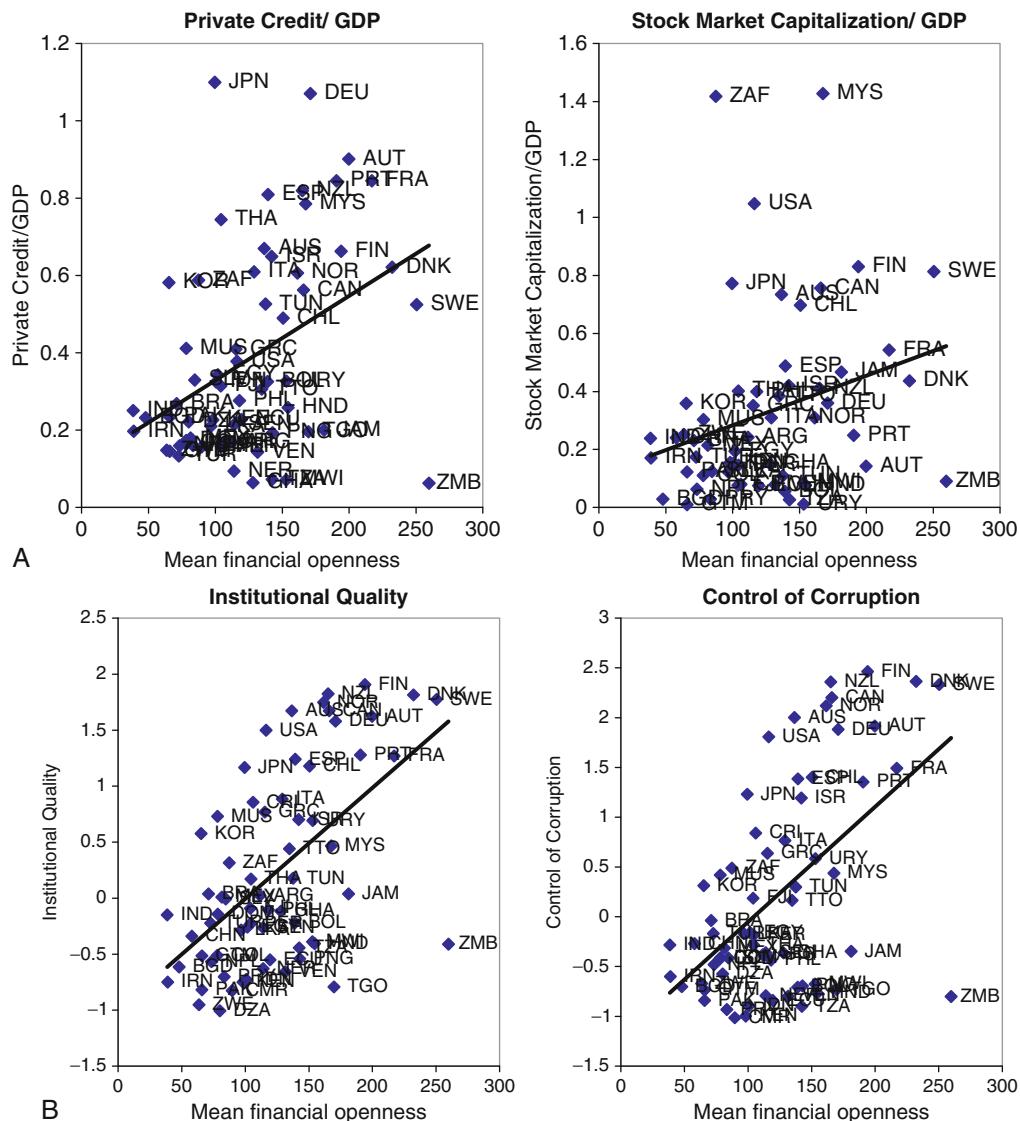


Figure 9 (A) Financial openness and financial development. (B) Financial openness and institutional quality.

(continued)

countries (Levine & Zervos, 1998), cross-country regressions indicate that financial openness contributes to equity market development only once at a moderate level of legal and institutional development has been attained (Chinn & Ito, 2006).¹⁹

The empirical evidence on financial globalization and institutional quality, while still sparse, does seem to indicate that financial globalization has helped some countries in

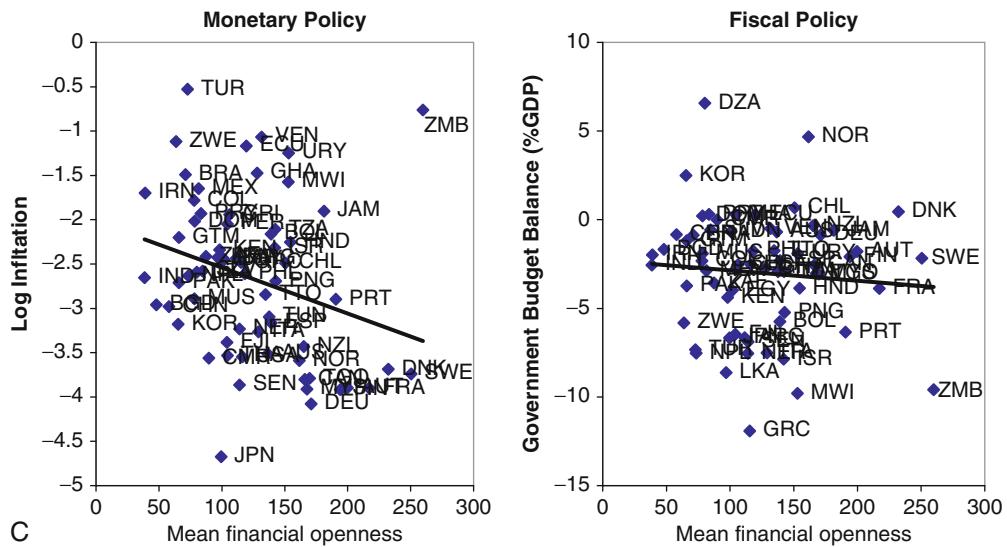


Figure 9—cont'd (C) Financial openness and macroeconomic policies. Notes: The financial integration data are based on a dataset constructed by Lane and Milesi-Ferretti (2006). The time period of analysis is 1985-2004. Financial Development data are taken from Beck and Al-Hussainy (2006). Private Credit refers to credit given to the private sector by deposit money banks and Stock Market Capitalization is defined as the value of listed shares. Institutional quality data are from Daniel Kaufmann, Aart Kraay and Massimo Mastruzzi (2005) and cover the period 1996-2004. Institutional Quality is the average of the following indicators: Voice and Accountability, Political Stability, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. Monetary and Fiscal data are from WDI, WEO and IFS. Inflation is defined as the annual change in CPI. Government Budget Balance is the difference between Government Revenues and Government Expenditures.

improving certain institutional features. For example, some countries have adjusted their corporate governance structures in response to demands from international investors (Cornelius & Kogut, 2003).²⁰ In addition, reforms to institutions take place mostly after financial integration and there appears to be a substantial improvement in the measures of law and order between partial and full liberalization (Kaminsky & Schmukler, 2003).

Financial integration also tends to have a disciplining effect on macroeconomic policies. Countries with higher levels of financial openness are more likely to generate better monetary policy outcomes in terms of lower inflation (Gupta & Yuan, 2008; Spiegel, 2008; Tytell & Wei, 2004). By contrast, there is little systematic evidence of a relationship between financial openness and better fiscal policies. Indeed, one must strike a cautionary note that, in practice, easy access to foreign finance may simply allow profligate governments to run larger budget deficits for a longer period without being limited by domestic financing constraints—so long as international investors are willing to finance these deficits.

How do the indirect benefits of financial integration translate into economic growth? They could enhance the growth outcomes through their impact on total factor productivity. If financial integration is to have a lasting effect on growth, it must be by moving economies closer to their production possibility frontiers by eliminating various distortions and creating efficiency gains, for example, in financial intermediation, technological adoption, etc.²¹ There appears to be a positive association between the extent of financial integration and productivity growth (Figure 10). Bonfiglioli (2008) and Kose, Prasad, and Terrones (2008a) have assembled some preliminary evidence suggesting that financial integration raises the TFP growth.

3.1.2.3 A Summary Although it is difficult to argue that the empirical evidence summarized here is decisive, the notion that financial globalization influences growth mainly through indirect channels has powerful implications for empirical analysis of its growth effects²² even after the effects take hold; however, they might be difficult to document. Standard growth

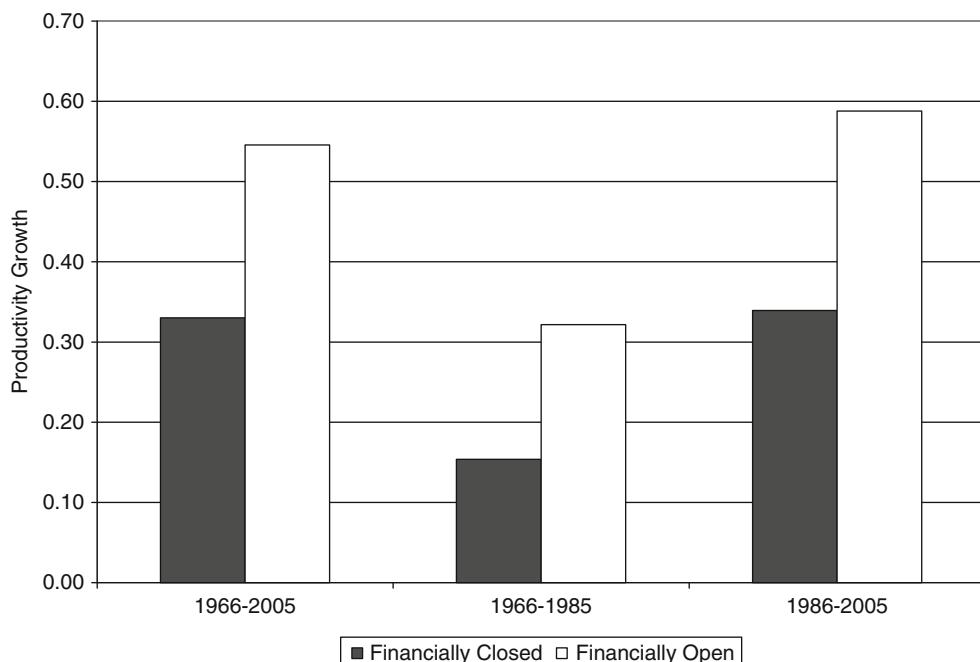


Figure 10 Financial openness and productivity growth (median values; in %). Notes: The measure of total factor productivity is based on a standard growth accounting exercise. Aggregate output is described by the Cobb-Douglas production function with labor, physical and human capital as factors of production. The degree of financial openness is based on the stock of liabilities relative to GDP. Financially open/closed sub-samples are defined relative to the median of the full sample.

regressions, used to evaluate the direct benefits of integration, include measures of institutional quality, financial sector development, quality of macroeconomic policies, etc. Yet, these may be the very channels through which financial integration generates growth benefits, making it difficult to disentangle the effects of financial integration. Moreover, the perspective of attaining collateral benefits through indirect channels could be helpful in designing policies to move forward on capital account liberalization. We will return to this topic later in this chapter when we discuss policy issues.

3.2 Macroeconomic volatility and patterns of international risk-sharing

3.2.1 Theory

In theory, the effects of financial integration on output volatility are ambiguous. Financial integration allows capital-poor countries to diversify away from their narrow production bases that are often agricultural or natural resource-dependent, thereby reducing the macroeconomic volatility. At a more advanced stage of development, however, trade and financial integration could together allow for enhanced specialization, as we have already noted. This could make middle-income developing countries more vulnerable to industry-specific shocks and thereby lead to higher output volatility (see Kose, Prasad, & Terrones, 2004).

If financial integration relies heavily on external debt, it could expose these countries to world interest rate shocks and, thus, to higher output volatility. For example, Rodrik and Velasco (2000) found that the ratio of short-term debt to reserves is a robust predictor of financial crises among emerging market economies. They report that countries with a larger short-term debt stock than reserves are three times more likely to experience a sudden and massive reversal in financial flows. Their results also indicate that the severity of crises becomes more acute as the exposure to short-term debt increases.

Theory, however, strongly predicts about the relationship between financial integration and consumption volatility. Since consumers and, by extension, economies are risk-averse, consumption theory tells us that they should desire to use financial markets to insure against income risk, thereby smoothing the effects of temporary idiosyncratic fluctuations in income growth on consumption growth. While the benefits of international risk-sharing could be quite large in theoretical models, the magnitudes of these benefits depend on various model-specific features.²³

Another prediction of theory, related to the consumption smoothing issue, concerns the cross-country comovement of major macroeconomic aggregates. In theory, the effect of increased financial integration on cross-country correlations of output growth is uncertain, as it depends on the nature of shocks and specialization patterns. However, financial integration should in theory help countries diversify away from country-specific risk and should, therefore, result in stronger comovement of consumption growth across countries.

3.2.2 Empirical evidence

There has been a well-documented decline trend in macroeconomic volatility in most of the major industrial economies since the mid-1980s, although the reasons for this decline are still a matter of debate. Output volatility seems to have been on a declining trend in emerging market and developing economies as well. However, the existing evidence based on papers using a variety of regression models, different country samples, and time periods leads to the conclusion that there is no systematic empirical relationship between financial openness and output volatility. This is not surprising since, as noted earlier, there is no consistent theoretical prediction across different models about how financial integration affects output volatility.²⁴

Kose, Prasad, and Terrones (2003b) note that, during the 1990s, average declines in output growth volatility were smaller for emerging markets than for either industrial or low-income developing economies. More importantly, they find that the ratio of consumption growth volatility to income growth volatility increased during the recent period of globalization for emerging market economies (and remained flat for the other two groups). What is surprising is not just that the volatility of consumption rose (perhaps because of crises experienced by some of these economies) but that it increased *more* than income volatility.²⁵ This result contradicts the presumed theoretical benefit of financial integration—that it allows countries to share income risk and smooth consumption.

In a related paper, Kose, Prasad, and Terrones (2008b) examined the risk-sharing implications of financial integration by focusing on the cross-country correlations of output and consumption. They report that, notwithstanding the prediction of conventional theoretical models, the financial globalization should foster increased risk-sharing across all countries; there is no evidence that this is true for developing countries. Even for the group of emerging market economies—which have become far more integrated into global markets than other developing countries—financial globalization has not improved the degree of risk-sharing (Figure 11).²⁶

Both of these papers reach sobering conclusions about the stability and risk-sharing implications of financial globalization, but as we discuss in the later sections their findings also depend on some country-specific conditions and the level and composition of financial flows.

3.2.3 Crises as special cases of volatility

The proliferation of financial crises is often viewed as one of the defining aspects of the intensification of financial globalization over the last two decades. Furthermore, the fact that recent crises have affected mainly emerging market economies has led to these phenomena being regarded as hallmarks of the unequal distribution of globalization's benefits and risks (Desai, 2003). This raises a challenging set of questions about whether the nature of crises has changed over time, what factors increase the vulnerability to crises, and whether such crises are an inevitable concomitant of financial globalization. These crises can be regarded as particularly dramatic episodes of volatility. While the research on the impact of financial

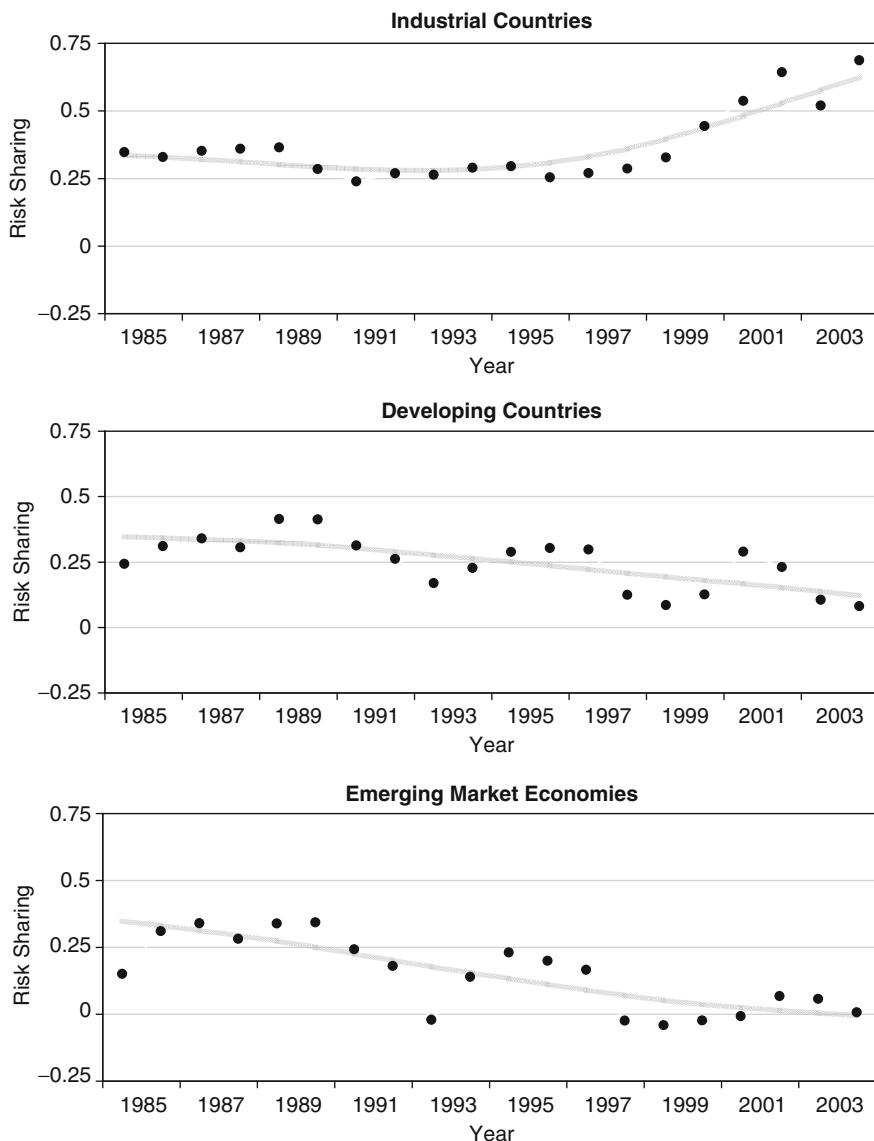


Figure 11 Evolution of risk sharing. Notes: The degree of risk sharing is based on a regression of a country's idiosyncratic consumption growth on its idiosyncratic output growth (Obstfeld 1995). These regressions are estimated for each country over rolling nine-year periods. World aggregates are calculated using industrial-country data. Medians for each group of countries are reported. The continuous line is the HP filter trend. Emerging Market Economies are part of the group of Developing Countries.

integration on volatility and risk sharing has resulted in somewhat negative results, recent research analyzing the effects of integration on crises has painted a different picture.

Some papers that have analyzed the effects of capital controls on susceptibility to financial crises have found that countries *with* capital controls are in fact more subject to crises. But this could simply be because of a “selection effect”—often it is countries with poor macroeconomic fundamentals that put controls in place to try and insulate themselves from crises. Glick, Guo, and Hutchison (2006) address this issue—they find that capital account openness reduces the probability of currency crises, even after controlling for selection bias in terms of how macroeconomic policies influence the existence of capital controls.²⁷ The relationship between capital controls and crises could also reflect the fact that some of the countries are actually more integrated in terms of *de facto* measures of integration (capital flight) and that capital controls, therefore, do not insulate them from crises. We provide a detailed analysis of the macroeconomic implications of capital controls later in the chapter.

Edwards (2005) examines this issue using a new measure of *de jure* financial openness that attempts to capture the intensity of capital account restrictions. He looks at two manifestations of external crises—sudden stops of inflows and current account reversals—and finds no evidence that countries with higher capital mobility tend to have a higher incidence of crises. In subsequent work, Edwards (2008) concludes that there is no evidence that the output costs of currency crises are smaller in countries that restrict capital mobility.

While currency crises have been emphasized in the literature on the risks of capital account liberalization, it is worth noting that banking crises account for about one-third of financial crises over the last three decades and that their frequency increased in the 1980s and 1990s. Banking crises tend to be more disruptive and generally have larger adverse effects on output growth than currency crises. Glick and Hutchison (2001) find little evidence that capital account liberalization by itself affects vulnerability to banking crises; moreover, the adverse effects of banking crises seem to be weaker for countries with open capital accounts.²⁸

In sum, there is little formal empirical evidence to support the oft-cited claims that financial globalization in and of itself is responsible for the spate of financial crises that the world has seen over the last three decades.²⁹ Of course, as we will discuss in more detail below, the interaction between capital account liberalization and other policy choices (e.g., fixed exchange rate regimes that are not well supported by other macroeconomic policies) could, under certain circumstances, spell trouble for a developing economy.

4. ECONOMIC POLICIES AND GROWTH OUTCOMES

Researchers have explored a number of avenues to reconcile the strong theoretical prediction that financial integration should boost long-run growth and reduce the risks of consumption instability in developing economies with weak empirical evidence. Some authors have argued that countries that do not have the right initial conditions

associated with certain structural and macroeconomic factors can experience growth surges due to financial integration but they inevitably experience crises, which pulls down their long-run growth. Other authors have argued that financially integrated developing countries that lack these factors are not able to derive the full benefits of financial integration even if they can escape crises.

Kose et al. (2006) pull these two lines of argument together to describe the conditioning variables that influence the relationship between financial integration and growth as a set of “threshold conditions.” These threshold conditions help determine the nature of policy measures that could improve the growth and stability benefits of financial globalization. They include an economy’s structural features—the extent of financial sector development, institutional quality, and trade integration—and also the macroeconomic policy framework.

Before getting into the details of theoretical arguments and empirical findings in the literature, we first present some preliminary observations on whether there are obvious threshold effects in the data based on recent work by Kose, Prasad, and Taylor (2008).³⁰ In particular, our interest is in whether, within the groups of emerging markets and other developing countries, the levels of certain conditioning variables are positively associated with economic performance. Table 1 compares unconditional and conditional growth rates over the period 1975–2004 for countries that are above or below the within-group sample medians for various variables that have been put forward as threshold variables in the related theoretical and empirical literature.

The main findings from this table can be summarized as follows: First, unconditional growth rates in emerging market countries are greater for those countries with higher (within-group above-median) levels of illustrative threshold indicators for financial development, trade openness, institutional quality, and macro policies, although this difference is not always statistically significant. These effects are less pronounced in other developing countries. However, for institutional quality the pattern is reversed. Second, for conditional growth rates, the patterns are less pronounced, although across many indicators the positive association of growth with higher thresholds remains. Third, the difference between the growth rates of emerging markets and other developing countries is more pronounced for the sub-samples with higher levels of thresholds.

Our preliminary exploration lends support to the notion of various factors related to a country’s structural characteristics and macroeconomic policy choices, playing a role in the relationship between financial openness and growth. However, clearly it abstracts from issues relating to endogeneity or the nature of any threshold relationship. To address these and other related issues, we provide a review of the theoretical and empirical literature on the importance of structural features and macroeconomic policies in the next section. For each of these factors, we review the underlying theoretical arguments and survey the relevant empirical evidence.

Table 1 Output growth and threshold factors (in %)

	Unconditional growth		Conditional growth	
	Emerging markets	Other developing countries	Emerging markets	Other developing countries
Overall	2.28 (1.94)	0.82 (0.65)	0.44 (0.53)	-0.16 (-0.04)
<i>Financial sector development</i>				
By private credit to GDP	High	3.16 (3.11)	0.66 (0.45)	0.73 (0.67)
	Low	1.49 (1.41)	0.98 (0.88)	0.18 (0.50)
	Difference in means	1.67*	0.33	0.56
<i>Institutional quality</i>				
By average WBGI institutional quality index	High	2.42 (1.88)	1.22 (0.85)	0.39 (0.42)
	Low	2.17 (1.94)	0.42 (0.41)	0.48 (0.63)
	Difference in means	0.25	0.80*	-0.09
				1.06**
<i>Trade openness</i>				
By <i>de facto</i> trade openness	High	2.74 (2.66)	1.26 (0.71)	0.51 (0.52)
	Low	1.87 (1.20)	0.63 (0.87)	0.38 (0.66)
	Difference in means	0.87	0.63	0.13
				0.66

Macroeconomic policies

By S.D. of CPI inflation	Low	3.38 (3.37)	1.51 (1.54)	1.07 (0.97)	0.40 (0.38)
	High	1.08 (1.15)	0.22 (0.35)	-0.26 (-0.24)	-0.84 (-0.81)
	Difference in means	2.30***	1.29***	1.33***	1.24***
By government expenditure to revenue ratio	Low	2.68 (2.88)	1.28 (1.16)	0.43 (0.53)	0.28 (0.39)
	High	1.54 (1.33)	0.53 (0.49)	0.16 (0.01)	-0.30 (-0.12)
	Difference in means	1.13	0.75	0.27	0.59

Notes: Output growth is the mean (median) annual growth of GDP. The time period of analysis is 1975–2004. The symbols ***, **, and * indicate statistical significance at 10%, 5%, and 1%, respectively of a *t*-test of mean equality across subsamples. High/low subsamples are defined relative to medians within groupings. Conditional growth is residual from cross-section regression of growth on log initial GDP per capita, average investment to GDP, average years schooling, and average population growth rate.

5. STRUCTURAL CHARACTERISTICS AND ECONOMIC POLICIES

5.1 Financial sector development

5.1.1 Theory

There is a strong theoretical presumption that the financial sector development not only enhances the growth benefits associated with financial globalization but also reduces the vulnerability to crises. It is intuitive that well-developed domestic financial markets are instrumental in efficiently allocating foreign financial flows to competing investment projects (Wurgler, 2000). A number of more formal models have been developed to analyze the effects of capital account liberalization in economies with limited financial development. Domestic and international collateral constraints could play a particularly important role in financially underdeveloped economies where access to arm's length financing is limited. Caballero and Krishnamurthy (2001), Aghion, Bacchetta, and Banerjee (2004), Mendoza, Quadrini, and Rios-Rull (2007), and Aoki, Benigno, and Kiyotaki (2007) show how, in different theoretical settings, the interaction of these constraints can lead to unpredictable and, occasionally, adverse effects of capital account liberalization.

Financial development also has a direct impact on macroeconomic stability in financially open economies. Sudden changes in the direction of capital flows tend to induce or exacerbate boom-bust cycles in developing countries that lack deep and well-functioning financial sectors (Aghion & Banerjee, 2005; Caballero & Krishnamurthy, 2001). Moreover, inadequate or mismanaged domestic financial sector liberalizations have been a major contributor to crises associated with financial integration (Mishkin, 2006). The lack of well-developed financial markets also appears to be a key reason in explaining the positive association between financial integration and the relative volatility of consumption growth documented by Kose et al. (2003b).³¹

5.1.2 Empirical evidence

There has been a large empirical literature analyzing the role of financial development in determining the impact of financial integration on economic growth (see Table 2). The main indicators of financial development used are private credit and stock market capitalization to GDP, although these might be better described as indicators of financial depth. In terms of financial openness, the main indicators used are either inward foreign direct investment (FDI) or measures of capital controls while equity flows and liberalizations of equity markets are also employed in some cases.

Using a large sample of developing countries over the period 1970–1995, Hermes and Lensink (2003) find that, in order to enjoy the growth benefits of FDI, a threshold level of financial sector development is a prerequisite. While more than half of the countries in their sample (mostly in Latin America and Asia) appear to meet the necessary threshold, almost all of the countries in sub-Saharan Africa, with their relatively

Table 2 Summary of empirical studies on structural factors and growth (financial development)

Study	No. of econ./ period	Econometric methodology	Dependent variable	Financial openness variable	Interaction/ threshold variables	Interaction/ threshold approach	Main findings on interaction effect
<i>1. FDI as financial openness measure</i>							
Hermes and Lensink (2003)	67 developing (1970-1995)	Cross-section OLS	Growth of real per capita GDP	Gross FDI inflows to GDP	Private bank credit to GDP	Linear	<i>Positive significant coefficient on interaction of FDI with FD variables</i>
Alfaro et al. (2004)	71 (1975-1995) for banking variables 50 (1980-1995) for stock market variables	Cross-section OLS plus IV	Growth of real per capita GDP	Net FDI inflows to GDP	Banking: liquid liabilities to GDP; commercial bank assets to commercial bank plus central bank assets; private credit to GDP; private bank credit to GDP. Stock market: value traded to GDP; stock market capitalization to GDP	Linear	<i>Positive significant coefficient on interaction of FDI with FD variables.</i> <i>Robust to additional controls and IV estimation.</i>

Continued

Table 2 Summary of empirical studies on structural factors and growth (financial development)—Cont'd

Study	No. of econ./ period	Econometric methodology	Dependent variable	Financial openness variable	Interaction/ threshold variables	Interaction/ threshold approach	Main findings on interaction effect
Durham (2004)	Up to 62 (1984–1998)	Cross-section OLS	Growth of real per capita GDP	FDI flows from 1979–1983 (OECD and IFS measures) and net portfolio equity inflows from US from 1979–1983	Stock market capitalization to GDP	Linear	<i>Mixed results.</i> Interaction coefficient positive significant for FDI using OECD data and with net portfolio equity inflows but insignificant with FDI using IFS data
Carkovic and Levine (2005)	Up to 68 (1960–1995)	Cross-section OLS and 5-yearly panel dynamic system GMM	Growth of real per capita GDP	Gross FDI inflows to GDP	Private credit by financial intermediaries to GDP	Linear	<i>Mixed results.</i> Interaction coefficient positive significant in cross-section but not significant in panel system estimation

2. Other FO measures

Kraay (1998)	n.a. ^a (1985-1997)	Cross-section OLS and IV. Event study	Growth of real per capita GDP	AREAER share; Quinn index; gross capital flows to GDP	M2 to GDP; domestic private credit to GDP; 1-average number of banking crises per year; freedom to undertake “nontraditional” banking activities	Linear (for regressions) sample split (above/below median) for event study	<i>Little evidence of interaction effects. Linear interaction coefficients either insignificant or mixed sign. Similar results with event studies</i>
Arteta et al. (2003)	Up to 62 (1973-1992)	Cross-section and subperiod panel pooled OLS	Growth of real per capita GDP (PPP)	Initial value of Quinn capital account liberalization index	Initial liquid liabilities to GDP	Linear	<i>Coefficient on FD interaction term insignificant</i>
Bekaert et al. (2005)	Up to 95 (1980-1997)	5-yearly panel (overlapping periods) GMM	5-year average growth rate of real per capita GDP	<i>De jure</i> international equity market liberalization	Private credit to GDP Stock market turnover	Sample split: Above/below median	<i>Countries with higher FD have significantly higher growth gain postliberalization</i>
Hammel (2006)	13 (1982-1995)	FE panel 3-year pre- and postliberalization	3-year growth in real sectoral	Bekaert et al. <i>de jure</i> international	Stock market capitalization to GDP	Sample split: various (base result	<i>Positive sig. coefficient on interaction of sectoral external</i>

Continued

Table 2 Summary of empirical studies on structural factors and growth (financial development)—Cont'd

Study	No. of econ./ period	Econometric methodology	Dependent variable	Financial openness variable	Interaction/ threshold variables	Interaction/ threshold approach	Main findings on interaction effect
			value added	equity market liberalization		for >10% GDP)	<i>finance dependence and liberalization for countries with stock cap. over 10% GDP (insig. if median split used)</i>
Prasad et al. (2007)	83 (<i>this is in an aggregate analysis—is it same in sectoral?</i>) (1980–1990)	Country and industry FE	Growth in real sectoral value added	Stock liabilities and gross and net flow liabilities to GDP: FDI; FDI and portfolio. <i>De jure</i> : Chinn-Ito; Edwards	Private sector credit to GDP	Sample split: below median	<i>Generally negative significant coefficient on interaction of external finance dependence of industry and FO measure for countries with below median FD</i>
Coricelli et al. (2008)	31 European economies, 1996–2004	Annual panel dynamic GMM	Growth of real per capita GDP	Stock of external liabilities and assets plus	Private sector credit to GDP and stock market capitalization	Sample split: various (10% of	<i>Evidence supportive of nonlinear interaction with</i>

liabilities to GDP: total, FDI; portfolio and other flows	plus private sector credit to GDP	GDP gradations) ^b	<i>coefficient on financial integration positive for financial development measures in the range 60–150% of GDP</i>
---	---	---------------------------------	---

Notes:

Samples and methodology: Details in table relate to sections of study when interaction effects are examined (and hence, may differ from other parts of papers).

Data for Hermes and Lensink (2003) from tables 3 and 4, Alfaro et al. from tables 4, 6 and 7, Durham (2004) from tables 1, 2, and 3, Carkovic and Levine (2005) from table 8.4 and 8.6, Kraay (1998) from tables 10–12, Arteta et al. (2001) from table 4, Bekaert et al. (2005) from table 8, Hammel (2006) from table 7 and Prasad et al. (2006) from tables 7 and 8.

Financial openness de jure measures: AREAER share: proportion of years in which countries had liberalized capital accounts based on the binary variable from AREAER; Quinn: *de jure* capital account liberalization measure based on Quinn (1997); Chinn-Ito: AREAER based measure of capital controls from Chinn and Ito, 2006; Edwards index from Edwards (2005). See Kose et al. (2006) for more details.

Unless indicated, country samples include industrial and developing economies.

GDP noted as PPP where explicitly indicated in paper.

Significant if at least 10% significance level.

Note: Unclear if Durham uses inflows from paper but from country summary statistics it looks like FDI inflows.

^aNumber of countries in interaction regressions not indicated by Kraay (1998). Financial openness measures are available for 117 countries for IMF AREAER share measure, 64 for Quinn liberalization measure, and 94 for gross capital flows measure.

weak financial systems, are below this level. Alfaro, Chanda, Kalemli-Ozcan, and Sayek (2004), Carkovic and Levine (2005), and Durham (2004) also find that the growth impact of FDI is stronger in economies with well-developed financial sectors. However, the implied financial development thresholds for a positive financial openness coefficient vary substantially within and across these four studies (Kose, Prasad, & Taylor, 2008). For example, the credit to GDP thresholds (from cross-section regressions) vary from 13% to 48%. This is likely to reflect different time and country samples, and also different credit measures employed.

Financial sector development also appears to improve the growth benefits of equity flows. For example, Bekaert, Harvey, and Lundblad (2005) and Hammel (2006) find that growth following equity market liberalizations (which allow foreign investors to invest in domestic stock markets) is higher for countries with higher private credit/stock market turnover and stock market capitalization, respectively.³²

Another major benefit of financial sector development is its positive impact on macroeconomic stability, which in turn has implications for the volume and composition of capital flows. In theory, by expanding the scope of diversification possibilities, developed financial markets moderate the effects of shocks and help reduce macroeconomic volatility.³³ Economic crises in emerging markets have repeatedly demonstrated the importance of deep and well-supervised domestic financial markets during the process of financial integration. Mishkin (2006) discusses how, after capital account liberalization, excessive risk taking by domestic banks played a major role in triggering the financial crises in Mexico in 1994 and many East Asian countries in 1997. Ishii et al. (2002) documented that countries with stronger financial systems generally avoided crises following capital account liberalization. However, countries with underdeveloped and poorly supervised financial markets suffered financial crises after liberalizing their capital accounts. Recent empirical work also finds that in countries with deeper domestic financial markets, financial integration is indeed associated with lower consumption growth volatility (Eozenou, 2006; IMF, 2007).

5.2 Institutional quality

5.2.1 Theory

Institutional quality has also received considerable attention as an important structural factor in the relation between financial openness and growth. The quality of corporate and public governance, the legal framework, the level of corruption, and the degree of government transparency can affect the allocation of resources in an economy. Since capital inflows make more resources available, the quality of institutions therefore matters much more for financially open economies. For instance, postmortems of the Asian financial crisis have pinned a significant portion of the blame on crony capitalism that was encouraged and facilitated by corruption and weak public governance

(Haber, 2002; Krueger, 2002). Indeed, an intermediate degree of financial openness with selective capital controls may be most conducive to crony capitalism, as it gives politically well-connected firms preferential access to foreign capital (Johnson & Mitton, 2003). We will discuss this issue later in the context of capital controls as a policy instrument.

Weak protection of property rights in poor countries means that foreign financing may not be directed to long-gestation, investment-intensive, and low-initial profitability projects (including infrastructure) where such financing could be particularly useful, given the domestic financing constraints (see Rajan & Zingales, 1998). Some authors have argued that, while factors such as weak macro policies are indeed precursors of crises, the deep determinants of bad macroeconomic and structural policies can in fact be traced back to weak institutions (Acemoglu, Johnson, Robinson, & Thaicharoen, 2003). These models imply that there may be important interactions among the threshold conditions themselves in determining the growth and volatility effects of financial integration.

5.2.2 Empirical evidence

Empirical evidence suggests that institutional quality appears to play an important role in determining not just the outcomes of financial integration but also the level of *de facto* integration. Furthermore, institutional quality also appears to have a strong influence on the composition of inflows into developing economies, which is another channel through which it affects macroeconomic outcomes.

A number of empirical studies find that better institutions appear to enhance the growth benefits of capital account liberalization. A range of indicators for both institutions and financial openness has been employed in the empirical literature, with the interaction terms having varying degrees of significance across studies (see Table 3). Kraay (1998) and Quinn and Toyoda (2006) argue that there is little evidence of interaction effects while Bekaert et al. (2005) and Chanda (2005) are more supportive. Klein (2005) finds that only intermediate levels of institutional quality are associated with a positive correlation between growth and capital account liberalization. This hints at the possibility of nonlinear threshold effects. Chanda (2005) finds that the cross-country relationship between capital controls and growth depends on the degree of ethnic heterogeneity, which he interprets as a proxy for rent-seeking and common pool problems. For countries with more heterogeneity (more competing groups), capital controls lead to greater inefficiencies and lower growth.³⁴

As an alternative to using indices of institutional quality, a country's level of income has been used as a proxy for an overall institutional development and interacted in a similar manner with financial openness measures. These studies report mixed results. Edwards (2001) and Edison, Klein, Ricci, and Sløk (2004) find evidence of a positive, significant linear interaction and an inverted U-shaped relationship, respectively. However, other papers examining these linkages, such as Arteta, Eichengreen, and Wyplosz

Table 3 Summary of empirical studies on structural factors and growth (institutions)

Study	No. of econ./ period	Econometric methodology	Dependent variable	Financial openness variable	Interaction/ threshold variables	Interaction/ threshold approach	Main findings on interaction effect
Kraay (1998)	n.a. ^a (1985-1997)	Cross-section OLS and IV. Event study	Growth of real per capita GDP	AREAER share; Quinn index; gross capital flows to GDP	Macro policy (weighted av. of gov. deficit and inflation); black market premium; ICRG corruption; ICRG bureaucracy quality	Linear (for regressions) sample split (above/below median) for event study	<i>Little evidence of interaction effects.</i> Linear interaction coefficients either insignificant or generally negative. Similarly no evidence from event studies
Arteta et al. (2003)	Up to 62 (1973-1992)	Cross-section and subperiod panel pooled OLS	Growth of real per capita GDP (PPP)	Initial value of Quinn index	ICRG law and order	Linear	<i>Positive sig. interaction coefficient in pooled sample (and for 1973-1981 and 1982-1987 but insig. for 1988-1992)</i>
Durham (2004)	Up to 62 (1984-1998)	Cross-section OLS	Growth of real per capita GDP	FDI flows (OECD and IFS) and net portfolio equity inflows from US from 1979-1983	Regulation; property rights; corruption	Linear	<i>Mixed results with positive interaction effects not robustly significant across FO measures</i>

Bekaert et al. (2005)	Up to 95	5-yearly panel (overlapping periods)	5-year average growth rate of real per capita GDP	<i>De jure</i> international equity market liberalization	Legal measures; institutional variables; investment condition variables ^b	Sample split: Above/below median	<i>English legal origin, high investment profile, anti-director rights and accounting standards associated with sig. higher growth gain postliberalization. Judicial measures, economic risk and creditor rights insig.</i>
Chanda (2005)	Up to 82	Cross-section OLS (1975–1995)	Growth of real per capita GDP	AREAER share; Freedom House measure	Ethnolinguistic homogeneity	Linear	<i>Interaction coefficient on capital controls and ethno-linguistic homogeneity is positive and sig.</i>
Klein (2005)	Up to 71	Cross-section OLS and IV, nonlinear LS (1976–1995)	Growth of real per capita GDP	AREAER share	Institutional quality (average of 5 ICRG measures) ^c	Cubic; quadratic spline, quintile dummies	<i>Intermediate levels of institutional quality associated with a positive relation between growth and capital account liberalization</i>
Quinn and Toyoda (2006)	Up to 85	5-yearly panel FE and system GMM (1955–2004)	Growth of real per capita GDP	Level of Quinn index	Ethnic fractionalization, black market premia, ICRG	Linear	<i>Conclude that capital account liberalization effects generally direct over</i>

Continued

Table 3 Summary of empirical studies on structural factors and growth (institutions)—Cont'd

Study	No. of econ./ period	Econometric methodology	Dependent variable	Financial openness variable	Interaction/ threshold variables	Interaction/ threshold approach	Main findings on interaction effect
				bureaucratic quality, ICRG corruption ^d			<i>this period</i> with interaction effects generally insignificant (or exhibiting contrary effects in different subperiods)

Notes:

Samples and methodology: Details in table relate to sections of study when interaction effects are examined (and hence, may differ from other parts of papers). Data for Kraay (1998) from tables 10–12, Arteta et al. (2003) from table 4, Durham (2004) from tables 1, 2, and 3, Bekaert et al. (2005) from table 8, Chanda (2005) from tables 3 and 4, Klein (2005) from table 3, and Quinn and Toyoda (2006) from table 15.

Financial openness de jure measures: See notes to Table 2.

Note: Information for Arteta et al. (2003) taken from NBER working paper no. 8414 version.

Unclear if Durham uses inflows from paper but from country summary statistics it looks like FDI inflows. Also Durham uses a business regulation index from Levine 2000 but in Levine paper it is an index of regulatory restrictions on commercial bank activities which is from 1 to 4. Property rights index is cited as from La Porta et al. (2000)—Corruption index from Knack and Keefer 1995—should be ICRG corruption index.

^aNumber of countries in interaction regressions not indicated by Kraay (1998). Financial openness measures are available for 117 countries for IMF AREAER share measure, 64 for Quinn liberalization measure, and 94 for gross capital flows measure.

^bBekaert et al. (2005) legal measures include legal origin, judicial efficiency, speed of process; institutions include ICRG summary index (sum of three International Country Risk Group (ICRG) indices on bureaucratic quality, corruption, and, law and order); investment conditions included ICRG economic risk index, ICRG investment profile index, anti-director rights, creditor rights and accounting standards.

^cKlein institutional quality measure is an average of ICRG measures for 1984–1995 for bureaucratic quality, corruption, expropriation risk, risk of repudiation of government contracts and rule of law.

^dQuinn and Toyoda also examine the interaction of their capital account measure with banking crises and financial crises and find no significant interaction effects.

(2003) and Quinn and Toyoda (2006), both of which use *de jure* measures, and Carkovic and Levine (2005), which uses FDI flows as a measure of financial openness, fail to find evidence of income levels influencing the growth effect of financial openness.

Moreover, better institutional quality increases the level of inflows. In particular, governance and institutional indicators seem to have a quantitatively significant influence on FDI inflows. Based on the distribution of US multinational firms around the world, Hines (1995) reports that American companies tend to invest less in destination countries where levels of corruption are higher. Using bilateral stocks of FDI from 12 OECD source countries to 45 host countries, Wei (2001) shows that countries' corruption levels are negatively associated with inward FDI (Figure 12). An increase in the corruption level from that of Singapore to that of Russia has the same negative effect on FDI as raising the marginal corporate tax rate by as much as 50 percentage points. Moreover, for any given level of corruption, less centralized and more arbitrary types of corruption tend to discourage FDI even more strongly. Better governance also appears to lead to more equity inflows (see Gelos & Wei, 2005).

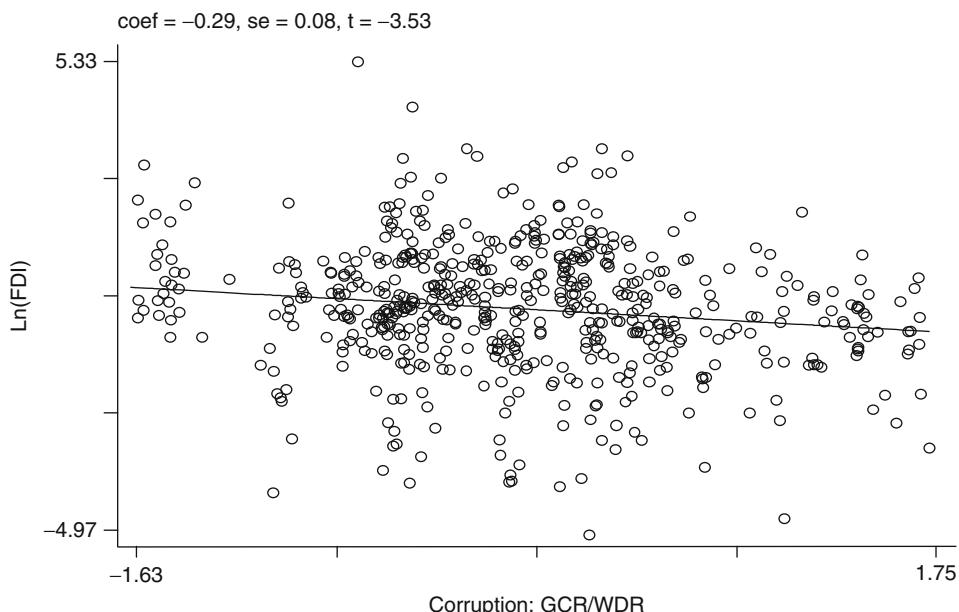


Figure 12 Corruption and foreign direct investment. Notes: Bilateral foreign direct investment from 14 major source countries to 41 host countries, averaged over 1996–1998. Index of host country corruption is derived by combining the measures from the Global Competitiveness Report (World Economic Forum and Harvard University, 1997) and World Development Report (World Bank 1997). More details can be found in Wei (2001).

There is a considerable body of evidence that institutions affect the structure of a country's capital inflows in a systematic way. This has important consequences for the outcomes associated with financial globalization since the composition of inflows seems to have strong predictive power for currency crashes. In particular, the share of FDI in a country's total capital inflow is negatively associated with the probability of a currency crisis (see, e.g., Frankel & Rose, 1996; Frankel & Wei, 2005). Other dimensions of composition are the maturity structure of external debt (the greater the share of short-term debt, the more likely a crisis) and the currency denomination of external debt (the greater the share of foreign currency debt, the more likely a crisis).³⁵

Wei (2001) and Wei and Wu (2002) suggest that countries with better public institutions are more likely to attract more direct investment relative to bank loans. These authors provide evidence based on total inflows (based on data from the IMF's Balance-of-Payments Statistics) and bilateral flows from source to destination countries (based on bilateral FDI data from the OECD and bilateral bank lending data from the BIS) (Figure 13).

Faria and Mauro (2005) find that better institutional quality helps tilt a country's capital structure toward FDI and portfolio equity flows which tend to bring more collateral benefits of financial integration. These authors find that, in a cross-section of

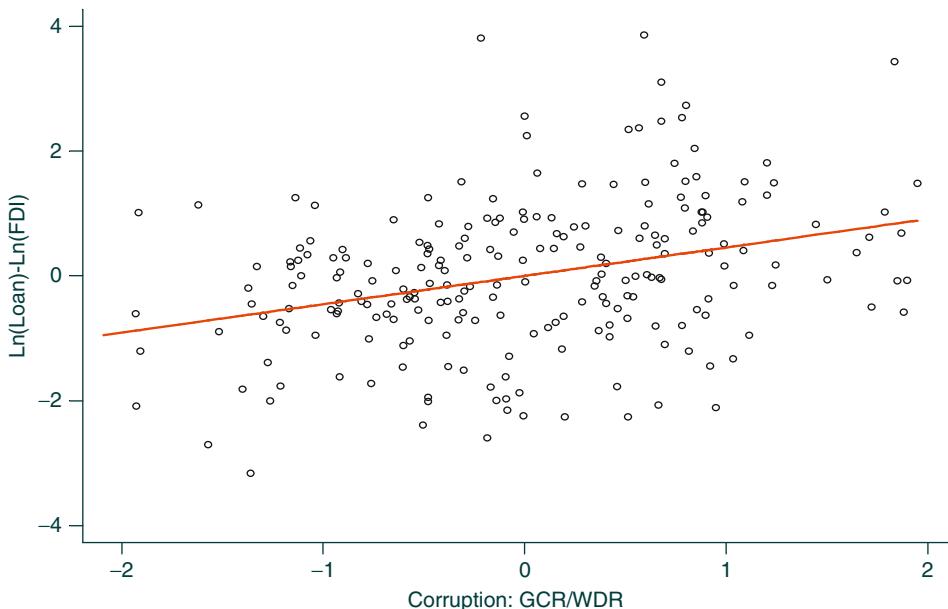


Figure 13 Corruption and the composition of capital flows. *Notes:* The index of host country corruption is derived by combining the measures from the Global Competitiveness Report (World Economic Forum and Harvard University, 1997) and World Development Report (World Bank 1997). Controlling for recipient country fixed effects, size, level of development, policy incentives and restrictions on FDI, geographic and linguistic connections.

emerging markets and other developing countries, equity-like liabilities as a share of countries' total external liabilities (or as a share of GDP) are positively associated with indicators of institutional quality.³⁶

5.3 Trade openness

5.3.1 Theory

Trade integration improves the growth and stability benefits of integration through various channels. First, trade integration reduces the probability of crises associated with sudden stops and current account reversals. Economies that are less open to trade have to undergo larger real exchange rate depreciations for a given current account adjustment, face more severe balance sheet effects stemming from depreciations, and, as a result, are more likely to default on their debt. This creates a link between the probability of sudden stops and the likelihood of default, implying that more open economies are less vulnerable to financial crises.³⁷ In addition, trade integration is also expected to mitigate the adverse growth effects of financial crises and facilitate recoveries from crises. It could help an economy to continue servicing its debt and export its way out of a recession since a given exchange rate depreciation would have a larger impact on its export revenues than in a less open economy.

Moreover, trade integration in general seems to be less risky than financial integration. For example, while trade integration can apparently proceed well even in the absence of financial integration, financial integration in the absence of trade integration could lead to a misallocation of resources. Eichengreen (2001) notes that, under these circumstances, capital inflows may be directed to sectors in which a country does not have a comparative advantage. Martin and Rey (2006) constructs a model in which trade integration has a positive growth effect, but financial integration can lead to asset price crashes and financial crises. They argue that costs associated with international trade in goods and assets alone could increase the vulnerability of developing countries to financial crises.³⁸ The model has a clear implication—consistent with the received wisdom—that developing countries should liberalize trade in goods before trade in financial assets.

5.3.2 Empirical evidence

There are many papers validating the traditional preference for liberalizing trade flows ahead of financial flows, but the empirical evidence that trade integration significantly affects the relationship between financial integration and growth is mixed (see Table 4). Using trade openness (exports plus imports to GDP) with FDI inflows, Balasubramanyam, Salisu, and Sapsford (1996) found a positive coefficient on FDI for a sub-sample of countries with higher imports to GDP but Carkovic and Levine (2005) do not find significant interaction effects. Gupta and Yuan (2008) use sectoral-level data and find that there is higher growth following international equity market liberalizations in

Table 4 Summary of empirical studies on structural factors and growth (trade openness measures)

Study	No. of econ./ period	Econometric methodology	Dependent variable	Financial openness variable	Interaction/ threshold variables	Interaction/ threshold approach	Main findings on interaction effect
Balasubramanyam et al. (1996)	Up to 46 developing economies (1970–1985)	Cross-section OLS and IV	Growth of real per capita GDP (PWT)	FDI to GDP	Imports to GDP (classify economy as import substituting IS or export promoting EP)	Regression estimated separately for the two sample splits	Significant difference across IS or EP samples. <i>Coefficient on FDI positive significant for EP countries but insignificant for IS.</i>
Arteta et al. (2003)	Up to 60 (1973–1992)	Cross-section and subperiod panel pooled OLS	Growth of real per capita PPP GDP	Initial value of Quinn index	Overall Sachs Warner (SW) openness indicator plus subcomponents of tariff/nontariff barriers and black market premium	Linear	Interaction with SW generally positive significant but insignificant if SW level added. Interaction with trade barriers measure insignificant. <i>Results supportive of positive relation between growth and capital account liberalization contingent on absence of large</i>

							<i>black market premium</i> (interpreted as absence of macro imbalances).
Carkovic and Levine (2005)	Up to 67 (1960-1995)	Cross-section OLS and 5-yearly panel dynamic system GMM	Growth of real per capita GDP	Gross FDI inflows to GDP	Exports plus imports to GDP	Linear	<i>Interaction effect insignificant in OLS regressions.</i> <i>Positive significant results for panel not robust to inclusion of other controls.</i>
Gupta and Yuan (2008)	31 emerging economies (1981-1998)	Annual sector level panel, country and year FE.	Growth of real sectoral value added	Liberalization of stock market to foreign investors.	“Trade competitiveness,” that is, ratio of industry to total output of that industry across sample	Linear	Growth postliberalization is significantly higher in industries which are more trade competitive.

Notes: Samples and methodology: Details in table relate to sections of study when interaction effects are examined (and hence, may differ from other parts of papers). Data for Balasubramanyam et al. (1996) from table 1, for Arteta et al. (2003) from tables 5–7, for Carkovic and Levine (2005) from table 8.6 and for Gupta and Yuan (2008) from table 5. Significant if at least 10% significance level.

Note: Information for Arteta et al. (2003) taken from NBER working paper no. 8414 version.

those sectors that are more trade competitive (defined as the ratio of annual exports in each industry to total annual output of that industry across all sample countries).

Citing the lack of empirical evidence from a couple of empirical studies focusing on Morocco and Venezuela, Rodrik (1999) argues that FDI has no extra benefit to host country development.³⁹ Moran (2005) contests this argument, noting that both of these countries practiced import-substitution based trade policies during the periods analyzed in these papers. He provides several case studies showing that the full benefits of FDI are realized only in an environment with minimal distortions from trade barriers and other protectionist policies.

However, there appears to be evidence favoring other channels emphasized by the theory. For example, the negative impact of trade openness on the likelihood of sudden stops is indeed empirically important. Calvo, Izquierdo, and Mejía (2004) show that trade openness makes countries less vulnerable to financial crises, including sudden stops and currency crashes; controlling for the endogeneity of trade strengthens this effect. Frankel and Cavallo (2004) and Cavallo (2007) report similar findings. They conclude that a 10 percentage point increase in trade openness reduces the probability of a sudden stop by about 30%.

Does trade integration play an important role during the recovery phases from crises? Calvo and Talvi (2005) claim that the collapse of capital flows to Argentina and Chile in the 1990s had a small impact on Chile since it is more open to trade flows.⁴⁰ Recent research also confirms that, among countries that have experienced sudden stops and current account reversals, those that are more open to trade suffer smaller growth declines.⁴¹ For example, Edwards (2005) reports that a decline in trade openness by roughly 30 percentage points increases the negative effect of a current account reversal on growth by approximately 1.2 percentage points.

Recent research also analyzes how trade and financial integration affect the negative relationship between volatility and economic growth. Although countries prone to higher macroeconomic volatility would be expected to show worse growth performance than more stable ones, this interpretation does not seem to be entirely borne out by the data.⁴² In particular, while emerging market countries affected by the recent financial crises faced episodes of high output volatility, they actually posted much better growth rates on average during the past two decades than other developing countries. In particular, during the period 1986–2005, while a number of emerging markets experienced financial crises, their average growth of output was more than two times higher than that of other developing economies. Does this mean that, in a period of rising globalization, the negative relationship between volatility and growth has changed?

Recent research addresses this question by studying the relationship between growth and volatility in a large sample of countries over the past four decades (Kose, Prasad, & Terrones, 2005, 2006). The results indicate that while the negative relationship between

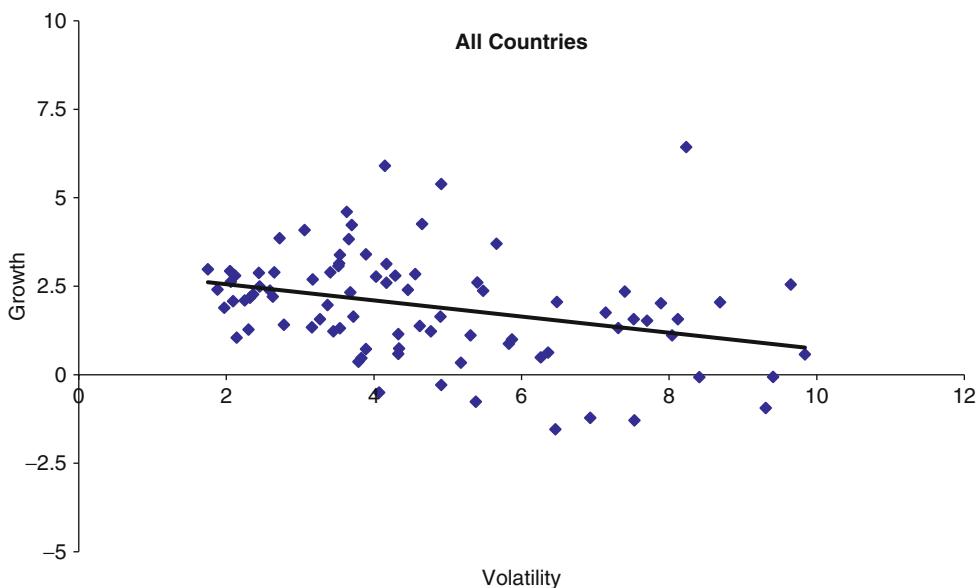


Figure 14 Output growth and volatility (simple correlations). Notes: Growth is the average annual growth of real GDP. Volatility is the standard deviation of annual real GDP growth. The time period of analysis is 1960–2000.

growth and volatility reported by previous research for the period 1960–1985 has persisted into the 1990s, when it is broken down by country groups, it is far from homogeneous (Figure 14). The relationship appears positive for advanced countries—indicating that, for countries in advanced stage of development and integration into the global economy, volatility is not necessarily associated with lower growth (Figure 15). Among developing countries, it is positive on average for emerging market economies and negative for the other developing countries that have not participated as much in the process of globalization.

Both trade and financial integration appear to have played important roles in changing the nature of the volatility and growth relationship for emerging market economies. The relationship between growth and volatility is negative before trade liberalization and positive after liberalization (Figure 16). For financial integration, there is a similar, although less strong, result. In other words, there is suggestive evidence from emerging market economies that both trade and financial integration change the sign of the relationship between volatility and growth.

Regression analysis also suggests that although volatility is still negatively associated with growth, higher trade and financial integration make this relationship weaker. In other words, economies that are more integrated into the global economy have the ability to withstand higher levels of volatility with less adverse effects on growth.⁴³

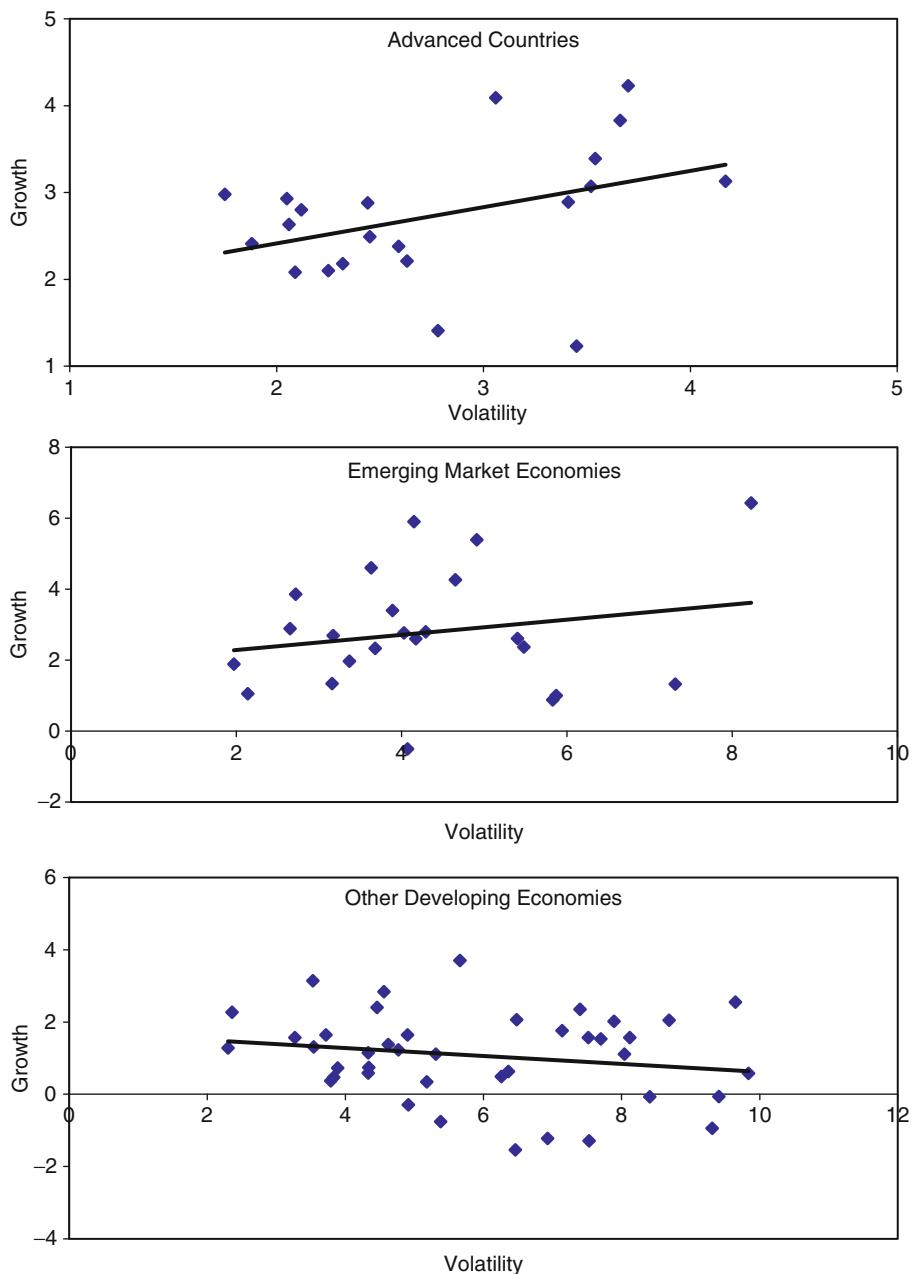


Figure 15 Output growth and volatility (simple correlations). Notes: Growth is the average annual growth of real GDP. Volatility is the standard deviation of annual real GDP growth.

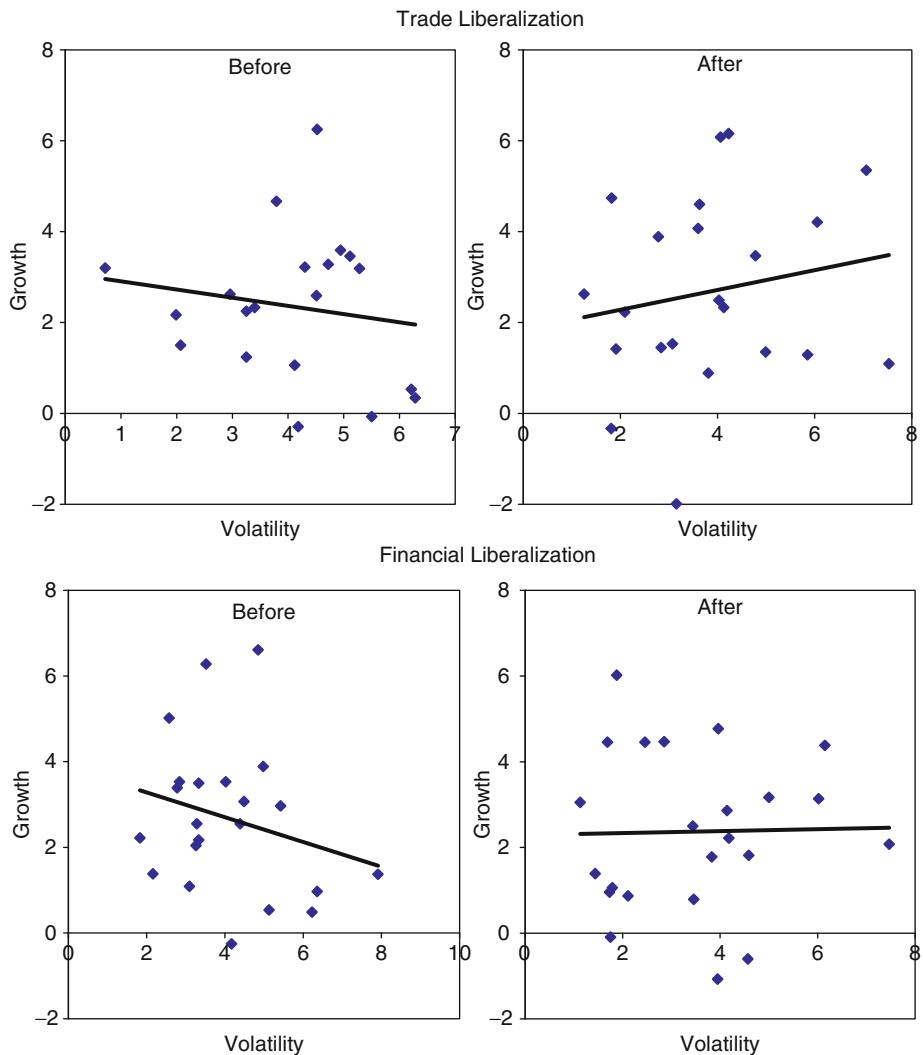


Figure 16 Output growth and volatility in emerging economies (before and after trade, and financial liberalization, simple correlations). Source: Growth is the average annual growth of real GDP. Volatility is the standard deviation of annual real GDP growth. The dates of official liberalization are determined on the basis of stock market liberalization and removal of restrictions on foreign investment based on the Bekaert, Harvey, and Lundblad (2005).

Moreover, they argue that these effects may be quantitatively important. When comparing the growth/volatility performance of advanced and developing countries, it is found that the higher levels of trade/financial openness of emerging markets could, under some assumptions, account for about 40% of the observed difference in average growth rates between these country groups. Overall, these findings suggest that the forces of trade and financial integration could help reduce the adverse impact of volatility on economic growth.

6. MACROECONOMIC POLICIES

6.1 Theory

Financial integration is expected to produce better growth outcomes when it is supported by good macroeconomic policies, which includes fiscal, monetary, and exchange rate policies (Eichengreen, 2000). Moreover, weak or incompatible policies can increase the risk of crises from an open capital account. For instance, the combination of a fixed exchange rate and an open capital account has been implicated in a number of currency crises (Obstfeld & Rogoff, 1995; Wyplosz, 2004). Similarly, managing the effects of capital inflows can be especially complicated in developing economies with large fiscal deficits and procyclical fiscal policy (Kaminsky, Reinhart, & Végh, 2004). All of this suggests that opening the capital account creates incentives for policymakers to maintain stable policies as discussed in Section 3. This logic has led to the proposition that capital account liberalization can serve as a commitment device for disciplining macroeconomic policies (Bartolini & Drazen, 1997; Gourinchas & Jeanne, 2006). A different view is propounded by Neut and Velasco (2003) who argue that such a commitment device could actually backfire in the presence of uncertainty—the possibility of unavoidable debt defaults due to large adverse shocks could result in a conservative policymaker having lower credibility.

6.2 Evidence

There is a large literature tying the quality of policies to macroeconomic performance, vulnerability to crises, and the level and composition of inflows. For example, using the black market premium on the domestic currency as an indicator of the extent of macro imbalances, Arteta et al. (2003) report evidence of threshold effects related to good macro policies in generating *positive* growth effects of financial openness. These positive effects are present only when macroeconomic imbalances that lead to inconsistencies between the administered exchange rate and other policies have first been eliminated (i.e., if there is no large black market premium). Mody and Murshid (2005) examine how policies affect the relationship between financial flows and domestic investment growth. Using a composite variable of macroeconomic policy quality constructed by the World Bank, they find that financial flows have a stronger impact on investment growth in countries with better macro policies. In models of early warning systems, proxies for exchange rate and monetary policies appear to be important for predicting financial crises.⁴⁴

These results are further supported by a large volume of case studies. For example, IMF (2007) undertook a study analyzing a number of countries' experiences with the process of capital account liberalization. It concludes that while the speed of liberalization process does not appear to affect the crisis propensity, countries with increasing inflation, expansionary fiscal policies and rising current account deficits are more likely to experience a financial crisis than those with relatively lower inflation, strong fiscal positions, and low current account deficits. These results confirm the earlier findings by Ishii

et al. (2002) in the country case study which underscores the importance of stable macro policies for averting crises in countries with open capital accounts.

With respect to fiscal policy, for obvious reasons there has been a large research program focusing on the importance of fiscal prudence. Countries that consistently face problems associated with government debt (referred to as “serial defaulters” by Reinhart & Rogoff, 2004) are more likely to benefit from financial globalization if their governments simultaneously take policy measures to avoid an excessive buildup of debt. In a related paper, Reinhart, Rogoff, and Savastano (2003) study the concept of “debt intolerance,” which manifests itself in the extreme duress many emerging market economies experience at overall debt levels that would seem quite manageable by the standards of the advanced industrial economies. They conclude that, for debt-intolerant countries, mechanisms to limit borrowing through institutional change on the debtor side can greatly limit the risks associated with financial integration.⁴⁵

The design of macroeconomic policies naturally becomes more complex during the periods of large capital inflows.⁴⁶ Cardarelli, Elekdag, and Kose (2007) examine the macroeconomic outcomes associated with various policy responses in over 100 episodes of large net capital inflows in a number of countries during the past two decades. Their results emphasize the importance of employing disciplined fiscal and monetary policies to cope better with the macroeconomic effects of large capital inflows.⁴⁷ For example, keeping government spending along a steady path—rather than engaging in excessive spending during inflow episodes—appears to mitigate the adverse effects of large inflows as it helps reduce upward pressures on both aggregate demand and the real exchange rate. With respect to monetary policies, they document that a higher degree of resistance to exchange rate pressure during the inflow period and a greater degree of sterilization of the resulting increase in money supply were not associated with lower real appreciation or with better postinflow growth performance.

6.2.1 Monetary and exchange rate policies

There has been a growing research program analyzing the links between financial globalization and exchange rate policy.⁴⁸ The choice of appropriate exchange rate regime has received much attention in the literature (see the chapter by Levy Yeyati & Sturzenegger (2009) in this handbook). An open capital account is likely to put a greater burden on other policies and structural features of the economy (e.g., product and labor market flexibility) to support a fixed exchange rate. In particular, for economies with weak financial systems, an open capital account and a fixed exchange rate regime are not an auspicious combination. Indeed, there is a compelling case to be made that rigid exchange rate regimes can make a country more vulnerable to crises when it opens its capital markets. For example, Prasad, Rumbaugh, and Wang (2005) survey a number of industrial and developing country experiences showing that the combination of capital account liberalization and a fixed exchange rate regime have often ended in forced and messy exits to more flexible exchange rate regimes. It can be argued that, in the absence of *de facto* or *de jure* fixed rates,

most of the crises of the 1990s, from Mexico to Asia to Russia to Brazil, might have been much less virulent, or might even have been avoided entirely.

However, the literature does not imply that fixed exchange rates are necessarily a problem for countries that are at early stages of domestic financial development or that are inappropriate prior to international capital market liberalization. Husain, Moday, and Rogoff (2005), using the *de facto* approach to classify exchange rate regimes developed by Reinhart and Rogoff (2004), find that pegged exchange rate regimes confer some advantages such as lower inflation upon developing countries that do not have much exposure to international capital. For emerging markets, standard measures of macroeconomic performance are not systematically associated with the nature of the exchange rate regime, but the likelihood of financial crises is higher for countries with pegged or nearly pegged exchange rates. Husain, Mody and Rogoff attribute the latter result under a regime with “hard commitment” to the inability to adapt to changed circumstances, the incentives of economic agents including entrepreneurs and financial intermediaries to undertake risky activities on the presumption that exchange rates will not change, and speculative pressures from investors who seek to test the commitment.⁴⁹

Wyplosz (2004) highlights the difficulties and risks associated with maintaining currency pegs when the capital account is open. As a short-term strategy for developing economies, he recommends a combination of a soft peg or managed exchange rate regime along with well-designed limit on capital mobility. Maintaining either a free float or a hard peg along with capital account openness requires a strong commitment to fostering good institutions, especially with respect to financial market regulation and supervision.

In a recent paper, Rodrik (2007a, 2007b) suggests that an undervalued real exchange rate helps promote economic growth. His main argument is that the tradable sector in a developing economy is more likely to suffer from various institutional and market failures than the nontradable sector. A subsidy in the form of an undervalued exchange rate is a useful offset to these problems. He notes that there are a variety of policies through which countries can keep their real exchange rates undervalued. Among other policy measures, his list also includes imposition of controls on inflows, liberalization of capital outflows, and intervention in foreign exchange markets in order to sustain exchange rate undervaluation.

7. DIRECTION OF CAPITAL FLOWS AND ECONOMIC GROWTH

7.1 Empirical evidence

As discussed earlier in Section 3, standard economic theory predicts that financial flows should, on net, flow from richer to poorer countries. That is, it should flow from countries that have more physical capital per worker—and hence, where the returns to capital are lower—to those that have relatively less capital—and hence, greater unexploited investment opportunities. In principle, this movement of capital should make poorer countries better off by giving them access to more financial resources that they can then invest in physical capital and improve their growth prospects.

However, the actual volumes of such flows do not come anywhere near what the baseline models predict, as famously emphasized by Lucas (1990).⁵⁰ Remarkably, as shown by Prasad, Rajan, and Subramanian (2007), this paradox has, if anything, intensified over time as financial globalization has picked up momentum.⁵¹ The average income, relative to the United States, of capital-exporting countries has fallen well below that of capital-importing countries (Figure 17). In other words, capital has been flowing from poor to rich countries implying that

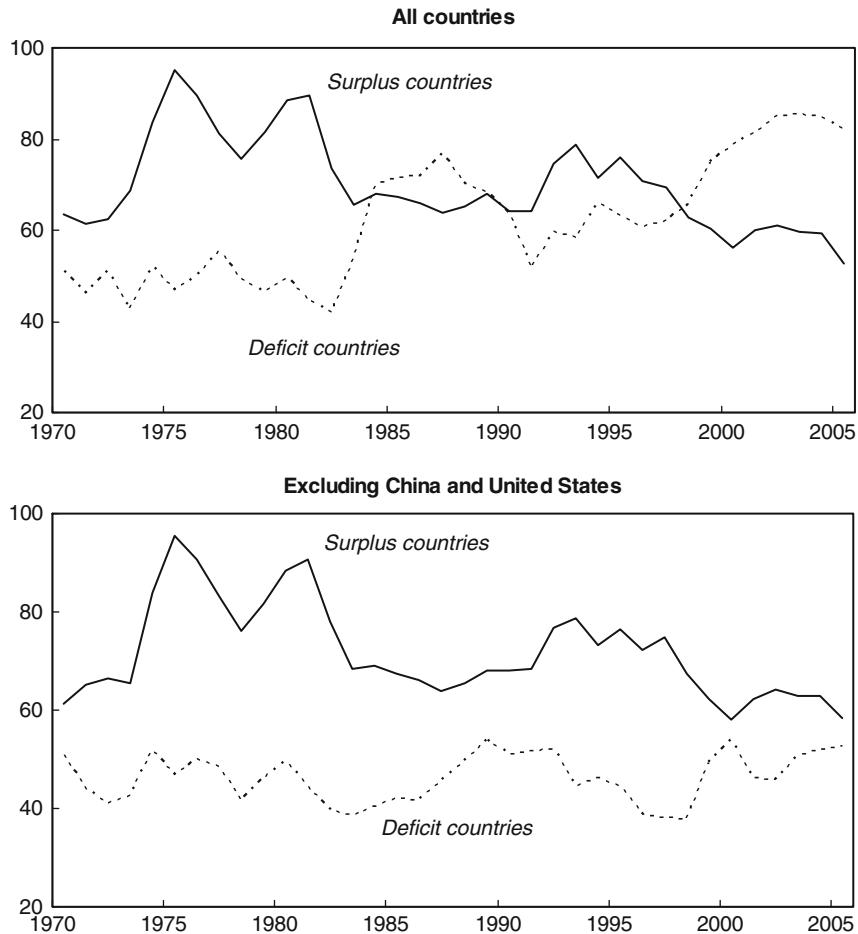


Figure 17 Relative GDP per capita of capital exporters and capital importers (percent of highest GDP per capita in indicated year). Notes: Each observation is the average GDP per capita (weighted by the country's share of the total current surplus or deficit) of countries in the WEO database with current account surpluses or deficits in the indicated year, expressed as a percentage of GDP per capita in the country with the highest GDP per capita that year. GDP per capita is adjusted for purchasing power parity. Raw data from the WEO database. The period of analysis is 1970-2005.

increasing the quantum of financing available for investment is not the channel through which financial globalization delivers its benefits for developing countries.

Prasad et al. (2007) provide further evidence why the direct channel emphasized by the standard theory does not work. In particular, they examine the long-run relationship between current account balances and growth. Countries that borrow more from abroad should be able to invest more (since they are less constrained by domestic saving) and, therefore, should grow faster. Surprisingly, for their sample of nonindustrial countries, the correlation between growth and current accounts using data averaged over a long period for each country is positive (Figure 18). In other words, developing countries that have relied *less* on foreign finance have grown faster in the long run. That is not to say there are no episodes where nonindustrial countries grow fast and run large current account deficits—East Asia before the crisis is a clear counter example. But, looking beyond a few short-run foreign-funded booms (and possibly busts), on average, and in the long run, nonindustrial countries that grow the fastest have not depended much on foreign finance.

7.2 How to interpret these findings?

How does one interpret the finding that there is a positive correlation between the current account surplus and a country's growth rate? One possible explanation is that the relationship reflects and is driven by domestic savings, which is either determined by deeper forces or generated through growth itself. After all, if foreign inflows responded

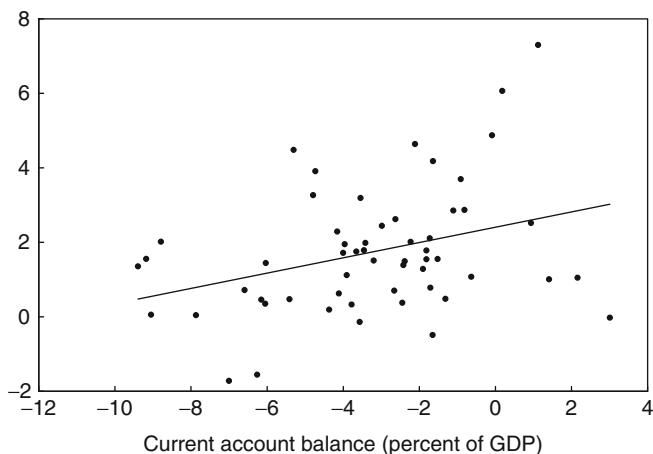


Figure 18 Growth in GDP per capita and level of current account balances (growth in GDP per capita; % a year). Notes: Data are for the fifty-six non-industrial countries in the core sample. Raw data from the Penn World Tables and the World Bank, World Development Indicators. The period of analysis is 1970-2004.

largely to investment opportunities, there should be an unambiguously negative relationship between growth and the current account.

Indeed, it turns out that the positive correlation is driven largely by national savings. That is, nonindustrial countries that have higher savings for a given level of investment experience higher growth. Of course, investment in high-saving countries could also be higher, so high domestic savings does not imply low reliance on foreign savings. As expected, countries with higher levels of investment do fare better than those with lower levels (Figure 19). What is interesting is that countries that had high investment ratios *and* lower reliance on foreign capital (lower current account deficits) grew faster—on average, by about 1% a year—compared with countries that had high investment but also a greater degree of reliance on foreign capital.

One explanation for the positive correlation between the current account surplus and a country's growth rate is that higher growth is associated with—and itself generates—higher domestic savings. In other words, fast growing countries may *need* less foreign capital. The problem is that, typically as countries grow (i.e., when they experience a positive productivity shock) they should want to consume more (because they are richer) and invest more (because of the investment opportunities). Thus, the correlation should, if anything, be negative.

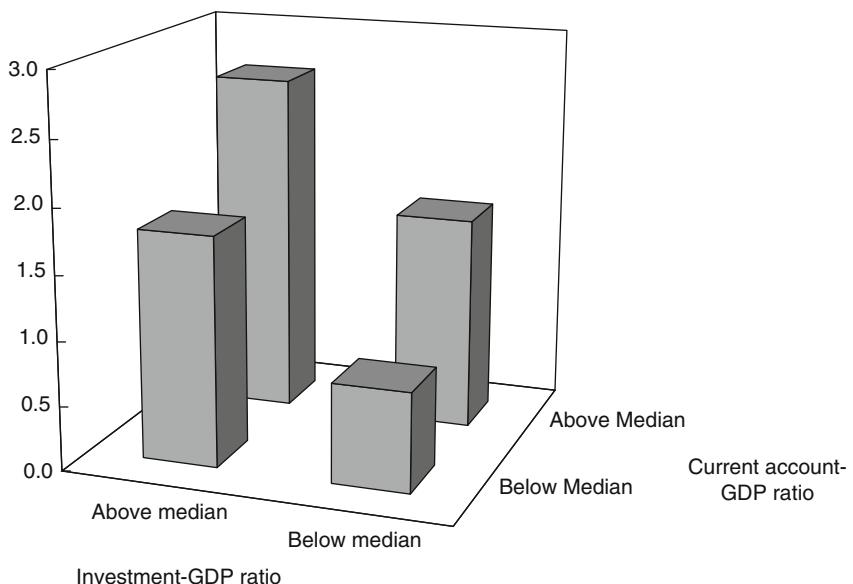


Figure 19 Growth in GDP per capita and levels of investment and the current account (growth in GDP per capita; % a year). Notes: Data are for the fifty-nine non-industrial countries in the entire sample plus Bangladesh. All data are period averages. Raw data from the World Bank, World Development Indicators. The period of analysis is 1970-2004.

This is where the financial system—especially an underdeveloped one—can play a role. If the financial sector were deep and efficient, a sustained increase in productivity would not only result in more investment (as firms borrow to take advantage of investment opportunities), but also in more consumption as consumers borrow to consume in anticipation of their higher income. Conversely, a weak financial sector could translate a sustained increase in the productivity of certain sectors into weaker investment growth and greater savings growth. Corporate investment could be limited to the funds firms generate internally from past investment, while consumers save much of the increased income stemming from the increase in productivity because they cannot borrow in anticipation of higher future income.

Another possibility is that weak financial systems may not help in efficiently intermediating foreign capital. This too could result in the lack of a positive relationship between flows of foreign capital and higher growth. Consistent with the views that foreign capital may not be needed nor be helpful because of weak financial systems, the positive correlation between the current account balance and growth turns out to be stronger for the group of countries with less well-developed financial systems. In these countries, the range of profitable investment opportunities, as well as private consumption, for those that experience growth episodes, may be constrained by financial sector impediments, so investment can be financed largely through domestically generated savings.

Excessive reliance on foreign capital may also have harmful consequences. It can lead to currency appreciation and, in some circumstance, overvaluation (a level of the exchange rate that is higher than the level warranted by economic fundamentals). In turn, this could hurt competitiveness and exports in key sectors like manufacturing. Recent analyses of growth episodes suggest that a dynamic manufacturing sector is a key to long-run growth. Thus, reduced reliance on foreign capital—and the avoidance of overvaluation—may help the development of an export-oriented manufacturing sector.

7.3 Are developing countries savings-constrained?

The results discussed above are consistent with Rodrik's (2007a) view that developing economies are investment-constrained rather than savings-constrained. That is, the conventional notion that access to foreign capital would expand investment opportunities in developing countries does not seem to be borne out by the data. Moreover, when capital inflows lead to exchange rate overvaluation, the net effect on growth can be negative. Rodrik (2007b) goes further, however, to suggest that a systematic policy of keeping the exchange rate undervalued by maintaining a relatively closed capital account and intervening in the foreign exchange market when necessary is good for a developing country's growth. Prasad et al. (2007) find no evidence in support of this proposition—that is, overvaluation does seem to be bad for growth but it is not obvious that undervaluation is good for growth.⁵²

7.4 Policy implications

What does all this mean for policies toward financial integration? Any discussion on the merits of financial integration is likely to be very specific to a country. These results suggest that—insofar as the domestic financial sector is underdeveloped and there is a need to avoid exchange rate appreciation caused by inflows—greater caution toward certain forms of foreign capital inflows might be warranted. At the same time, financial openness may itself be needed to spur domestic financial development and to reap the benefits that financial flows and better growth opportunities provide.

How can this tension be resolved? One approach might be a firm—and hopefully credible—commitment to integrate financial markets at a definite future date, thus giving time for the domestic financial system to develop without possible adverse effects from capital inflows, even while giving participants the incentive to prepare for it by suspending the sword of future foreign competition over their heads. A recent example of this is the Chinese approach of trying to spur banking reform by committing to open up their banking sector to foreign competition as part of their obligations for accession to the World Trade Organization. Another possibility is to focus on specific collateral benefits that may be relevant for a particular country and to try to design an appropriate capital account liberalization program that would deliver that limited set of benefits.

8. CAPITAL CONTROLS AS A POLICY TOOL

Although financial flows can potentially lead to long-term growth benefits, heavy capital inflows may pose significant challenges to macroeconomic stability. A key policy question for developing economies is to determine how to effectively manage these large inflows. Capital controls are one of the more controversial choices available to policymakers during periods of large capital flows. Countries employ control measures to attain a variety of policy objectives, such as discouraging capital inflows to reduce upward pressures on the exchange rate, reducing the risk associated with the sudden reversal of inflows, and maintaining some degree of monetary policy independence (Magud & Reinhart, 2007). After a brief overview of the different types of capital controls and their measurement, this section provides a summary of macroeconomic and microeconomic implications of capital controls.

8.1 Implementation and measurement issues

Although capital controls cover a wide range of measures regulating inflows and outflows of foreign capital, they generally take two broad forms: direct (or administrative) and indirect (or market-based) controls. Direct controls are associated with administrative measures, such as direct prohibitions and explicit limits on the volume of transactions. For example, Malaysia introduced a set of direct capital controls in 1998

involving various quantitative restrictions on cross-border trade of its currency and credit transactions. Indirect capital controls include explicit or implicit taxation of financial flows and differential exchange rates for capital transactions. For example, in order to discourage capital inflows Chile imposed an implicit tax in 1991 in the form of an unremunerated reserve requirement (URR) on specified inflows for up to 1 year. These controls were substantially relaxed in 1998 (see Magud & Reinhart, 2007).

Using a variety of capital control measures, a large literature has studied their macroeconomic and microeconomic implications. However, irrespective of their type, it is a challenge to effectively quantify the extent of capital controls as discussed in Section 2. In particular, it would be desirable to capture the degree of enforcement of capital controls. Moreover, the impact of a measure would depend on a broad assessment of the openness of the capital account.

8.2 Macroeconomic implications of capital controls

The literature assessing whether capital controls have attained their stated macroeconomic objectives is, at best, mixed.⁵³ It is hard to draw a set of general results as most of the studies are based on country cases (Ariyoshi et al., 2000). Overall, the studies suggest that controls on inflows did not affect the volume of net flows in most countries, although it seems that they were able to temporarily tilt the composition toward longer maturities in a few cases (Chile after 1991, see Edwards & Rigobon, 2005).⁵⁴ Even in cases where a narrow range of objectives were met, controls had only temporary effects as market participants eventually found ways to circumvent them. Clearly, however, there is a cost to market participants of evading such controls, which effectively acts as a tax on inflows. An open question is whether this tax simply imposes a distortion cost without affecting the volume of flows significantly, or if the tax is large enough to reduce flows materially at least in the short run.

Cardarelli et al. (2007) analyze the effectiveness of capital controls of various policy responses in over 100 episodes of large net private capital inflows in a group of emerging market countries and advanced economies since 1987. They find that episodes characterized by tighter controls on inflows are associated with narrower current account deficits and lower net private inflows, including lower net FDI flows (Figure 20). While stricter inflow controls are accompanied by lower-postinflow growth and a larger appreciation of the currency, these distinctions are not statistically significant. In contrast, inflation rates have been significantly higher in episodes with tighter controls.

Does having capital controls in place reduce the vulnerability to financial crises and sudden stops? Episodes that ended in an abrupt reversal of net inflows do not seem to be associated with lower capital controls (Figure 21). On the contrary, although the differences are not statistically significant, Cardarelli et al. (2007) report that episodes that ended abruptly were associated with somewhat stricter inflow controls. These findings are consistent with results documented in Section 3 about the smaller likelihood of crises in countries with more open capital account regimes.⁵⁵

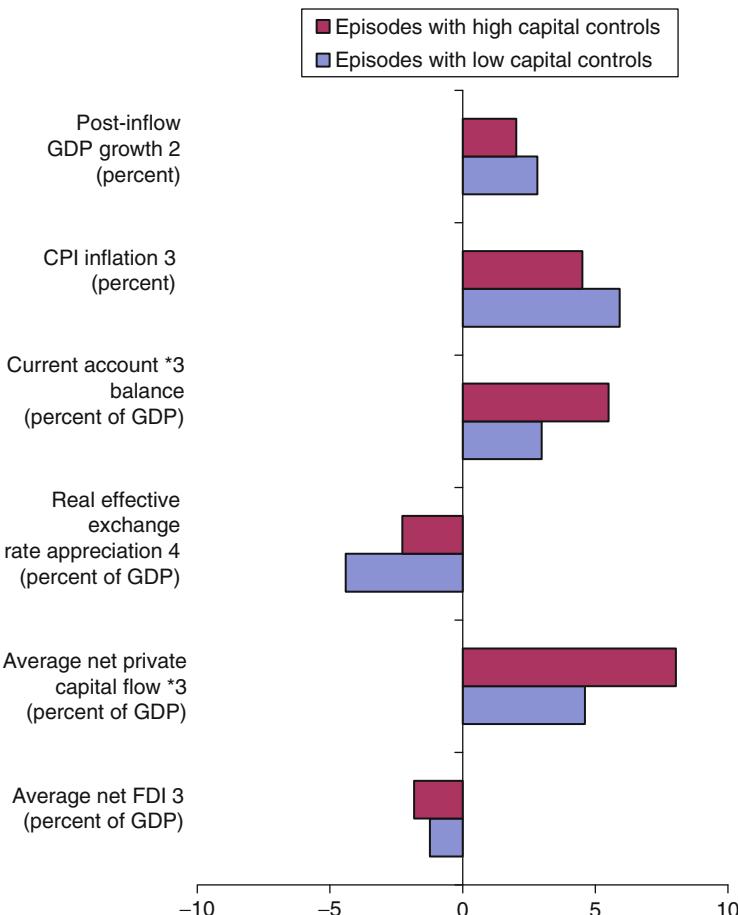


Figure 20 Controls on capital inflows and selected macroeconomic indicators. 1. Values reported are medians for the two groups of episodes. Episodes with high (low) capital controls are those with above (below) median values of the capital controls index discussed in the text, where higher (lower) values indicate tighter (looser) regulation of inflows. The asterisk (*) indicates that the difference between medians is significant at a 10 percent confidence level or better. 2. Average real GDP growth in the two years after an episode minus average during the episode. 3. Average during the episode. 4. Cumulative change during the episode. Data Sources: IMF, Annual Report on Exchange Arrangements and Exchange Restrictions; IMF, Balance of Payments Statistics; and IMF staff calculations.

8.3 Microeconomic implications of capital controls

Although the literature analyzing the macroeconomic implications of capital controls is unable to produce conclusive evidence, recent studies using micro data find that controls result in significant efficiency costs at the level of individual firms or sectors (see Forbes, 2005a for a survey). These costs often manifest themselves through indirect

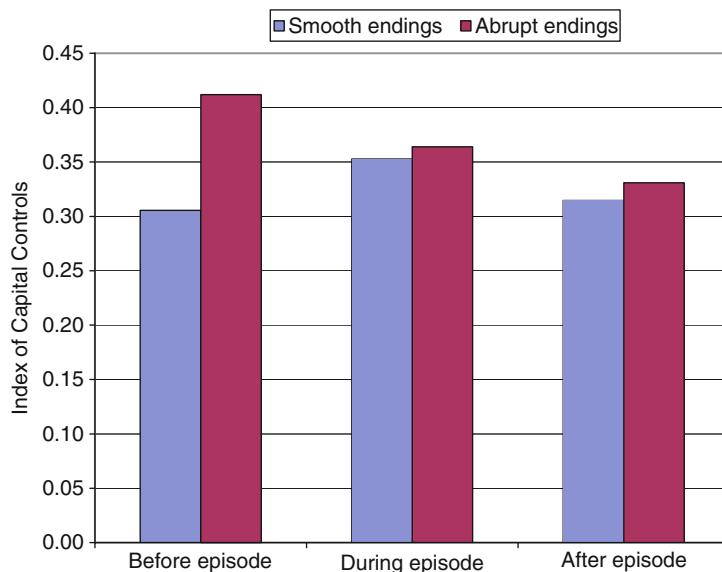


Figure 21 Endings of episodes and controls on capital inflows. Notes: Median values across all completed episodes using the index of capital controls discussed in the text, where higher values indicate tighter regulation of inflows. “Before” denotes averages of the index in the two years before the episode. “After” denotes averages of the index in the two years after the episode. An episode is considered to end “abruptly” if the ratio of net private capital inflows to GDP in the year after the episode terminates is more than 5 percentage points of GDP lower than at the end of the episode—closely following the definition of “sudden stops” in the literature. An episode is also considered to finish abruptly if its end coincides with a currency crisis, that is, with a steep depreciation of the exchange rate. Data Sources: IMF, Annual Report on Exchange Arrangements and Exchange Restrictions; and IMF staff calculations.

channels. For example, capital controls result in an increase in the cost of capital, reduce market discipline, create distortions in the behavior of firms and individuals, and impose substantial administrative costs on the government. There is also some recent evidence suggesting that capital controls involving exchange rate transactions act as nontariff trade barriers reducing the volume of trade flows.

8.3.1 Higher cost of capital

Some recent studies argue that the Chilean capital controls increased the financial constraints for smaller firms by making it more difficult and expensive for them to raise capital (Forbes, 2005b). The cost of capital is higher for multinationals as well when controls are in place (Desai, Foley, & Hines, 2004). For example, multinational affiliates located in countries with capital controls are found to face interest rates that are about 5 percentage points higher than affiliates of the same parent company borrowing locally in countries without capital controls. The wedge arises because capital controls

typically result in costs of avoidance of those controls as well as higher domestic interest rates. The cross-country investment patterns of multinationals suggest that the level of FDI inflows into a country is adversely affected by capital controls because of their impact on the cost of capital.

8.3.2 Reduced market discipline

By insulating the economy from competitive forces, capital controls reduced the market discipline among Malaysian firms and created a screen for cronyism (Johnson & Mitton, 2003). The cost of these controls was quite significant because the ability of the government to provide subsidies to politically connected firms increased with the imposition of controls.

8.3.3 Distortions in firm/individual behavior and cost on government

Capital controls can cause distortions in the behavior of firms (and individuals) with the objective of evading the controls. For example, when the Argentine government imposed capital controls at the end of 2001, in order to evade the controls investors came up with a variety of schemes leading to a large volume of capital outflow.⁵⁶ Since the controls often have to be continually updated in order to close loopholes and limit evasion, they involve significant administrative and monitoring costs (Forbes, 2005a).

8.3.4 Controls as nontariff barriers

In countries with extensive capital controls, firms and individuals often try to circumvent them by mis-invoicing imports and exports, or both. In response, authorities have to implement various inspections and impose extensive reporting requirement to minimize such leakages. These increase the cost of engaging in international trade even for those firms that do not intend to evade capital controls (Wei & Zhang, 2006). Such controls have economically and statistically significant negative effects on the volume of trade. For example, a one standard deviation increase in the controls on foreign exchange transactions reduces trade by the same amount as a rise in tariff by 10.8–11.3 percentage points.

While policymakers aim at moderating the volume and/or volatility of certain types of capital flows with the help of these controls, evidence presented in this section suggests that their macroeconomic impact has been at best. Moreover, evidence indicates that they have been associated with substantial microeconomic costs.

9. HOW TO APPROACH FINANCIAL INTEGRATION?

A number of papers have attempted to reconcile the disparity between theory and empirical evidence on the benefits of financial globalization by suggesting that the costs—including crises—are in the nature of growing pains that will recede once globalizing economies achieve fuller integration.⁵⁷ This finding partly lines up with the results about financial integration generating collateral benefits and thereby eventually

having a positive impact on economic growth. Similarly, Martinez, Tornell, and Westermann (2004) argue that crises are the price that must be paid to attain rapid growth in the presence of contract enforceability problems. These authors present some evidence that developing economies that have registered higher growth rates have typically experienced boom-bust cycles (also see Bussiere & Fratzscher, 2004; Kaminsky & Schmukler, 2003).

These papers reflect the notion that financial globalization carries a short-run cost—one that must inevitably be paid if a developing country, which typically has weak institutions and a fragile financial sector, wants to move on to a high-growth path. And, putting this together with the literature that finds that financial globalization could serve as a useful catalyst for improving domestic institutions and financial markets, it appears that developing countries face a Hobson's choice. Globalize and improve growth prospects at a cost of vulnerability to painful crises. Or not globalize and bear the cost of being stuck in a low-growth environment. Is there a way out?⁵⁸

The reality is that developing economies may ultimately have little choice but to accept financial globalization since staying closed could become increasingly costly in terms of foregone long-term economic welfare, both in absolute and relative terms. There are some approaches that countries could adopt to try and acquire some of the benefits of globalization, say by opening up to trade flows and liberalizing trade-related capital account transactions. Trade integration may deliver some of the benefits of globalization, including preparing the ground for financial integration. But this has its limits since trade integration creates channels for *de facto* financial integration.

The collateral benefits perspective proposed in this chapter does suggest a way for moving forward on capital account liberalization, if policymakers in a given country feel that it could be beneficial but are concerned about the attendant risks. If one can identify which reform priorities are the key ones for a particular country, then one can design an approach to capital account liberalization that could generate specific benefits while minimizing the associated risks. For instance, Prasad and Rajan (2008) propose an approach to control the capital account liberalization for economies trying to manage their exchange rates while experiencing large inflows. Their approach, which would essentially involve securitizing inflows via private mutual funds that would invest abroad, would generate benefits such as development of domestic financial markets (through the issuance and trading of securities) and would also give domestic agents access to international portfolio diversification opportunities. But the outflows would be controlled both in terms of quantity and timing, thereby reducing the risks. This could mitigate the problem noted by Bhagwati (1998) that, once opened, capital accounts can subsequently be difficult to close even if circumstances should warrant it.⁵⁹

10. CONCLUSION

Our synthesis of the literature on financial globalization and economic policies points to some major complications during the transition from low to high levels of financial integration. Financial globalization can, in principle, play a catalytic role in generating an array of collateral benefits that boost long-run growth and welfare. These collateral benefits could include development of the domestic financial sector, improvement in institutions, and better macroeconomic policies. By contrast, the macroeconomic evidence on the benefits and effectiveness of capital controls is at best mixed while some recent studies based on firm-level data indicate that controls appear to lead to various costs at the micro level. The implications of these results for policies toward capital account liberalization are complicated by the existence of threshold conditions. Full-fledged opening of the capital account in the absence of essential supporting conditions can vitiate the realization of any benefits, while making a country more vulnerable to sudden stops of capital flows. Economic policies designed to foster these necessary supporting conditions, while beneficial in their own right, could also be instrumental in more effectively utilizing the growth and stability gains stemming from financial integration.

We obviously do not have a silver bullet to offer based on our reading of the literature as our analysis suggests that the relationship between financial integration and economic policies is a complex one and that there are inescapable tensions inherent in evaluating the risks and benefits associated with financial globalization. Although there is evidence in support of our broad conclusions, even these often need to be tailored to take into account country-specific circumstances in light of these tensions. Nevertheless, it is essential to see financial integration not just as an isolated policy goal but also as part of a broader package of reforms and supportive macroeconomic policies.

It is becoming increasingly more sensible for developing countries to shift their focus to how they will manage the process of financial liberalization rather than whether they should liberalize at all (see Prasad, 2008 for a discussion in the context of India). There are at least a couple of compelling reasons for this: First, capital accounts will become more open so long as there are strong incentives for cross-border flows of capital. Increasing global financial flows will inevitably result in *de facto* opening of the capital account, irrespective of the capital control regime. Hence, it may be best for policymakers in emerging market economies to take steps to actively manage the process of financial integration—rather than just try to delay or push back against the inevitable—in order to improve the benefit-cost tradeoff. Otherwise, policymakers may be stuck with the worst of all possible worlds—the distortions created by *de jure* capital controls and the complications of domestic macroeconomic management that are a consequence of increasing cross-border flows.

Second, given the balance of risks will vary over time, the global economic environment and the circumstances of individual countries may create windows of opportunity

for countries to pursue financial integration. For instance, private capital flows in the last few years are increasingly taking the form of FDI or portfolio equity flows, both of which are less volatile and more beneficial than portfolio debt flows. A number of emerging market economies have accumulated large stocks of foreign exchange reserves, and have also become more open to trade, which has substantially reduced the risks related to sudden stops or reversals of capital inflows, and also mitigated risks of contagion. A country that has shifted the terms of the debate to “how” from “whether” can take advantage of these windows of opportunity to press for further liberalization.

Our findings suggest that, in exploring appropriate policies with respect to the liberalization of financial account, policy makers need to consider not only the relevant threshold conditions, but also the collateral benefits associated with integration. In light of this conclusion, how can policymakers approach the process of capital account liberalization? The collateral benefits perspective suggests a way for moving forward on capital account liberalization. If one can identify which reform priorities are the key ones for a particular country, then one can design a controlled approach to liberalization that could generate specific benefits while minimizing the associated risks. This is preferable to a complete opening of the capital account, which could also deliver that benefit, but at the cost of the possible worsening of the benefit-risk tradeoff of financial opening. This approach also provides a broader analytical framework within which one can incorporate country-specific features and initial conditions into the design of appropriate capital account liberalization programs.

Nonetheless, the design of an analytical framework is easy, or the risks of financial integration are small, and that countries should rush headlong into it. Indeed, one of the main lessons of recent episodes of capital account liberalization is that, once the taps are opened to capital flows, it can be very difficult to shut them. Moreover, allowing financial integration to get too far ahead of other policy reforms—especially, the domestic financial sector reforms and greater exchange rate flexibility—could have potentially devastating consequences if there were to be sudden shifts in international investor sentiment. There are also substantial inefficiencies in international financial markets, which remain far from complete in terms of the range of available instruments for sharing risk and are still beset by informational asymmetries, herding behavior, and other such pathologies.⁶⁰

Where can research help sharpen such policy conclusions? First, it is imperative to extend the research program on measuring financial openness. Although it is clear that different countries have adopted widely differing approaches to financial globalization, existing measures of cross-country differences are so crude that it is difficult to evaluate the macroeconomic outcomes of various policies in many cases.

Future research should focus on the indirect benefits of financial globalization that ultimately express themselves in productivity growth and macroeconomic stability. Early research that emphasized how financial globalization can help enhance physical capital accumulation in developing countries was clearly misplaced. The links between certain

aspects of open capital accounts (e.g., unrestricted foreign bank entry) and domestic financial sector development have been analyzed extensively, but evidence on other indirect benefits is limited. Future studies on these issues would help understand the quantitative impact of various indirect channels associated with financial globalization and can potentially lead to a better framework in evaluating alternative policies.

Another promising research avenue is a more detailed analysis of threshold effects—especially, the relative importance of different threshold conditions and the tradeoffs among them for a country that wishes to open up its capital account. How to balance the risks stemming from the absence of certain supporting conditions against the collateral benefits to be gained from financial integration is a pressing policy question. Future research needs to focus on the design of a unified framework for analyzing a variety of thresholds and their implications for the process of financial integration.

Our findings also suggest that it is difficult to make strong statements about the potential role of economic policies in shaping the growth and stability outcomes of financial integration using macroeconomic data. Further research based on industry- and firm-level data as well as event and case studies may provide more informative insights about the design of economic policies that could improve the benefits of financial integration.

End Notes

- *. This chapter is prepared for the *Handbook of Development Economics*, edited by Dani Rodrik and Mark R. Rosenzweig. We thank Stijn Claessens, Marco Terrones, and our discussant, Dani Rodrik, for their detailed suggestions. We also thank conference participants for providing helpful comments at the Handbook Preconference. Dionysis Kaltis and Yusuke Tateno provided excellent research assistance. The views expressed in this chapter are solely those of the authors and do not necessarily reflect those of the IMF or IMF policy.
- 1. For some other recent surveys on financial globalization, see Eichengreen (2001), Prasad et al. (2003), Kose et al. (2006), Mishkin (2006), Henry (2007), and Obstfeld (2007).
- 2. Financial globalization refers to rising global linkages through cross-border financial flows. Financial integration refers to an individual country's linkages to international capital markets. In this chapter, we use these terms interchangeably.
- 3. Share measures have been created by Grilli and Milesi-Ferretti (1995), Rodrik (1998), and Klein and Olivei (2006). Finer measures of openness based on the AREAER have been developed by Quinn (1997, 2003), Miniane (2004), Chinn and Ito (2006), Mody and Murshid (2005), and Edwards (2005). Edison and Warnock (2003) constructed measures of capital account restrictions related to just equity flows. Some recent studies consider more disaggregated measures based on the AREAER descriptions (see Schindler, 2008). Bekaert and Harvey (2000) and Henry (2000a) compile dates of equity market liberalizations for developing countries.
- 4. Edwards (2005) notes that binary measures suggest similar levels of capital account restrictions in Chile, Mexico, and Brazil during the period 1992–1994. In fact, Mexico had a rather open capital account, Brazil employed a complex set of controls on capital flows, and there were some controls on short-term flows in Chile.

5. A further complication is that, despite the extensive coverage of the IMF's annual AREAER publication, there could be other regulations that effectively act as capital controls but are not counted as controls. For an extensive discussion of these issues, see Kose et al. (2006).
6. Other quantity-based measures of integration include price-based measures of asset market integration and saving-investment correlations (see Karolyi & Stulz, 2003; Obstfeld & Taylor, 2004). Related to the price-based approach, researchers also employ various interest parity conditions (see Edison et al., 2002; Frankel, 1992). There are, however, serious problems in using these measures as they are difficult to operationalize and interpret for an extended period of time and for a large group of countries.
7. The *de facto* measures of financial integration that we use here draw upon the pioneering work of Lane and Milesi-Ferretti (2006), who have constructed an extensive dataset of gross liabilities and assets for 145 countries covering the period 1970–2004.
8. A much earlier wave of financial globalization, which took place between 1880 and 1914, has been analyzed by Bordo, Taylor, and Williamson (2003), Obstfeld and Taylor (2004), and Mauro, Sussman, and Yafeh (2006).
9. While debt financing remains the most important source of inflows for advanced economies, FDI now accounts for almost half of the total inflow into developing economies. Equity flows have become quite important for emerging markets, accounting for almost 12% of inflows, while this category still remains virtually nonexistent for other developing economies, reflecting their underdeveloped stock markets.
10. Akin and Kose (2008) document a variety of stylized facts about the evolution of trade and financial linkages, their underlying determinants and the changing nature of growth dynamics around the world.
11. Henry (2007) argues that, even in the context of the basic neoclassical model, the financing channel should imply only a temporary, rather than permanent, pickup in growth from financial integration. It is not clear, however, how important this nuance is likely to be empirically in studies that look at growth experiences over periods of just 2–3 decades.
12. The rapid growth in emerging economies over the past two decades have led to speculations about a possible shift in the center of global growth fueling questions about the implications of increased international linkages for the North–South growth dynamics. Akin and Kose (2008) examine these issues and provide empirical evidence about the changing nature of growth linkages across the Northern and Southern economies.
13. Endogeneity between financial integration and growth remains a potentially problematic issue in studies that find a positive association between these variables. Some authors have attempted to deal with this problem by using lagged measures of financial integration and GMM techniques in panel regressions. However, this problem may ultimately be intractable in macroeconomic data; looking at more disaggregated data may be one way out.
14. For example, in a much-cited study, Rodrik (1998) finds that capital account liberalization has no significant effect on economic growth. His analysis is based on a very coarse binary measure of capital controls. Employing a finer and more informative version of the same *de jure* openness measure, Quinn and Toyoda (2006) document a positive association between capital account liberalization and economic growth.
15. Bekaert, Harvey, and Lundblad (2005) conclude that equity market liberalizations increase long-term GDP growth by about 1%, a remarkably strong effect. There is also evidence, consistent with the predictions of international asset pricing models, that stock market liberalizations reduce the cost of capital and boost investment growth. For evidence on the first point, see Stulz (1999a, 1999b), Bekaert and Harvey (2000), Henry (2000a), and Kim and Singal (2000). On the latter, see Henry (2000b) and Alfaro and Hammel (2007).
16. Henry (2007) provides a detailed survey of this literature.

17. For a survey of the empirical literature on the risks associated with short-term debt, see Berg, Borenstein and Pattillo (2004).
18. See Claessens, Demirguc-Kunt, and Huizinga (2001), Levine (2001), Claessens and Laeven (2004), Clarke, Cull, Martinez Peria, and Sanchez (2003), Goldberg (2008), and Schmukler (2004). However, see Levy Yeyati and Micco (2007) for a contrarian view based on a sample of Latin American economies.
19. Baltagi, Demetriades, and Law (2007) documents that financial openness has a positive impact on the development of local financial markets in countries with low levels of trade openness. There is also some evidence that increased usage of international equity markets may have potentially negative spill-over effects for other domestic firms in terms of stock turnover (Levine & Schmukler, 2006).
20. Poor public governance (as measured by severity of bureaucratic corruption or lack of government transparency) discourages inward FDI and portfolio equity inflows (Doidge, Karolyi, & Stulz, 2005; Gelos & Wei, 2005; Wei, 2001).
21. Recent literature has emphasized the importance of TFP growth as the main driver of long-term GDP growth (see, e.g., Gourinchas & Jeanne, 2006; Hall & Jones, 1999; Jones & Olken, 2005).
22. See Rodrik and Subramanian (2008) for a skeptical view on the literature on indirect benefits.
23. In particular, the welfare gains depend on the volatility of output shocks, the rate of relative risk aversion, the risk-adjusted growth rate and the risk free interest rate in these models (see the discussion in Lewis, 1999; Obstfeld & Rogoff, 2004, chapter 5; van Wincoop, 1999). Recent research convincingly shows that the higher volatility that developing countries experience implies that they can potentially reap large benefits from international risk-sharing arrangements (see Pallage & Robe, 2003).
24. See Razin and Rose (1994), Easterly, Islam, and Stiglitz (2001), and Buch, Dopke, and Pierdzioch (2005).
25. Bekaert, Harvey, and Lundblad (2006) find that, following equity market liberalizations, there is a decline in consumption volatility. These results differ from those of Kose et al. (2003b) due to differences in the definitions of financial integration, the measures of consumption volatility, data samples, and methodologies. The results of Bekaert, Harvey, and Lundblad (2006) suffer from the same problems noted about their work on the impact of equity market liberalizations on economic growth.
26. A number of papers show that the synchronicity of national business cycle fluctuations (in both industrial countries and emerging market economies), and the relative importance of global (and/or) regional factors for these fluctuations has increased during the period of globalization (see Kose, Otrok, & Prasad, 2008; Kose, Otrok, & Whiteman, 2008). Imbs (2006) documents that financial integration has led to higher cross-country consumption and output correlations among industrialized economies. Kose et al. (2003a) document changes in output co-movement across a broader group of industrial and developing economies and link these changes to financial integration. Contrary to the predictions of theory, they document that, on average, cross-country correlations of consumption growth did not increase in the 1990s, precisely when financial integration would have been expected to result in better risk-sharing opportunities for developing economies.
27. These authors use a binary capital account openness indicator based on the IMF's AREAER. Whether this relationship holds up with de facto measures remains to be seen.
28. On the output costs of banking crises, see Hutchison and Noy (2005) and Bonfiglioli and Mendicino (2004).
29. The evidence cited on this point by some prominent critics of financial globalization in fact turns out to be about how domestic financial sector liberalization, rather than financial integration, has in some cases precipitated financial crises (see end note 5 in Stiglitz, 2004).
30. A key empirical issue is about the definition of thresholds. Kose et al. (2008) provide a comprehensive analysis of threshold effects in the process of financial integration. They employ three different

approaches: (i) A linear interaction between financial openness and the threshold variable. This approach just tests for whether the level of a particular variable affects the marginal effect of financial openness on growth; (ii) A quadratic interaction that allows for nonlinearity in the effect of the threshold variable; and (iii) A high-low cutoff based on the sample median of a threshold variable. This exogenously sets the threshold but it does provide a simple way of testing if the level of a particular variable matters in terms of the quantitative effect of openness on growth outcomes.

31. For instance, Levchenko (2005) and Leblebicioglu (2006) consider dynamic general equilibrium models where only some agents have access to international financial markets. In both models, capital account liberalization leads to an increase in the volatility of aggregate consumption since agents with access to international financial markets stop participating in risk-sharing arrangements with those who do not have such access.
32. Using broader measures of financial openness, Prasad, Rajan, and Subramanian (2007) find evidence of high/low interaction effects among nonindustrial countries while Kraay (1998) and Arteta et al. (2003) find little evidence of interaction effects.
33. See Easterly, Islam, and Stiglitz (2001), Beck, Lundberg and Majnoni (2001), Denizer, Iyigun, and Owen (2002), and Larrain (2004).
34. IMF (2007) provides empirical evidence about the conditioning role of institutional quality in governing the relationship between financial integration and consumption volatility. In particular, the study reports that financial integration is often associated with higher consumption volatility in countries with relatively poor institutional structures.
35. Hausmann and Fernández-Arias (2000) provide a contrarian view on the determinants and implications of the composition of flows to developing countries. Albuquerque (2003) argues that financially constrained countries are likely to get more FDI than other types of flows since it is harder to expropriate—not because it is more productive or intrinsically less volatile. Ju and Wei (2006) provide a framework to reconcile the results of these two papers with those of other authors arguing that it is crucial to distinguish between property rights institutions and financial institutions.
36. Their measure of institutional quality is an average of six indicators—voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption—from Kaufmann, Kraay, and Mastruzzi (2003). Faria and Mauro instrument the institutional index using settler mortality (Acemoglu, Johnson, and Robinson, 2001) and ethno-linguistic fragmentation. The IV approach reaffirms their basic conclusion.
37. See Calvo et al. (2004) and Frankel and Cavallo (2004).
38. Obstfeld and Rogoff (2001) emphasize the importance of frictions related to trade costs (broadly defined) for explaining a number of puzzles in international macroeconomics.
39. The studies cited by Rodrik are Haddad and Harrison (1993) and Aitken and Harrison (1999). These two papers examine the role of horizontal spillovers—productivity spillovers from foreign firms to domestic firms in the same sector—in transmitting the productivity benefits of FDI to Morocco and Venezuela. Lipsey (2004) also notes that Morocco and Venezuela were relatively closed to trade during the periods covered by the panel datasets used in these studies.
40. Kose, Meredith, and Towe (2005) argue that trade integration has made the Mexican economy more resilient to shocks and contributed to its faster recovery from the 1994–1995 peso crisis than from the 1982 debt crisis.
41. See Edwards (2004, 2005), Desai and Mitra (2004), and Guidotti, Sturzenegger, and Villar (2004).
42. Ramey and Ramey (1995), Aghion and Banerjee (2005), and Aizenman and Pinto (2006) document the negative relationship between growth and volatility.
43. While FDI flows help dampen the adverse impact of volatility on economic growth, other types of flows do not appear to have a significant effect on the relationship between volatility and growth (Kose, Prasad, et al., 2005).

44. See Berg, Borenstein, and Patillo (2004). Eichengreen, Rose, and Wyplosz (1995) show that these issues are relevant for more advanced economies as well. Using quarterly panel data for 20 OECD countries over the period 1959–1993, they document that high money and credit growth as well as large deficits in current account and fiscal positions tend to raise the probability of devaluations.
45. There are, of course, various ways to limit fiscal excess in developing countries. These include legislation of fiscal limits and greater transparency of public accounts (see Obstfeld, 2007).
46. A number of studies have examined the policy responses to capital inflows, focusing mainly on the experience of a few countries during the 1990s (see Calvo et al., 1994; Edwards, 2000; Fernández-Arias and Montiel, 1996; Glick, 1998; Montiel, 1999; Reinhart and Reinhart, 1998).
47. Their findings provide helpful guidance on what has worked, and not worked, in the past. The appropriate policy response to large capital inflows depends on a variety of country-specific circumstances, including the nature of the underlying inflows, the stage of the business cycle, the fiscal policy situation, and the quality of domestic financial markets also matters.
48. These authors provide an extensive survey of the effects of monetary and exchange rate policies on economic growth. For analyses of the impact of financial globalization on monetary policy, see also Obstfeld (2007), Rogoff (2007), and Woodford (2007).
49. These authors also find that banking crises are more likely under rigid exchange rate regimes. They note that this result contradicts that of Ghosh, Gulde, and Wolf (2003) and trace the differences to the latter authors' use of a *de jure* exchange rate regime classification.
50. Lucas himself offered a new growth model based on increasing returns to human capital to explain what was then a low volume of net flow to developing countries, though recent work has tended to focus more on the financial channel emphasized contemporaneously by Gertler and Rogoff (1990). Caballero, Farhi, and Gourinchas (2006), Mendoza et al. (2007), and Alfaro, Kalemli-Ozcan, and Volosovych (2007) argue that institutional failures more generally may lead to capital flow reversals. Reinhart and Rogoff (2004) suggest that recurrent defaults and financial crises in developing countries may depress the investment. Gordon and Bovenberg (1996) focus on the role played by information asymmetries.
51. Gourinchas and Jeanne (2006) document that net capital flows from rich to poor countries often end up in countries with relatively low productivity. They argue that this constitutes a major challenge for the standard models often employed to evaluate the implications of financial integration.
52. Prasad et al. (2007) estimate cross-country growth regressions with separate slopes on the exchange rate valuation variable for countries with overvalued and undervalued exchange rates. They find weak evidence of the asymmetric effect described in the text.
53. Magud and Reinhart (2007) argue that the literature analyzing the macroeconomic implications of capital controls using aggregate data has only limited value added since the studies in this literature suffer from various problems, including the use of heterogeneous samples, differences in methodologies, and multiple definitions of outcomes associated with the success of controls. Based on a study of the Malaysian experience, Abdelal and Alfaro (2003) argue that countries can rarely control international capital flows by imposing controls.
54. Moreover, stricter controls on outflows appeared to reduce net capital flows and allow more independent monetary policy in Malaysia after 1998, but there is little support for such outcomes in other countries (Magud & Reinhart, 2007). Kaplan and Rodrik (2001) provide evidence in support of the benefits of capital controls for monetary policy independence in Korea, Thailand, Indonesia and Malaysia.
55. Another policy used by some countries to cope with large net inflows was the removal of controls on outflows. Evidence based on the wave of inflows during the 1990s suggests that elimination of controls on outflows has often led to larger inflows. In a recent paper, Aizenman and Noy (2008) report that controls on capital account transactions have no impact on the volume of flows. Liberalizing

- outflow restrictions may attract heavier inflows by sending a positive signal to markets, increasing the investor confidence, and thereby fuelling even larger inflows (Bartolini & Drazen, 1997), which is supported by evidence based on several countries (Reinhart & Reinhart, 1998).
56. In particular, investors evaded them by purchasing Argentine stocks for pesos, converting the stocks into ADRs, and then selling the ADRs in New York for dollars that could be deposited in the US bank accounts.
 57. For instance, Krugman (2002) has argued that “...growing integration does predispose the world economy toward more crises, mainly because it creates pressures on governments to relax financial restrictions that in earlier decades made 1990s-style financial crises much less likely.” He goes on to say that “In the long run, integration may solve the problems it initially creates.”
 58. Rodrik (2007a) emphasizes the importance of creating an efficient “policy space” that could help address the risks associated with trade and financial integration. Hausmann, Rodrik, and Velasco (2008) argue that it is critical to identify the major constraint(s) that hamper economic growth and then propose solutions to eliminate them. Based on this idea, they develop a methodology of “growth diagnostics.” In light of the results of these diagnostics, Rodrik (2004) concludes that industrial policies, which are partly shaped by public institutions, play an important role in promoting economic development. Aghion and Durlauf (2007) criticize this approach and present an alternative one based on standard growth regressions. Zettelmeyer (2006) provides a brief evaluation of the growth diagnostics approach in the context of the Latin American experience.
 59. Some recent papers provide contradictory conclusions about the nature of policy responses to financial integration. For example, Rodrik and Subramanian (2008) conclude that the benefits of financial globalization are hard to document and it would be useful to consider policies to restrict capital inflows, if country-specific conditions deserve such a response. In contrast, Mishkin (2008) argues that, in order to attain better standards of living, emerging market economies need to become more integrated with the global financial system while employing policies that can improve their institutional frameworks to facilitate the growth and stability enhancing effects of international financial flows.
 60. For instance, as we discussed, efficient international risk sharing could yield enormous welfare gains, especially for developing countries, which have more volatile GDP growth rates than industrial countries. But, the degree of international risk sharing remains limited, mainly for want of suitable instruments. Despite the allure of GDP-indexed bonds as a device for sharing risk among countries, their use remains very limited. Griffith-Jones and Sharma (2006) discuss the potential benefits of such instruments and the factors that have limited their prevalence so far.

References

- Abdelal, R., & Alfarol, L. (2003). Capital and control: Lessons from Malaysia. *Challenge*, 46(4), 36–53.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401.
- Acemoglu, D., Johnson, S., Robinson, J. A., & Thaicharoen, Y. (2003). Institutional causes, macroeconomic symptoms: Volatility, crises and growth. *Journal of Monetary Economics*, 50(1), 49–123.
- Aghion, P., Bacchetta, P., & Banerjee, A. (2004). Financial development and the instability of open economies. *Journal of Monetary Economics*, 51(6), 1077–1106.
- Aghion, P., & Banerjee, A. (2005). *Volatility and growth: Clarendon lectures in economics*. New York: Oxford University Press.
- Aghion, P., & Durlauf, S. (2007). *From growth theory to policy design*. Working Paper. Harvard University.
- Aitken, B. J., & Harrison, A. E. (1999). Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *American Economic Review*, 89(3), 605–618.
- Aizenman, J., & Pinto, B. (Eds.), (2006). *Managing economic volatility and crisis*. New York: Cambridge University Press.

- Aizenman, J., & Noy, I. (2008). Links between trade and finance—A disaggregated analysis. In: S. Edwards & M. G. P. Garcia (Eds.), *Financial markets volatility and performance in emerging markets*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Akin, C., & Kose, M. A. (2008). Changing nature of north-south linkages: Stylized facts and explanations. *Journal of Asian Economics*, 19(1), 1–28.
- Albuquerque, R. (2003). The composition of international capital flows: Risk sharing through foreign direct investment. *Journal of International Economics*, 61(2), 353–383.
- Alfaro, L., Chanda, A., Kalemli-Ozcan, S., & Sayek, S. (2004). FDI and economic growth: The role of local financial markets. *Journal of International Economics*, 64(1), 89–112.
- Alfaro, L., & Hammel, E. (2007). Capital flows and capital goods. *Journal of International Economics*, 72(1), 128–150.
- Alfaro, L., Kalemli-Ozcan, S., & Volosovych, V. (2007). Capital flows in a globalized world: The role of policies and institutions. In: S. Edwards (Ed.), *Capital controls and capital flows in emerging economies: Policies, practices, and consequences*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Aoki, K., Benigno, G., & Kiyotaki, N. (2007). *Adjusting to capital account liberalization*. Working Paper. London School of Economics.
- Ariyoshi, A., Habermeier, K., Laurens, B., Otker-Robe, I., Canales-Kriljenko, J. I., & Kirilenko, A. (2000). Capital controls: Country experiences with their use and liberalization. IMF Occasional Paper 190. Washington: International Monetary Fund.
- Arteta, C., Eichengreen, B., & Wyplosz, C. (2003). When does capital account liberalization help more than it hurts? In E. Helpman & E. Sadka (Eds.), *Economic policy in the international economy: Essays in honor of Assaf Razin*. Cambridge, UK: Cambridge University Press.
- Balasubramanyam, V. N., Salisu, M., & Sapsford, D. (1996). Foreign direct investment and growth in EP and IS countries. *Economic Journal*, 106(434), 92–105.
- Baltagi, B., Demetriades, P., & Law, S. H. (2007). *Financial development, openness and institutions: Evidence from panel data*. Working Paper. Syracuse University.
- Bartolini, L., & Drazen, A. (1997). Capital-account liberalization as a signal. *American Economic Review*, 87(1), 138–154.
- Beck, T., Lundberg, M., & Majnoni, G. (2001). Financial intermediary development and growth volatility: Do intermediaries dampen or magnify shocks? *Journal of International Money and Finance*, 25(7), 1146–1167.
- Bekaert, G., & Harvey, C. R. (2000). Foreign speculators and emerging equity markets. *Journal of Finance*, 55(2), 565–613.
- Bekaert, G., Harvey, C. R., & Lundblad, C. (2005). Does financial liberalization spur growth? *Journal of Financial Economics*, 77(1), 3–55.
- Bekaert, G., Harvey, C. R., & Lundblad, C. (2006). Growth volatility and equity market liberalization. *Journal of International Money and Finance*, 25(3), 370–403.
- Berg, A., Borensztein, E., & Pattillo, C. (2004). Assessing early warning systems: How have they worked in practice? IMF Working Paper 04/52. Washington: International Monetary Fund.
- Bhagwati, J. (1998). The capital myth. The difference between trade in widgets and dollars. *Foreign Affairs*, 77(3), 7–12.
- Bonfiglioli, A. (2009). Financial integration, productivity and capital accumulation. *Journal of International Economics*, Elsevier, 86(2), 337–355.
- Bonfiglioli, A., & Mendicino, C. (2004). Financial liberalization, banking crises, and growth: Assessing the links. Working Paper Series in Economics and Finance 567. Stockholm School of Economics.
- Bordo, M., Taylor, A. M., & Williamson, J. G. (2003). *Globalization in historical perspective*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Brecher, R. A., & Diaz-Alejandro, C. F. (1977). Tariffs, foreign capital, and immiserizing growth. *Journal of International Economics*, 7, 317–322.
- Buch, C. M., Döpke, J., & Pierdzioch, C. (2005). Financial openness and business cycle volatility. *Journal of International Money and Finance*, 24, 744–765.

- Bussiere, M., & Fratzscher, M. (2004). Financial openness and growth: Short-run gain, long-run pain? ECB Working Paper No. 348.
- Caballero, R., Farhi, E., & Gourinchas, P. O. (2006). *An equilibrium model of global imbalances and low interest rates*. NBER Working Papers 11996. Cambridge, MA: National Bureau of Economic Research.
- Caballero, R. J., & Krishnamurthy, A. (2001). International and domestic collateral constraints in a model of emerging market crises. *Journal of Monetary Economics*, 48(3), 513–548.
- Calvo, G., Izquierdo, A., & Mejía, L. F. (2004, June). On the empirics of sudden stops: The relevance of balance-sheet effects. In Federal Reserve Bank of San Francisco Proceedings.
- Calvo, G., Leiderman, L., & Reinhart, C. M. (1994). The capital inflows problem: Concepts and issues. *Contemporary Economic Policy*, 12(3), 54–66.
- Calvo, G., & Talvi, E. (2005). Sudden stop, financial factors, and economic collapse in Latin America: Learning from Argentina and Chile. NBER Working Papers 11153. Cambridge, MA: National Bureau of Economic Research.
- Carderelli, R., Kose, M. A., & Elekdag, S. (2009). Capital Inflows: Macroeconomic Implications and Policy Responses. *IMF Working Paper*, WP/09/40.
- Carkovic, M., & Levine, R. (2005). Does foreign direct investment accelerate economic growth? In T. H. Moran, E. M. Graham & M. Blomström (Eds.), *Does foreign investment promote development?* (pp. 195–220). Washington: Institute for International Economics.
- Cavallo, E. A. (2007). *Output volatility and openness to trade: A reassessment*. RES Working Papers 4518. Washington: Inter-American Development Bank, Research Department.
- Chanda, A. (2005). The influence of capital controls on long run growth: Where and how much? *Journal of Development Economics*, 77, 441–466.
- Chari, A., & Henry, P. B. (2004). Risk sharing and asset prices: Evidence from a natural experiment. *Journal of Finance*, 59(3), 1295–1324.
- Chari, A., & Henry, P. B. (2008). Firm-specific information and the efficiency of investment. *Journal of Financial Economics*, 87(3), 636–655.
- Chinn, M., & Ito, H. (2006). What matters for financial development? Capital controls, institutions and interactions. *Journal of Development Economics*, 81(1), 163–192.
- Claessens, S., Demirgüç-Kunt, A., & Huizinga, H. (2001). How does foreign entry affect domestic banking markets? *Journal of Banking and Finance*, 25(5), 891–911.
- Claessens, S., & Laeven, L. (2004). What drives bank competition? Some international evidence. *Journal of Money, Credit, and Banking*, 36(3), 563–583.
- Clarke, G., Cull, R., Martinez Peria, M. S., & Sanchez, S. M. (2003). Foreign bank entry: Experience, implications for developing economies, and agenda for further research. *World Bank Research Observer*, 18(1), 25–59.
- Collins, S. M. (2007). Comments on Financial globalization, growth, and volatility in developing countries by Eswar Prasad, Kenneth Rogoff, Shang-Jin Wei, and M. Ayhan Kose. In: A. Harrison (Ed.), *Globalization and poverty*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Coricelli, F., Masten, A., and Masten, I. (2008) “Non-linear growth effects of financial development: Does financial integration matter?” *Journal of International Money and Finance*, 27(2008), 295–313.
- Cornelius, P.K., & Kogut, B. (Eds.), (2003). *Corporate governance and capital flows in a global economy*. New York: Oxford University Press.
- Denizer, C., Iyigun, M. F., & Owen, A. L. (2002). Finance and macroeconomic volatility. *Contributions to Macroeconomics*, 2(1), Article 7.
- Desai, M. A., Foley, C. F., & Hines, J. R., Jr. (2004). *Capital controls, liberalizations and foreign direct investment*. NBER Working Papers 10337. Cambridge, MA: National Bureau of Economic Research.
- Desai, P. (2003). *Financial crisis, contagion, and containment: From Asia to Argentina*. Princeton: Princeton University Press.
- Desai, P., & Mitra, P. (2004). Why do some countries recover more readily from financial crises? Paper presented at the *IMF Conference in honor of Guillermo A. Calvo*. Washington: International Monetary Fund.

- Dodge, C., Karolyi, A., & Stulz, R. (2005). Why do countries matter so much for corporate governance? *Journal of Financial Economics*, 86(1), 1–39.
- Durham, J. B. (2004). Absorptive capacity and the effects of foreign direct investment and equity foreign portfolio investment on economic growth. *European Economic Review*, 48(2), 285–306.
- Easterly, W., Islam, R., & Stiglitz, J. E. (2001). Shaken and stirred: Explaining growth volatility. In B. Pleskovic & N. Stern (Eds.), *Annual World Bank Conference on Development Economics*. Washington: World Bank.
- Edison, H. J., Klein, M., Ricci, L., & Sløk, T. (2004). *Capital account liberalization and economic performance: Survey and synthesis*. IMF Staff Papers, 51(2) Washington: International Monetary Fund.
- Edison, H. J., Levine, R., Ricci, L., & Sløk, T. (2002). International financial integration and economic growth. *Journal of International Money and Finance*, 21(6), 749–776.
- Edison, H. J., & Warnock, F. E. (2003). A simple measure of the intensity of capital controls. *Journal of Empirical Finance*, 10(1–2), 81–103.
- Edwards, S. (Ed.), (2000). *Capital flows and the emerging economies: Theory, evidence, and controversies*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Edwards, S. (Ed.), (2001). *Capital mobility and economic performance: Are emerging economies different?* NBER Working Papers 8076. Cambridge, MA: National Bureau of Economic Research.
- Edwards, S. (Ed.), (2004). *Financial openness, sudden stops, and current account reversals*. NBER Working Papers 10277. Cambridge, MA: National Bureau of Economic Research.
- Edwards, S. (Ed.), (2005). *Capital controls, sudden stops, and current account reversals*. NBER Working Papers 11170. Cambridge, MA: National Bureau of Economic Research.
- Edwards, S. (2008). Financial openness, currency crises and output losses. In: S. Edwards & M. G. P. Garcia (Eds.), *Financial markets volatility and performance in emerging markets*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Edwards, S., & Rigobon, R. (2005). *Capital controls, exchange rate volatility and external vulnerability*. NBER Working Papers 11434. Cambridge, MA: National Bureau of Economic Research.
- Eichengreen, B. J. (2000). Taming capital flows. *World Development*, 28(6), 1105–1116.
- Eichengreen, B. J. (2001). Capital account liberalization: What do cross-country studies tell us? *World Bank Economic Review*, (15), 341–365.
- Eichengreen, B. J., Hausmann, R., & Panizza, U. (2006). The pain of original sin. In B. Eichengreen & R. Hausmann (Eds.), *Other people's money*. Chicago: University of Chicago Press.
- Eichengreen, B., Rose, A., & Wyplosz, C. (1995). Exchange market mayhem: The antecedents and aftermath of speculative attacks. *Economic Policy*, 10(21), 249–312.
- Ezoenou, P. (2006). *Financial integration and macroeconomic volatility: Does financial development matter?* Working Paper. European University Institute.
- Faria, A., & Mauro, P. (2005). *Institutions and the external capital structure of countries*. IMF Working Paper 04/236. Washington: International Monetary Fund.
- Fernandez-Arias, E., & Montiel, P. J. (1996). The surge in capital inflows to developing countries: An overview. *World Bank Economic Review*, 10(1), 51–77.
- Fischer, S. (1998). Capital account liberalization and the role of the IMF. In *Should the IMF pursue capital-account convertibility? Princeton Essays in International Finance 207* (pp. 1–10). Princeton: Princeton University Press.
- Forbes, K. (2005a). The microeconomic evidence on capital controls: No free lunch. NBER Working Papers 11372. Cambridge, MA: National Bureau of Economic Research.
- Forbes, K. (2005b). Capital controls: Mud in the wheels of market efficiency. *Cato Journal*, 25(1), 153–166.
- Frankel, J. (1992). Measuring international capital mobility: A review. *American Economic Review*, 82(2), 197–202.
- Frankel, J., & Cavallo, E. A. (2004). *Does openness to trade make countries more vulnerable to sudden stops or less? Using gravity to establish causality*. NBER Working Papers 10957. Cambridge, MA: National Bureau of Economic Research.
- Frankel, J., & Rose, A. K. (1996). Currency crashes in emerging markets: An empirical treatment. *Journal of International Economics*, 41(3–4), 351–366.

- Frankel, J., & Wei, S. J. (2005). Managing macroeconomic crises: Policy lessons. In J. Aizenman & B. Pinto (Eds.), *Managing economic volatility and crises: A practitioner's guide*. New York: Cambridge University Press.
- Gelos, R. G., & Wei, S. J. (2005). Transparency and international portfolio holdings. *Journal of Finance*, 60(6), 2987–3020.
- Gertler, M., & Rogoff, K. (1990). North-South lending and endogenous domestic capital market inefficiencies. *Journal of Monetary Economics*, 26(2), 245–266.
- Ghosh, A., Gulde, A. M., & Wolf, H. (2003). *Exchange rate regimes: Choices and consequences*. Cambridge, MA: MIT Press.
- Glick, R. (Ed.), (1998). *Managing capital flows and exchange rates: Perspectives from the Pacific Basin*. Cambridge, MA: Cambridge University Press.
- Glick, R., Guo, X., & Hutchison, M. (2006). Currency crises, capital-account liberalization, and selection bias. *Review of Economics and Statistics*, 88(4), 698–714.
- Glick, R., & Hutchison, M. (2001). Banking and currency crises: How common are twins? In R. Glick, R. Moreno & M. M. Spiegel (Eds.), *Financial crises in emerging markets*. New York: Cambridge University Press.
- Goldberg, L. (2009). *Understanding banking sector globalization*. IMF Staff Papers, 56(1), 171–197.
- Gordon, R., & Bovenberg, A. L. (1996). Why is capital so immobile internationally? Possible explanations and implications for capital income taxation. *American Economic Review*, 86(5), 1057–1075.
- Gourinchas, P. O., & Jeanne, O. (2005). *Capital mobility and reform*. IMF Manuscript. Washington: International Monetary Fund.
- Gourinchas, P. O., & Jeanne, O. (2006). The elusive gains from international financial integration. *Review of Economic Studies*, 73(3), 715–741.
- Griffith-Jones, S., & Sharma, K. (2006). *GDP-indexed bonds: Making it happen*. DESA Working Paper 21. United Nations.
- Grilli, V., & Milesi-Ferretti, G. M. (1995). Economic effects and structural determinants of capital controls. *IMF Staff Papers*, 42(3), 517–551. (Washington: International Monetary Fund).
- Guidotti, P. E., Sturzenegger, F., & Villar, A. (2004). On the consequences of sudden stops. *Economía*, 4(2), 171–214.
- Gupta, N., & Yuan, K. (2008). On the growth effects of liberalizations. *Review of Financial Studies*, (forthcoming).
- Haber, S. (2002). “The Political Economy of Crony Capitalism.” In Stephen Haber ed., *Crony Capitalism and Economic Growth in Latin America: Theory and Evidence* (Hoover Institution Press, 2002), xi–xxi.
- Haddad, M., & Harrison, A. (1993). Are there positive spillovers from direct foreign investment? Evidence from panel data for Morocco. *Journal of Development Economics*, 42(1), 51–74.
- Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, 114(1), 83–116.
- Hammel, E. (2006). *Stock market liberalization and industry growth*. Manuscript. Harvard Business School.
- Hausmann, R., & Fernández-Arias, E. (2000). Foreign direct investment: Good cholesterol? IADB Working Paper 417. Washington: Inter-American Development Bank.
- Hausmann, R., Rodrik, D., & Velasco, A. (2008). Growth diagnostics. In J. Stiglitz & N. Serra (Eds.), *The Washington consensus reconsidered: Towards a new global governance*. New York: Oxford University Press.
- Henry, P. B. (2000a). Stock market liberalization, economic reform, and emerging market equity prices. *Journal of Finance*, 55(2), 529–564.
- Henry, P. B. (2000b). Do stock market liberalizations cause investment booms? *Journal of Financial Economics*, 58(1–2), 301–334.
- Henry, P. B. (2007). Capital account liberalization: Theory, evidence, and speculation. *Journal of Economic Literature*, 45(4), 887–935.
- Hermes, N., & Lensink, R. (2003). Foreign direct investment, financial development and economic growth. *Journal of Development Studies*, 40(1), 142–163.
- Hines, J. R., Jr. (1995). Forbidden payment: Foreign bribery and American business after 1977. NBER Working Papers 5266. Cambridge, MA: National Bureau of Economic Research.

- Husain, A., Mody, A., & Rogoff, K. (2005). Exchange rate regime durability and performance in developing versus advanced economies. *Journal of Monetary Economics*, 52(1), 35–64.
- Hutchinson, M., & Noy, I. (2005). How bad are the Twins? Output costs of currency and banking crises. *Journal of Money, Credit and Banking*, 37(4), 725–752.
- Imbs, J. (2006). The real effects of financial integration. *Journal of International Economics*, 68(2), 296–324.
- International Monetary Fund. (2007). Reaping the benefits of financial globalization. Policy Discussion Paper. Washington: International Monetary Fund.
- Ishii, S., Habermeier, K., Laurens, B., Leimone, J., Vadasz, J., & i Canales-Kriljenko, J. I. (2002). Capital account liberalization and financial sector stability. IMF Occasional Paper 211. Washington: International Monetary Fund.
- Javorcik, B. S. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review*, 94(3), 605–627.
- Johnson, S., & Mitton, T. (2003). Cronyism and capital controls: Evidence from Malaysia. *Journal of Financial Economics*, 67(2), 351–382.
- Jones, B. F., & Olken, B. A. (2005). Do leaders matter? National leadership and growth since World War II. *The Quarterly Journal of Economics*, 120(3), 835–864.
- Ju, J. M., & Wei, S. J. (2006). *A solution to two paradoxes of international capital flows*. NBER Working Papers 12668. Cambridge, MA: National Bureau of Economic Research.
- Kaminsky, G. L., Reinhart, C. M., & Vegh, C. A. (2004). *When it rains, it pours: Procylical capital flows and macroeconomic policies*. NBER Working Papers 10780. Cambridge, MA: National Bureau of Economic Research.
- Kaminsky, G. L., & Schmukler, S. (2003). *Short-run pain, long-run gain: The effects of financial liberalization*. NBER Working Papers 9787. Cambridge, MA: National Bureau of Economic Research.
- Kaplan, E., & Rodrik, D. (2001). *Did the Malaysian capital controls work?* NBER Working Papers 8142. Cambridge, MA: National Bureau of Economic Research.
- Karolyi, G. A., & Stulz, R. M. (2003). Are financial assets priced locally or globally? In G. M. Constantinides, M. Harris & R. M. Stulz (Eds.), *Handbook of the economics of finance, financial markets and asset pricing*. (Vol. 1B). Amsterdam: Elsevier.
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2003). Governance matters III: Governance indicators for 1996–2002. World Bank Policy Research Working Paper 3106. Washington: World Bank.
- Kim, E. H., & Singal, V. (2000). Stock market openings: Experience of emerging economies. *Journal of Business*, 73(1), 25–66.
- Klein, M. (2005). *Capital account liberalization, institutional quality and economic growth: Theory and evidence*. NBER Working Papers 11112. Cambridge, MA: National Bureau of Economic Research.
- Klein, M., & Olivei, G. (2006). *Capital account liberalization, financial depth, and economic growth*. Working Paper. Tufts University.
- Kose, M. A., Meredith, G., & Towe, C. (2005). How has NAFTA affected the Mexican economy? Review and evidence. In R. J. Langhammer & L. Vinhas de Souza (Eds.), *Monetary policy and macroeconomic stabilization in Latin America* (pp. 35–81). New York: Springer-Verlag.
- Kose, M. A., Otrok, C., & Prasad, E. S. (2008). Global business cycles: Convergence or decoupling? IMF Working Paper No. 08/143.
- Kose, M. A., Otrok, C., & Whiteman, C. (2008). Understanding the evolution of world business cycles. *Journal of International Economics*, 75, 110–130.
- Kose, M. A., Prasad, E. S., Rogoff, K., & Wei, S. J. (2009). Financial globalization: A reappraisal. IMF Working Paper 06/189. Washington: International Monetary Fund, *IMF Staff Papers*, 56(1), 8–62.
- Kose, M. A., Prasad, E. S., & Taylor, A. (2008). Thresholds in the process of international financial integration. Manuscript. IMF and Cornell University.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2003a). How does globalization affect the synchronization of business cycles? *American Economic Review*, 93(2), 57–63.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2003b). Financial integration and macroeconomic volatility. *IMF Staff Papers*, 50, 119–142.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2004). Volatility and comovement in a globalized world economy: An empirical exploration. In H. Siebert (Ed.), *Macroeconomic policies in the world economy* (pp. 89–122). New York: Springer-Verlag.

- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2005). Growth and volatility in an era of globalization. *IMF Staff Papers*, 52, 31–63. Washington: International Monetary Fund.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2006). How do trade and financial integration affect the relationship between growth and volatility? *Journal of International Economics*, 69, 176–202.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2009a). Does openness to international financial flows raise productivity growth? *Journal of International Money and Finance*, 28(4), 554–580.
- Kose, M. A., Prasad, E. S., & Terrones, M. E. (2009b). Does financial globalization promote risk sharing? *Journal of Development Economics*, 89(2), 258–270.
- Kraay, A. (1998). In search of the macroeconomic effects of capital account liberalization. Manuscript. World Bank (unpublished).
- Krueger, A. O. (2002). Why crony capitalism is bad for economic growth. In S. Haber (Ed.), *Crony capitalism and economic growth in Latin America: Theory and evidence*. Stanford: Hoover Institution Press.
- Krugman, P. (2002). Crises: The price of globalization? Paper presented at 2000 Symposium on Global Economic. Federal Reserve Bank of Kansas City.
- Lane, P. R., & Milesi-Ferretti, G. M. (2006). *The external wealth of nations mark II: Revised and extended estimates of foreign assets and liabilities, 1970–2004*. IMF Working Paper 06/69. Washington: International Monetary Fund.
- Larraín, B. (2004). *Financial development, financial constraints, and the volatility of industrial output*. Public Policy Discussion Paper 04–6. Federal Reserve Bank of Boston.
- La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, Vishny & Robert (2000). Investor protection and corporate governance. *Journal of Financial Economics*, 58, 3–27.
- Leblebicioglu, A. (2006). *Financial integration, credit market imperfections and consumption smoothing*. Working Paper. North Carolina State University.
- Levchenko, A. A. (2005). *Financial liberalization and consumption volatility in developing countries*. *IMF Staff Papers*, 52(2). Washington: International Monetary Fund.
- Levine, R. (2001). International financial integration and economic growth. *Review of International Economics*, 9(4), 684–698.
- Levine, R. (2005). Finance and growth: Theory and evidence. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth*. (Vol. 1A, pp. 865–934). Amsterdam: Elsevier.
- Levine, R., & Schmukler, S. (2006). Internationalization and stock market liquidity. *Review of Finance*, 10 (1), 153–187.
- Levine, R., & Zervos, S. (1998). Capital control liberalization and stock market development. *World Development*, 26(7), 1169–1183.
- Levy Yeyati, E., & Micco, A. (2007). Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk. *Journal of Banking and Finance*, 31(6), 1633–1647.
- Levy Yeyati, E., Sturzenegger, F. (2009). The effect of monetary and exchange rate policies on development. In D. Rodrik & M. Rosenzweig (Eds.), *Handbook of development economics*. Amsterdam: Elsevier Vol. 5.
- Lewis, K. K. (1999). Trying to explain home bias in equities and consumption. *Journal of Economic Literature*, 37(2), 571–608.
- Lipsey, R. E. (2004). Home and host country effects of FDI. In: R. E. Baldwin & L. A. Winters (Eds.), *Challenges to globalization: Analyzing the economics*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Lucas, R. E. Jr, (1990). Why doesn't capital flow from rich to poor countries. *American Economic Review*, 80(2), 92–96.
- Magud, N., & Reinhart, C. (2007). Capital controls: An evaluation. In: S. Edwards (Ed.), *Capital controls and capital flows in emerging economies: Policies, practices, and consequences*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Martinez, L., Tornell, A., & Westermann, F. (2004). *The positive link between financial liberalization growth and crises*. NBER Working Papers 10293. Cambridge, MA: National Bureau of Economic Research.
- Martin, P., & Rey, H. (2006). Globalization and emerging markets: With or without crash? *American Economics Review*, 96(5), 1631–1651.

- Mauro, P., Sussman, N., & Yafeh, Y. (2006). *Emerging markets, sovereign debt, and International Financial Integration: 1870–1913 and today*. New York: Oxford University Press.
- Mendoza, E., Quadrini, V., & Rios-Rull, J. V. (2007). *Financial integration, financial deepness and global imbalances*. NBER Working Papers 12909. Cambridge, MA: National Bureau of Economic Research.
- Miniane, J. (2004). A new set of measures on capital account restrictions. *IMF Staff Papers*, 51(2), 276–308.
- Mishkin, F. S. (2006). *The next great globalization: How disadvantaged nations can harness their financial systems to get rich*. Princeton: Princeton University Press.
- Mishkin, F. S. (2009). Why we shouldn't turn our backs on financial globalization. *IMF Staff Papers*, 56(1), 139–170.
- Mitton, T. (2006). Stock market liberalization and operating performances at the firm level. *Journal of Financial Economics*, 81(3), 625–647.
- Mody, A., & Murshid, A. P. (2005). Growing up with capital flows. *Journal of International Economics*, 65 (1), 249–266.
- Montiel, P. J. (1999). Policy Responses to Volatile Capital Flows, unpublished. Available via the Internet: www.worldbank.org/research/interest/confs/past/paps1516/montiel2.pdf.
- Moran, T. H. (2005). How does FDI affect host country development? Using industry case studies to make reliable generalizations. In T. H. Moran, E. M. Graham & M. Blomström (Eds.), *Does foreign investment promote development?*. Washington: Institute for International Economics.
- Neft, A., & Velasco, A. (2003). *Tough policies, incredible policies*. NBER Working Papers 9932. Cambridge, MA: National Bureau of Economic Research.
- Obstfeld, M. (2009). *International finance and growth in developing countries: What have we learned?* IMF Staff Papers, 56(1), 63–111.
- Obstfeld, M., & Rogoff, K. (1995). Exchange rate dynamics redux. *Journal of Political Economy*, 103(3), 624–660.
- Obstfeld, M., & Rogoff, K. (2001). The six major puzzles in international macroeconomics: Is there a common cause? In B. S. Bernanke & K. Rogoff (Eds.), *NBER macroeconomics annual 2000*. Cambridge, MA: MIT Press.
- Obstfeld, M., & Rogoff, K. (2004). *Foundations of international macroeconomics*. Cambridge, MA: MIT Press.
- Obstfeld, M., & Taylor, A. M. (2004). *Global capital markets: Integration, crisis, and growth*. Cambridge, UK: Cambridge University Press.
- Pallage, S., & Robe, M. A. (2003). On the welfare cost of economic fluctuations in developing countries. *International Economic Review*, 44(2), 677–698.
- Prasad, E. S. (2009). Some new perspectives on India's approach to capital account liberalization. *Brookings India Policy Forum*, Vol. 5 (2008–2009), July.
- Prasad, E. S., & Rajan, R. (2008). A pragmatic approach to capital account liberalization. *Journal of Economic Perspectives*, 22(3), 149–172.
- Prasad, E. S., Rajan, R., & Subramanian, A. (2007). Foreign capital and economic growth. *Brookings Papers on Economic Activity*, 2007(1), 153–230.
- Prasad, E. S., Rogoff, K., Wei, S. J., & Kose, M. A. (2003). *Effects of financial globalization on developing countries: Some empirical evidence*. IMF Occasional Paper 220. Washington: International Monetary Fund.
- Prasad, E. S., Rumbaugh, T., & Wang, Q. (2005). *Putting the cart before the horse? Capital account liberalization and exchange rate flexibility in China*. IMF Policy Discussion Paper 05/1. Washington: International Monetary Fund.
- Prasad, E. S., & Wei, S. J. (2007). The Chinese approach to capital inflows: Patterns and possible explanations. In: S. Edwards (Ed.), *Capital controls and capital flows in emerging economies: Policies, practices, and consequences*. National Bureau of Economic Research Conference Report, Chicago: University of Chicago Press.
- Quinn, D. (1997). The correlates of changes in international financial regulation. *American Political Science Review*, 91, 531–551.
- Quinn, D. (2003). Capital account liberalization and financial globalization, 1890–1999: A synoptic view. *International Journal of Finance and Economics*, 8(3), 189–204.
- Quinn, D., & Toyoda, A. M. (2006). *Does capital account liberalization lead to growth?*. Working Paper. Georgetown University.

- Rajan, R., & Zingales, L. (1998). Financial dependence and growth. *American Economic Review*, 88(3), 559–586.
- Ramey, G., & Ramey, V. A. (1995). Cross-country evidence on the link between volatility and growth. *American Economic Review*, 85(5), 1138–1151.
- Razin, A., & Rose, A. K. (1994). Business-cycle volatility and openness: An exploratory cross-sectional analysis. In L. Leiderman & A. Razin (Eds.), *Capital mobility: The impact on consumption, investment, and growth* (pp. 48–76). Cambridge, UK: Cambridge University Press.
- Reinhart, C. M., & Reinhart, V. (1998). Some lessons for policymakers who deal with the mixed blessing of capital inflows. In M. Kahler (Ed.), *Capital flows and financial crises* (pp. 93–127). Ithaca, NY: Cornell University Press.
- Reinhart, C. M., & Rogoff, K. (2004). The modern history of exchange rate arrangements: A reinterpretation. *The Quarterly Journal of Economics*, 119(1), 1–48.
- Reinhart, C. M., Rogoff, K., & Savastano, M. A. (2003). Debt intolerance. *Brookings Papers on Economic Activity*, 1, 1–74.
- Rodrik, D. (1998). Who needs capital-account convertibility? In *Should the IMF pursue capital-account convertibility?* Princeton Essays in International Finance 207 (pp. 55–65). Princeton: Princeton University Press.
- Rodrik, D. (1999). *The new global economy and developing countries: Making openness work*. Washington: Overseas Development Council.
- Rodrik, D. (2004). *Industrial policy for the twenty-first century*. Working Paper. Harvard University.
- Rodrik, D. (2007a). *How to save globalization from its cheerleaders*. Working Paper. Harvard University.
- Rodrik, D. (2007b). *The real exchange rate and economic growth: Theory and evidence*. Working Paper. Harvard University.
- Rodrik, D., & Subramanian, A. (2009). Why did financial globalization disappoint? *IMF Staff Papers*, 56(1), 112–138.
- Rodrik, D., & Velasco, A. (2000). Short-term capital flows. In B. Plescovic & J. E. Stiglitz (Eds.), *Annual World Bank Conference on Development Economics 1999* (pp. 59–70). Washington: World Bank.
- Rogoff, K. (2007). Impact of globalization on monetary policy. *Federal Reserve Bank of Kansas City Proceedings*, 2006, 265–305.
- Schindler, M. (2009). Measuring financial integration: A new dataset. *IMF Staff Papers*, 56(1), 222–238.
- Schmukler, S. L. (2004). Financial globalization: Gain and pain for developing countries. *Federal Reserve Bank of Atlanta Economic Review*, 2nd Quarter 2004, pp. 39–66.
- Spiegel, M. M. (2009). Financial globalization and monetary policy discipline: A survey with new evidence from financial remoteness. *IMF Staff Papers*, 56(1), 198–221.
- Stiglitz, J. (2002). *Globalization and its discontents*. New York: W. W. Norton and Company.
- Stiglitz, J. (2004). Capital-market liberalization, globalization, and the IMF. *Oxford Review of Economic Policy*, 20(1), 57–71.
- Stulz, R. (1999a). International portfolio flows and security markets. In M. Feldstein (Ed.), *International capital flows* (pp. 257–293). Chicago: University of Chicago Press.
- Stulz, R. (1999b). *Globalization of equity markets and the cost of capital*. NBER Working Papers 7021. Cambridge, MA: National Bureau of Economic Research.
- Stulz, R. (2005). The limits of financial globalization. *Journal of Finance*, 60(4), 1595–1637.
- Summers, R. H. (2000). International financial crises: Causes, prevention, and cures. *American Economic Review*, 90(2), 1–16.
- Tytell, I., & Wei, S. J. (2004). *Does financial globalization induce better macroeconomic policies?* IMF Working Paper 04/84. Washington: International Monetary Fund.
- Van Wincoop, E. (1999). How big are potential welfare gains from international risk sharing? *Journal of International Economics*, 47(1), 109–135.
- Wei, S. J. (2001). Domestic crony capitalism and international fickle capital: Is there a connection? *International Finance*, 4(1), 15–46.
- Wei, S. J., & Wu, Y. (2002). Negative alchemy? Corruption, composition of capital flows, and currency crises. In S. Edwards & J. Frankel (Eds.), *Preventing currency crises in emerging markets* (pp. 461–501). Chicago: University of Chicago Press.

- Wei, S. J., & Zhang, Z. (2006). *Collateral damage: Exchange controls and international trade*. IMF Working Paper 07/08. Washington: International Monetary Fund.
- Woodford, M. (2007). Globalization and monetary control. NBER Working Papers 13329. Cambridge, MA: National Bureau of Economic Research.
- Wurgler, J. (2000). Financial markets and the allocation of capital. *Journal of Financial Economics*, 58(1–2), 187–214.
- Wyplosz, C. (2004). Financial instability in emerging market countries: Causes and remedies. *Paper presented at the Forum on Debt and Development (FONDAD) Conference. Stability, growth and the search for a new development agenda: Reconsidering the Washington Consensus*, Santiago, Chile.
- Zettelmeyer, J. (2006). Growth and reforms in Latin America: A survey of facts and arguments. IMF Working Paper 06/210. Washington: International Monetary Fund.

Postscript to “Financial Globalization and Economic Policies”

M. Ayhan Kose, Eswar Prasad, Kenneth Rogoff and Shang-Jin Wei*

The global financial crisis serves as a reminder of the risks of financial globalization. After grappling with surges of capital inflows earlier in this decade, many emerging market and developing economies experienced a sharp reversal of those inflows in late 2008 as a result of the crisis. Moreover, international financial linkages clearly served as a channel transmitting the financial turmoil from advanced countries to the shores of emerging markets. These developments will re-ignite the fierce debate about the merits of financial globalization and its effects on growth and stability, especially for emerging market and developing countries.

As the crisis is still unfolding, it is premature to undertake a detailed analysis of its implications for the debate on financial globalization. Nevertheless, there are two preliminary observations that are pertinent. First, the differential effects of the crisis across countries confirm that it is not just financial openness, but a country's structural features and its precrisis policy choices that have determined the crisis' overall impact on a country. Second, the crisis has not led to a resurgence of capital controls in emerging market economies.

1. OPENNESS, STRUCTURAL FEATURES AND POLICIES

The overall impact of the crisis on economic activity has differed across emerging markets, depending on a variety of factors.⁶¹ In addition to the role played by a country's openness to trade and financial flows, its structural features and precrisis policy choices have determined the extent of damage caused by the crisis on a country's economy and financial system. For example, some emerging European countries with large current account imbalances have suffered disproportionately more as bank-related flows were curtailed. Since commodity prices have registered sharp declines, a number of commodity producing countries in Africa, Latin America, and the Middle East have been hit very hard. As the demand for durable products from advanced countries have plummeted, the impact of the crisis has become more pronounced in some East Asian countries with heavy dependence on manufacturing exports.

In addition, the crisis has inflicted more pain on countries with weak financial systems and inconsistent macroeconomic policies. Some emerging European countries heavily relied on short-term external financing to sustain their credit booms earlier in

this decade.⁶² As global financial flows have dried up, these economies have been hit very hard because of the collapse of their already fragile banking systems. Moreover, their policy choices have become rather complex due to their less flexible exchange rate regimes. Although the pain inflicted by the crisis has been felt in all corners of the globe, countries with stronger policy frameworks, sound financial systems, and healthy balance sheet positions have been much more successful in mitigating its effects.

In addition, countries receiving more of their inflows in the form of FDI or portfolio equity inflows have tended to weather the crisis better than those receiving inflows largely in the form of bank lending—as noted earlier, some emerging European economies are a prime example of the latter. Recent research further emphasizes the important role of the composition of capital inflows in determining the extent of pain caused by the crisis on nonfinancial firms (see Tong & Wei, 2009).

2. POLICIES REGULATING CAPITAL ACCOUNT

Early observations suggest that most countries have shown a clear preference for maintaining their capital accounts open during the crisis. Only a small number of countries have resorted to policies to restrict cross-border movements of capital despite the turmoil in global financial markets. For example, Iceland introduced a comprehensive set of controls amid the collapse of its financial system. Among emerging market countries, only Nigeria and Ukraine have employed substantial controls to limit capital outflows. By contrast, some emerging market economies have in fact further loosened restrictions on inflows and outflows rather than curbing them.

The crisis is expected to lead to substantial changes in the regulation and supervision of domestic financial markets. However, it has yet to be seen how the crisis would affect the policies regulating capital account transactions and the process of financial globalization. In any event, the crisis has once again shown the risks of excessive reliance on short-term debt flows, especially in the context of weak financial systems and less flexible exchange rate regimes.

End Notes

- *. IMF, Cornell University, Harvard University, and Columbia University, respectively. The views expressed in this chapter are solely those of the authors and do not necessarily reflect those of the IMF or IMF policy.
- 61. International Monetary Fund (2009) provides an analysis of the effects of the crisis on a large number of countries. Ghosh, Chamon, Crowe, Kim, and Ostry (2009) analyze policies available to emerging market economies to cope with the adverse effects of the crisis.
- 62. Cardarelli et al. (2007) examine the implications of surges in private capital inflows in a group of emerging market countries over the past two decades. They conclude that the consequences of large

References

- Cardarelli, R., Kose, M. A., & Elekdag, S. (2009). Capital Inflows: Macroeconomic Implications and Policy Responses. *IMF Working Paper*, WP/09/40.
- Ghosh, A. R., Chamon, M., Crowe, C., Kim, J. I., & Ostry, J. D. (2009). *Coping with the crisis: Policy options for emerging market countries*. IMF Staff Position Note, SPN/09/08.
- International Monetary Fund. (2009, April). Global prospects and policies. *World Economic Outlook*, 1–62.
- Tong, H., & Wei, S.-J. (2009). The composition matters: Capital inflows and liquidity crunch during the 2007–09 global economic crisis. *IMF Working Paper*, No. WP/09/164).

International Migration and the Developing World

Gordon H. Hanson

University of California, San Diego and NBER

Contents

1. Introduction	4364
2. The Dimensions of International Migration	4366
2.1 Data sources	4367
2.2 International migration patterns	4368
2.3 Skilled emigration	4375
2.4 Networks and migration costs	4383
2.5 Discussion	4385
3. Impact of Emigration on Sending Countries	4386
3.1 Labor markets and fiscal accounts	4386
3.2 Remittances and return migration	4390
3.3 Information and the flow of ideas	4394
3.4 Discussion	4395
4. Immigration Policy Regimes	4396
4.1 Political economy of immigration policy	4396
4.2 The design of immigration policy regimes	4398
4.3 Discussion	4402
5. Final Discussion	4403
End Notes	4405
References	4408

Abstract

In this chapter, I discuss the recent academic research on international migration, focusing on the causes and consequences of emigration from developing countries and the motivations behind the restrictions imposed by the developed countries on immigration. My aim is to identify facts about international migration relevant to those concerned about why labor moves between countries, how these movements affect the countries that send these laborers, and why the receiving countries are so selective about the immigrants that they admit.

JEL classifications: F22, J61, O15

Keywords

migration
immigration
emigration
labor flows
development

1. INTRODUCTION

International migration is now recognized as an important mechanism for globalization. Between 1990 and 2005, the number of individuals living outside of their country of birth increased from 154 to 190 million, reaching a level equivalent to 3% of the world population (United Nations, 2005). While there are sizable labor flows between low-income countries,¹ it is the rising flows from low- to high-income countries that have attracted most attention from scholars and policy makers.

As workers migrate from Latin America to the United States, Africa to Europe, or Southeast Asia to Australia, there is a global shift in labor supply from labor-abundant to labor-scarce economies. Absent dynamic adjustment in capital or technology, labor flows tend to lift wages in sending countries and depress them in receiving ones (Aydemir & Borjas, 2007), helping reduce international differences in factor prices. Migrants enjoy substantial income gains from moving abroad (Rosenzweig, 2007), which they share with family members through remittances. International labor flows respond to economic and political shocks, smoothing labor-market adjustment to macroeconomic fluctuations. The surge in emigration from Mexico following the 1995 peso crisis, for instance, may have dampened the wage impact of the country's harsh economic contraction (Hanson & Spilimbergo, 1999).

Moving labor across borders creates a conduit for the global transmission of ideas. Returning migrants, including students who have gone abroad for their education, arrive home with news about advances in foreign technology, exposure to alternative political systems, and contacts with foreign business. As of 2006, 45 heads of government were products of US higher education (Spilimbergo, 2006). The migration of Indian engineers to Silicon Valley in the 1980s later paved the way for US firms to outsource business services to Bangalore and Hyderabad (Saxenian, 1999, 2002), much as overseas Chinese business networks have come to intermediate trade between mainland China and the rest of the world (Rauch & Trindade, 2002).

The contribution of international migration to arbitraging wage differences, reducing macroeconomic volatility, diffusing knowledge across borders, and facilitating trade suggests that there may be substantial welfare gains for letting labor flow between countries. Yet, most labor-importing countries tightly restrict admissions. While the right to emigrate is codified in international treaties,² the right to immigrate lacks similar support.

There has never been a Washington consensus on international migration. Rich country impediments to immigration contrast with their pro-liberalization stances on trade and investment (Hanson, Scheve, & Slaughter, 2007; Hatton & Williamson, 2005). Although OECD countries have lowered barriers to foreign trade and capital in recent decades, they have not commensurately reduced barriers to foreign labor.

In the academic literature, there is ambivalence about international migration, reflected in a dissensus on global migration policy. Labor economists debate whether immigration benefits the receiving countries (e.g., Borjas, 1999a; Card, 2005) while development economists disagree on whether emigration is good for sending countries (e.g., Bhagwati & Hamada, 1974; Stark & Wang, 2002). These disputes arise in part from concerns that migration may exacerbate distortions in factor markets. Without such distortions, unrestricted migration would unambiguously raise global income and welfare (Hamilton & Whalley, 1984). However, in labor-importing countries, the existence of social-welfare programs (financed by nonlump-sum taxes) may make a departure from free immigration the constrained optimum, especially where low-skilled labor would dominate labor inflows (Wellisch & Walz, 1998). In labor-exporting countries, human-capital externalities and subsidies to higher education may create a second-best justification for taxing skilled emigration (McHale, 2007).

Given the cognitive dissonance in economics about globalization (trade and capital flows are unambiguously good, labor flows are unambiguously complicated), it is perhaps no surprise that in their dealings with the developed world developing countries devote much more time to negotiating international trade and investment agreements than to discussing international migration. Should developing country policy makers be more optimistic about the capacity for emigration to raise living standards? Are there environments where emigration is particularly helpful or harmful? Is there scope for international coordination on migration policy?

In this chapter, I discuss recent academic research on international migration with the aim of evaluating the causes and consequences of emigration from developing countries and the motivations behind the developed-country restrictions on immigration. My goal is not to give an exhaustive account of the literature but rather to identify facts about international migration relevant to those concerned about why labor moves between countries, how these movements affect the sending-country economies, and why the receiving countries are so selective about the immigrants they admit.³

Section 2 begins with a brief discussion of the current patterns in international migration. International labor movements are on the rise, with considerable variation across sending countries in terms of how many individuals emigrate, who selects into migration, and where they go. In general, the more educated account for a disproportionate share of emigrants and they tend to move to countries that offer relatively high rewards for their skill. Distance, language, and migration networks also appear to shape international labor movements.

While there is an enormous literature on the consequences of immigration for receiving countries (and for the United States, in particular),⁴ research on impacts on sending countries, as discussed in Section 3, is less developed. Labor outflows tend to put an upward pressure on the wages in sending countries, with these effects concentrated in specific regions and among the young. Despite four decades of research on brain drain, we still do not know whether skilled emigration raises or lowers the stock of human capital in sending countries. It does appear that contacts with the outside world promote trade flows, technology adoption, and political openness. Recently, measured remittances have risen much faster than emigration. While these income flows have raised consumption, and perhaps educational investments, among family members at home, there is a confusing tendency in the literature to portray remittances as a causal factor in development rather than as a consequence of intra-household specialization.

With global migration policy being controlled by labor-importing countries, the determination of these policies is an important factor in the economic development of labor exporters. As discussed in Section 4, theoretical literature explains immigration restrictions as the result of political pressure by groups that would be hurt by labor inflows and concerns over the fiscal consequences of immigration. Less explored is the *design* of immigration policy (Cox & Posner, 2007). Legal immigration tends to be governed by quotas, whose number varies according to explicit selection criterion. Illegal immigration, which makes up a rising share of global labor flows, is governed by enforcement policies, which define an implicit selection criterion for unauthorized migrants by determining the smuggling fee they pay to enter a country and the wage discount they must endure on the job. The screening of entrants at the border highlights informational problems in identifying ‘desirable’ immigrants, which may offer a rationale for multilateral coordination on migration policy.

By way of conclusion, in Section 5, I identify gaps in the literature regarding the causes and consequences of emigration, outline directions for future research, and briefly discuss implications drawn from the literature for whether labor-exporting countries should view emigration as a mechanism for development.

2. THE DIMENSIONS OF INTERNATIONAL MIGRATION

There is a widespread perception that international migration is on the rise. Until recently, however, one would have been hard pressed to examine global labor movements for more than a handful of countries. It is only in this decade that cross-country data on emigrant stocks have become available, implying that research on international migration is still at an early stage. I begin this section with a review of data sources on the stock of international migrants and then address rates of emigration in sending countries, the magnitude of immigration in receiving countries, the correlates of bilateral migration flows, and issues related to the emigration of skilled labor.

2.1 Data sources

There have been several recent attempts to measure international migration.⁵ In an earlier effort, Carrington and Detragiache (1998) estimated emigration rates in 1990 for individuals with tertiary education from 61 source countries to OECD destination countries. Their approach was based on data that provide limited information on the educational attainment of immigrants, requiring them to use the characteristics of US immigrants to impute the schooling of immigrants in other receiving countries. Adams (2003) applied a similar methodology to estimate emigration rates for 24 large labor-exporting countries in 2000. In an attempt to cover a larger number of sending countries, the OECD (2003) listed the foreign-born population 15 years and older in 2000 by source country and education level (primary, secondary, tertiary, unknown) for each OECD country.⁶ As these population counts include some individuals still in school, the reported education of the foreign-born population may not reflect the immigrants' ultimate educational attainment. A further problem with the OECD data is that schooling is unknown for a substantial portion of the foreign-born population.

In an important recent work, Docquier and Marfouk (2006) extended the OECD data by constructing more complete estimates of the stocks of international migrants. They used the population census of 30 OECD countries in 1990 and 2000 to obtain the count of adult immigrants (25 years and older) by source country and level of education (primary, secondary, or tertiary schooling). They combine these counts with the size of adult populations and the fraction of adult populations with different levels of schooling from Barro and Lee (2000) to obtain emigration rates by education level and source country. The DM data contain 174 source countries in 1990 and 192 in 2000. While the set of source countries is comprehensive, the coverage of destination countries excludes those countries not in the OECD as of 2000.⁷

There are myriad problems in assembling the DM data, which underscore the complications involved in obtaining an accurate description of international migration.

- OECD countries differ in how they define immigrants. While most national censuses identify immigrants as individuals born abroad (with foreign citizenship at birth), a few countries (Germany, Greece, Italy, Japan, Korea) only identify immigrants who are noncitizens (though most immigrants appear not to have naturalized).
- In some countries (Belgium, Greece, Portugal) the census identifies immigrants but not their education levels, making it necessary to estimate the allocation of immigrants across education groups using data from household surveys.
- OECD countries differ in how they define educational attainment, with some (Austria, Germany) recording educational certification (e.g., a vocational degree) rather than the highest grade of schooling completed, which complicates the assignment of individuals to categories based on years of schooling.
- Some immigrants may have completed a portion of their schooling in the destination country, making emigration rates by education level difficult to interpret. Beine, Docquier, and Rapoport (2006a) address this issue by extending the DM data to

include counts of migrants based on the age at arrival in the destination country, with older-arriving immigrants likely having completed more of their schooling at home.

- Some immigrants are students who will return to their home countries after completing their education, suggesting a portion of the immigrant stock may not be permanent. DM addresses this issue in part by including only individuals 25 years and older, most of whom have completed their schooling.
- DM coverage of illegal immigrants is unknown. The immigrants in the DM data are those captured by national censuses, most of which seek to enumerate all long-term residents, regardless of their legal status. Still, illegal immigrants may be undercounted in censuses, due to tendencies to reside in irregular housing units or to avoid being identified by government authorities.⁸

With these caveats in mind, the DM data (and their extension by Beine et al., 2006a) are the most comprehensive available on international migration.

Also, noteworthy are several recent surveys that track individual migrants over time in specific sending or receiving countries, the two most important of which are the New Immigrant Survey (NIS) in the United States and the Mexican Family Life Survey (MXFLS). The United States and Mexico are, respectively, the world's largest labor-importing and labor-exporting country. The NIS (<http://nis.princeton.edu/>) is a random sample of individuals who received a US legal permanent residence visa (or green card) in 2003–2004 (and who were resurveyed in 2007). While the NIS excludes temporary and illegal immigrants, it does contain detailed information on the type of visa the immigrant has obtained (e.g., family-based, employment-based, or refugee-based), which is missing in most other data on international migration. MXFLS (www.mxfls.uia.mx/) is a random sample of 10,500 households in Mexico in 2002, whose members were located and resurveyed in 2005, wherever they happened to be living. This survey structure provides observations on individuals before and after they chose to migrate internally (within Mexico) or externally (to the United States).

2.2 International migration patterns

Low-income countries are an increasingly important source of migrants to high-income countries. Table 1 shows the share of the immigrant population in OECD countries by sending-country region.⁹ In 2000, 67.2% of immigrants in the OECD were from a developing country, up from 54.0% in 1990. This gain came almost entirely at the expense of Western Europe, whose share of OECD immigrants fell from 35.5% to 24.4%. Among developing sending regions, Mexico, Central America, and the Caribbean are the most important, accounting for 20.2% of OECD immigrants in 2000, up from 14.9% in 1990. Half of this region's migrants come from Mexico, which in 2000 was the source of 11.3% of OECD immigrants, making it by far and away the world's largest supplier of international migrants.¹⁰ As seen in Table 2, the next most important developing source countries for OECD immigrants are Turkey (3.5% of OECD immigrants), China, India, and the Philippines (each with 3.0%).

Table 1 Share of OECD immigrants by source region, 2000

Developing source region	Destination region				Change in all OECD
	All OECD	North America	Europe	Australia, Oceania	Immigrant share 1990-2000
Mexico, Central					
America, Caribbean	0.202	0.374	0.025	0.002	0.053
Southeast Asia	0.102	0.137	0.039	0.160	0.016
Eastern Europe	0.099	0.049	0.161	0.116	0.042
Middle East	0.063	0.032	0.113	0.029	0.001
South Asia	0.052	0.052	0.055	0.036	0.011
North Africa	0.044	0.009	0.098	0.018	-0.006
South America	0.041	0.050	0.031	0.035	0.010
Central, Southern					
Africa	0.036	0.021	0.061	0.021	0.007
Former Soviet Union	0.029	0.023	0.042	0.010	-0.002
Pacific Islands	0.004	0.003	0.001	0.027	0.000
Subtotal	0.672	0.750	0.626	0.454	0.132
High income source region					
Western Europe	0.244	0.152	0.336	0.368	-0.111
Asia, Oceania	0.055	0.062	0.018	0.156	-0.010
North America	0.029	0.037	0.020	0.023	-0.011
Subtotal	0.328	0.251	0.374	0.547	-0.132

Notes: This table shows data for 2000 on the share of different sending regions in the adult immigrant population of the entire OECD and of three OECD sub-regions. High Income North America includes Canada and the United States, and High Income Asia and Oceania includes Australia, Hong Kong, Japan, Korea, New Zealand, Singapore, and Taiwan.

The growing importance of lower-income countries in the supply of international migrants has contributed to an overall increase in labor flows into rich countries. Table 3 shows the share of the population that is foreign born in select OECD members. The size of the immigrant population varies across destinations, reflecting

Table 2 Share of OECD immigrants by source country, 2000

Source country	Destination region			
	All OECD	North America	Europe	Australia, Oceania
Mexico	0.113	0.219	0.001	0.000
United Kingdom	0.053	0.041	0.027	0.206
Italy	0.042	0.027	0.062	0.038
Germany	0.038	0.028	0.049	0.045
Turkey	0.035	0.003	0.085	0.005
India	0.030	0.038	0.023	0.018
China	0.030	0.039	0.009	0.066
Philippines	0.030	0.046	0.007	0.030
Vietnam	0.022	0.032	0.008	0.026
Portugal	0.022	0.011	0.040	0.002
Korea	0.021	0.025	0.002	0.075
Poland	0.020	0.019	0.024	0.010
Morocco	0.019	0.002	0.048	0.000
Cuba	0.015	0.028	0.002	0.000
Canada	0.015	0.025	0.004	0.006
France	0.014	0.007	0.027	0.005
United States	0.014	0.012	0.015	0.017
Ukraine	0.013	0.009	0.022	0.002
Spain	0.013	0.004	0.027	0.002
Greece	0.013	0.008	0.015	0.028
Serbia	0.013	0.003	0.027	0.009
Jamaica	0.012	0.019	0.006	0.000
Ireland	0.012	0.006	0.021	0.009
El Salvador	0.012	0.022	0.000	0.001
Netherlands	0.011	0.007	0.014	0.020

Notes: This table shows the share of adults immigrants in regions of the OECD accounted for by the 25 largest source countries in 2000.

Table 3 Share of foreign-born population in total population, OECD countries

	Change				
	1995	2000	2002	2004	1995-2004
Australia	23.0	23.0	23.2	23.6	0.6
Austria	–	10.5	10.8	13.0	–
Belgium	9.7	10.3	11.1	–	–
Canada	16.6	17.4	17.7	18.0	1.4
Czech Republic	–	4.2	4.6	4.9	–
Denmark	4.8	5.8	6.2	6.3	1.6
Finland	2.0	2.6	2.8	3.2	1.2
France (a)	–	10.0	–	–	–
Germany (b)	11.5	12.5	12.8	12.9	1.4
Greece (c)	–	10.3	–	–	–
Hungary	2.8	2.9	3.0	3.2	0.4
Ireland (d)	6.9	8.7	10.0	11.0	4.0
Italy (c)	–	2.5	–	–	–
Luxembourg	30.9	33.2	32.9	33.1	2.2
Mexico	0.4	0.5	–	–	–
Netherlands	9.1	10.1	10.6	10.6	1.6
New Zealand (d)	16.2	17.2	18.4	18.8	2.6
Norway	5.5	6.8	7.3	7.8	2.3
Poland	–	–	1.6	–	–
Portugal	5.4	5.1	6.7	6.7	1.3
Slovak Republic (c)	–	2.5	–	3.9	–
Spain (c)	–	5.3	–	–	–
Sweden	10.5	11.3	11.8	12.2	1.7
Switzerland	21.4	21.9	22.8	23.5	2.2

Continued

Table 3 Share of foreign-born population in total population, OECD countries—Cont'd

	Change				
	1995	2000	2002	2004	1995-2004
Turkey	—	1.9	—	—	—
United Kingdom	6.9	7.9	8.6	9.3	2.3
United States	9.3	11.0	12.3	12.8	3.5

Notes: (a) 2000 value is from 1999; (b) 2004 value is from 2003; (c) 2000 value is from 2001; (d) 1995 value is from 1996.

Source: *International Migration Outlook*, OECD (2006) edition.

differences in both their attractiveness and openness to international migrants. Aside from tiny Luxembourg, the countries with the largest immigrant presence in 2004 are Australia (23.6%), Switzerland (23.5%), New Zealand (18.8%), and Canada (18.0%). Next in line are the large economies of Germany (12.9%), the United States (12.8%), France (10.0%), and the United Kingdom (9.8%), with the United States alone hosting 40% of immigrants living in OECD countries. In the last decade, there have been substantial increases in foreign-born population shares in a number of rich countries, with the largest changes over 1995–2004 occurring in Ireland (4.0%), the United States (3.5%), New Zealand (2.6%), the United Kingdom (2.3%), Norway (2.3%), and Switzerland (2.2%).¹¹ Japan is at the other extreme of openness to immigration, with a foreign-born population of less than 1.5% of its total population.

There is abundant evidence that a rising share of labor inflows in rich countries are made up by illegal entrants, with data for the United States being the most extensive. Passel (2006) estimates that in 2005 illegal immigrants accounted for 35.2% of the US foreign-born population, up from 28.0% in 2000 and 19.3% in 1996. Of the 2005 population of illegal immigrants, 56% were from Mexico, implying that 60% of the population of Mexican immigrants in the United States was unauthorized.¹²

Apparent in Table 1 is a tendency for different destination regions to draw more heavily on migrants from particular source countries. Mexico, Central America, and the Caribbean together are the largest source region for North America, but send few migrants to other parts of the world; Eastern Europe is the most important developing source region for OECD Europe; and Southeast Asia is the most important developing source region for Australia and Oceania. Proximity clearly plays a role in bilateral migration flows, but, as we discuss below, so do other factors, including income.

There is substantial variation across countries in the propensity to emigrate. As of 2000, there were 25 countries (22 of which were developing nations) with 10% or

more of their adult population having migrated to the OECD, and another 24 countries (16 of which were developing) with emigration rates above 5%.¹³ At the other extreme, 54 countries (52 of which were developing) had emigration rates below 1%. Figure 1A plots emigration rates for countries against their log population densities in 2000, where the emigration rate is the fraction of a country's adult population that has migrated to an OECD country. There is a clear positive relation between emigration rates and population density, indicating that more densely populated countries tend to send a higher fraction of their population abroad.¹⁴

The relation between emigration rates and income is more complex, as seen in Figure 1B. There appears to be a threshold level of per capita GDP—of approximately \$3000 at 2000 PPP-adjusted prices—below which emigration rates are very low. Above this threshold, emigration is strongly decreasing in average income.

There is a growing literature on the correlation between international migration and income. In a recent contribution, Clark, Hatton, and Williamson (2007) regress the emigration flow to the United States on a large number of sending-country characteristics for a panel of 81 countries over the period 1971–1998. They calculate the emigration flow as the log ratio of US legal immigrants admitted to the source-country population.¹⁵ Consistent with Figure 1A, they find an inverted U in the relationship between sending-country average income and emigration, with emigration rates increasing in income at low income levels and decreasing in income at higher income levels. In their data, the per capita GDP level at which the emigration rate is at a maximum is 10% of US per capita GDP (or \$3400 at 2000 PPP-adjusted prices).

Clark et al. also find that migration flows to the United States are higher for countries that speak English, are geographically closer to the United States, and have large existing populations of US immigrants. They estimate an elasticity of emigration flows with respect to a distance of -0.20 to -0.28, which would imply that in moving from El Salvador (3400 km. from the United States) to Brazil (7700 km from the United States) emigration to the United States would fall by 20%.¹⁶ The positive correlation between current migration flows and lagged migration stocks may reflect migration networks—in which older migrants help newer migrants in becoming established in a destination country—or provisions in US immigration policy that favor relatives of US residents in the granting of entry visas.¹⁷

The regression approach in Clark et al. follows a tradition in the migration literature of estimating bilateral migration flows as a function of characteristics in the source and destination countries only.¹⁸ This imposes the assumption that opportunities for migration to other destinations do not affect bilateral flows for a given country pair. In theory, however, migration from, say, Ecuador to Spain should be affected not just by conditions in the two countries but also by what is happening in other potential destinations that Ecuadorian migrants might consider. The problem of other destinations is analogous to one that arises in the estimation of the gravity

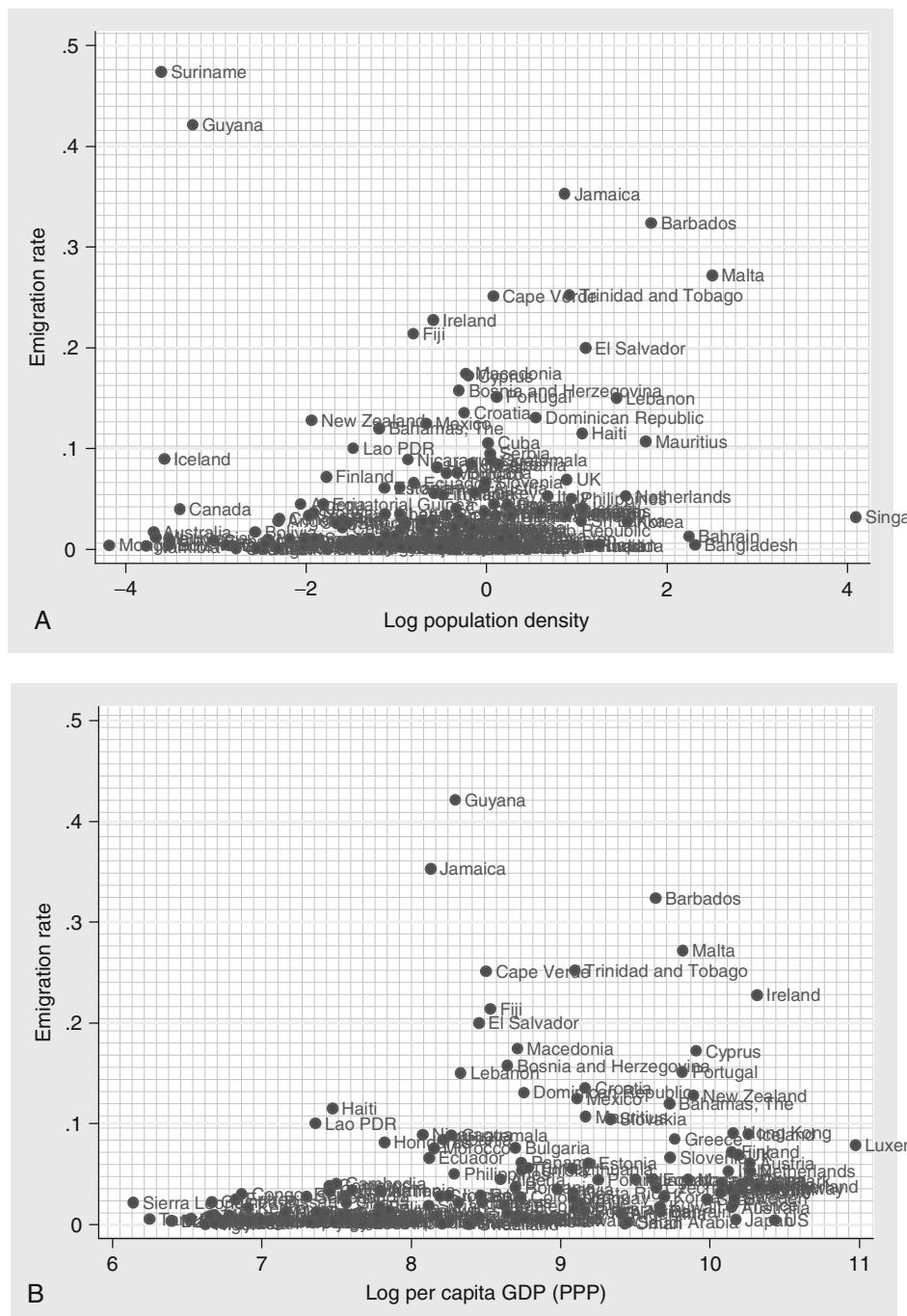


Figure 1 (A) Emigration rates and population density, 2000. (B) Emigration rates and per capita GDP, 2000.

model of international trade. Anderson and van Wincoop (2004) show that failing to control for the opportunities to buy from or sell to other trading partners—which they refer as the multilateral resistance to trade—results in biased estimates of the determinants of bilateral trade. As shown in Section 2.4, failing to control other migration opportunities could similarly produce biased estimates of the determinants of bilateral migration.

One attempt to address the issue of other destination countries is given by Mayda (2005), who examines bilateral migration between a large number of source countries and 14 OECD destination countries over the period 1980–1995. She regresses bilateral migration rates on income per capita in the source country, income per capita in the destination country, and average income per capita in other OECD destinations, among other control variables. Bilateral migration is increasing in destination-country income and decreasing in the income of other destinations, consistent with the idea that better economic conditions in third countries deflect migration away from a given destination. One limitation of this specification is that it relies on the assumption that average income in other destinations is a sufficient statistic for other migration opportunities.

2.3 Skilled emigration

2.3.1 Brain drain

Much of the literature on international migration focuses on the movement of skilled labor, whose departure may drain poor economies with scarce supplies of human capital. Figure 2 plots the emigration rate for adults with a tertiary education against the emigration rate for all adults. In 2000, there were 44 countries (41 developing) with emigration rates for the tertiary educated above 20%.

The Docquier and Marfouk data could overstate the extent of brain drain, since many tertiary emigrants might not have attained postsecondary education had they not had the opportunity to study abroad. To consider this possibility, Beine et al. (2006a) examine the age of arrival for tertiary emigrants in the DM data. They find that 68% of tertiary migrants arrive in the destination country at age 22 or older, 10% arrive between ages 18 and 21, and 9% arrive between ages 12 and 17, suggesting the majority of tertiary emigrants depart sending countries at an age at which they would typically have acquired at least some postsecondary education. Another way in which tertiary emigration rates could overstate brain drain is if individuals migrate for the purpose of obtaining their education and then spend some time working in the destination country before returning home. What appears to be brain drain would actually be brain circulation. In the case of migration to the United States, Rosenzweig (2006) finds that 20% of foreign university students remain in the United States.¹⁹

Brain drain is a concern where there are distortions in the acquisition of skill. Absent distortions, moving labor from a low-productivity to a high-productivity environment unambiguously raises global income (Benhabib & Jovanovic, 2007; Hamilton &

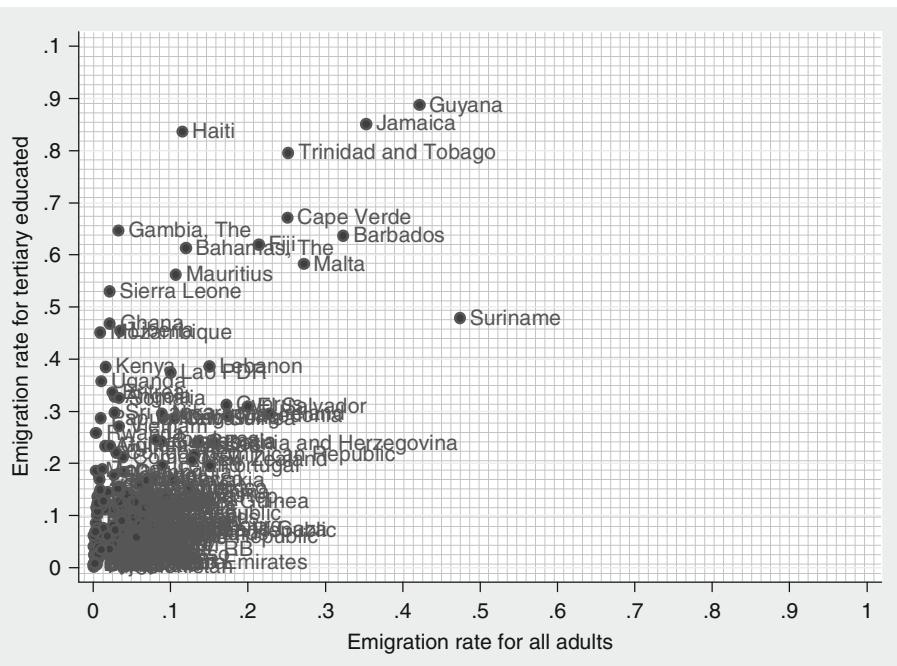


Figure 2 Emigration rates for the more educated, 2000.

Whalley, 1984). However, if there are positive externalities associated with learning (e.g., Lucas, 1988), then the social product of human capital exceeds its private product and the exodus of skilled labor from a country may have adverse consequences for its economic development (Bhagwati & Hamada, 1974; Grubel & Scott, 1966; McColloch & Yellen, 1977). Another negative impact of brain drain is that many individuals have their education subsidized by the state, meaning their emigration would deprive their origin country of tax contributions to offset the cost of their schooling (Bhagwati & Rodriguez, 1975).

Recent literature counters that the opportunity for emigration may actually increase the supply of human capital in a country, creating a brain gain (Stark, Helmenstein, & Prskawetz, 1997; Stark & Wang, 2002). The logic is that, with relatively high incomes for skilled labor in rich countries and uncertainty over who will succeed in emigrating, the option of moving abroad induces individuals to accumulate enough additional human capital to compensate for the loss in skill to labor outflows (Beine, Docquier, & Marfouk, 2001; Mountford, 1997).²⁰ Crucial for this argument is that the probability of emigrating is large enough to affect the expected return to investing in skill (and that this probability does not vary too much across individuals, such that many people believe they have a nontrivial chance of moving abroad).

Only a handful of empirical papers examine the relationship between emigration and human-capital accumulation. For a cross-section of countries, Beine, Docquier, and Rapoport (2006b) report a positive correlation between emigration to rich countries (measured by the fraction of the tertiary educated population living in OECD countries in 1990) and the increase in the stock of human capital (measured as the 1990–2000 change in the fraction of adults who have tertiary education).²¹ While this finding is consistent with emigration increasing the incentive to acquire education, the cross-section correlation between emigration and schooling is not well suited for causal inference about the impact of brain drain on educational attainment. Education and migration decisions are likely to be jointly determined, making each endogenous to the other. Valid instruments for migration are very difficult to find. For causal analysis, one would need to observe changes in human-capital accumulation in sending countries before and after they experienced unexpected and exogenous shocks in the opportunity to emigrate. Such experiments have yet to be found in the data.

2.3.2 Inequality and selection into migration

The relatively high propensity for the highly educated to migrate abroad is seen clearly in Figure 3, which plots the share of emigrants with tertiary education against the share

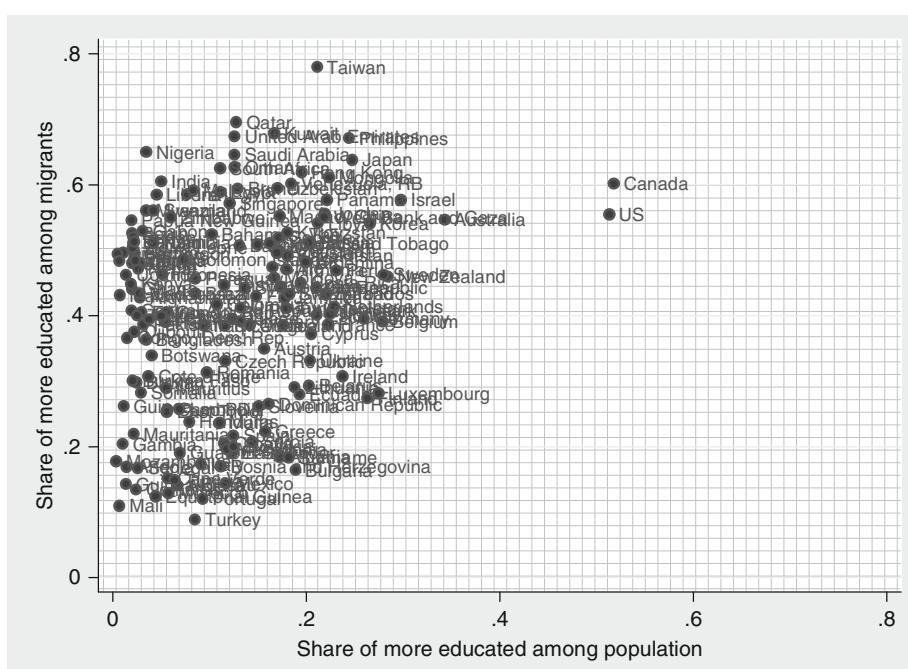


Figure 3 Selection of emigrants in terms of education, 2000.

of the general population with tertiary education in 2000. Nearly all points lie above the 45-degree line, indicating that in the large majority of countries emigrants are positively selected in terms of schooling.

Positive selection of emigrants is at odds with much recent empirical literature on international migration. In an influential line of work, Borjas (1987) uses the Roy (1951) model to show how migration costs and international variation in the premium for skill shape the incentive to migrate.²² In countries with low average wages and high wage inequality, as appears to be the case in much of the developing world, there is negative selection of emigrants. Those with the greatest incentive to relocate to rich countries (which tend to have high average wages and low wage inequality) are individuals with below-average skill levels in their home countries.

Much of the recent empirical research on Borjas' negative-selection hypothesis examines labor movements either from Mexico to the United States or Puerto Rico to the US mainland. Puerto Rican outmigrants tend to have low education levels relative to nonmigrants (Borjas, 2006; Ramos, 1992), consistent with migrants being negatively selected in terms of skill. Mexican emigrants, however, appear to be drawn more from the middle of the country's schooling distribution,²³ consistent instead with intermediate selection. Figure 3 suggests Mexico and Puerto Rico are exceptional cases. Positive selection of emigrants appears to be a nearly universal phenomenon.

Despite overwhelming evidence that emigrants are positively selected in terms of schooling, there is confusion in the literature over the relationship between income inequality and the incentive to emigrate. A common empirical approach is to explain bilateral migration using sending-country per capita GDP and income inequality (e.g., as measured by the GINI coefficient) relative to the receiving country (e.g., Clark et al., 2007; Mayda, 2005). A positive parameter estimate on the GINI coefficient is interpreted as an indicator that migrants are negatively selected in terms of skill.

However, this approach characterizes selection into migration only under restrictive conditions. In Borjas (1987), migration costs are assumed to be constant across individuals in time-equivalent units, in which case an individual will choose to migrate from source-country s to destination-country d as long as

$$\ln W_d - \ln W_s > \pi, \quad (1)$$

where π is the amount of labor time lost in migration (measured in terms of the sending-country wage). Expressing wages in Mincerian terms

$$W_k = e^{\mu_k + \delta_k z}, \quad (2)$$

where W_k is the wage in country k , μ_k is the base log wage (the log wage for an individual with zero skill) in country k , δ_k is the skill premium (the change in the log wage from acquiring an additional unit of skill) in country k , and z is an individual's skill level. We can then rewrite Eq. (1) as

$$(\mu_d - \mu_s) + z(\delta_d - \delta_s) > \pi, \quad (3)$$

which indicates that as long as $\mu_d - \mu_s > 0$ and $\delta_d - \delta_s < 0$ (i.e., source-country s has relatively low base wages and high wage inequality), the incentive to emigrate is decreasing in an individual's skill level.

Borjas (1991) shows that the result on negative selection in Eq. (3) is obtained only for a special version of the Roy model. Since the migration cost is constant in terms of labor time, the monetary cost of migration is higher for more-skilled individuals. Where the cost of migration is nonincreasing in skill—as for migration costs that are fixed in monetary units—migrants may be negatively or positively selected. Credit constraints in sending countries could make migration costs *decreasing* in skill. Suppose, for instance, education and migration are subject to a fixed monetary cost, credit-market imperfections make wealthier individuals subject to lower borrowing costs (e.g., Banerjee & Newman, 1993; Rapoport, 2002). Then, the wealthier will be more likely to become educated and more likely to migrate abroad (Assuncao & Carvalho, 2009). For Mexico, McKenzie and Rapoport (2007) find an inverted U-shaped relationship between migration and wealth, consistent with low-wealth individuals being too poor to afford migration and high-wealth individuals having an incentive not to leave.

More generally, it is not the Mincerian skill premium that shapes the skill composition of migrants, as in Eq. (3), but the overall reward to skill. This implication is present in the study by Rosenzweig (2007), who derives a Roy model of migration with moving costs that include components that are fixed in monetary units and time-equivalent units. It is the *level* difference in the reward to skill between source and destination that conditions migrant selectivity. Suppose in Nigeria someone with a primary education would earn \$1000 a year and someone with a tertiary education would earn \$10,000 a year, the comparable sums in the United States are \$20,000 and \$60,000, respectively. In Nigeria, the base wage is relatively low (the Nigeria-US log difference in base wages is -3) while the skill premium is relatively high (the Nigeria-US difference in the log wage ratio of high-skilled to low-skilled labor is 1.2), and the gross income gain to migration is much larger for the more educated person (\$50,000 vs. \$19,000). As long as migration costs are nondecreasing in skill and the marginal utility of income is not strongly decreasing, the incentive to emigrate would be much larger for the more educated person.

Is the positive selection of emigrants evident in Figure 3 due to cross-country differences in the income that accrues to skill? To see how one might answer this question, consider the difference in the incentive to migrate from country s to country d between an individual with skill level z and an individual with zero skill. Suppose that the cost of migrating from s to d is given by

$$C_{sd} + \theta_{sd}z + \epsilon_{sd}, \quad (4)$$

where C_{sd} is a fixed migration cost, θ_{sd} is the differential migration cost per unit of skill (which may be positive or negative), and ϵ_{sd} is a mean zero random migration cost (which for simplicity is assumed to be uncorrelated with z). Using Eq. (2), a skilled individual will be relatively likely to emigrate if

$$(e^{\mu_d + \delta_d z} - e^{\mu_s + \delta_s z}) - (e^{\mu_d} - e^{\mu_s}) > \theta_{sd}z. \quad (5)$$

Imposing the simplifying assumption that migration costs are skill neutral ($\theta_{sd} = 0$), the condition for positive selection in Eq. (5) becomes

$$e^{\mu_d - \mu_s} > \frac{e^{\delta_s z} - 1}{e^{\delta_d z} - 1}. \quad (6)$$

On the basis of Eq. (6), the selection of emigrants in terms of skill is ambiguous. Working in favor of positive selection is the presumed relatively high base wage in the destination country ($\mu_d > \mu_s$); working against positive selection is the presumed relatively high skill premium in the sending country ($\delta_d < \delta_s$).

In principle, with coefficients from Mincer wage regressions run in different countries, it would be possible to evaluate whether the condition for positive selection in Eq. (6) is satisfied. While Psacharopoulos and Patrinos (2004) report that the Mincerian skill premium has been estimated for at least 82 countries, it is difficult to compare the magnitudes of the reported estimates, as there is enormous variation across studies in control variables, skill definitions, and sample restrictions. One recent exception is the study by Hanushek and Zhang (2006), who use micro data from the International Adult Literacy Survey to estimate wage regressions for individuals in 13 countries. Their results suggest the Mincerian return to an additional year of schooling is 0.064 in the United States, 0.076 in Chile, and 0.080 in Poland. All else equal, these estimates indicate that there would be more tertiary educated than primary educated emigrants from Poland (Chile) to the United States, as long as the US base wage was at least 1.5 (1.3) times that in Poland (Chile).

To obtain a crude estimate of relative base wages, let GDP in country k be given by $Y_k = K_k^{1-\alpha} (\sum_i \lambda_{ik} L_k)^\alpha$, where K_k is the capital stock, L_k is the number of workers

(under the assumption that each worker supplies one person year of labor), and λ_{ik} is the productivity of the i th worker relative to a worker with no skill, which is assumed proportional to the number of years of education. The base wage (i.e., that for a worker with $\lambda_{ik} = 1$) can be written as $W_k = (Y_k/L_k)_k^{-1}$, where Y_k/L_k is GDP per worker and k is mean years of schooling. Using this formulation, in 2000 the estimated ratio of base wages (in PPP terms) for the United States and Poland is 2.7 and for the United States and Chile is 1.8, suggesting both high-inequality Chile and low-inequality Poland satisfy the condition in Eq. (6) for emigrants to the United States to be positively selected. Figure 3 shows that emigrants from both countries are in fact positively selected in terms of schooling.²⁴

Rosenzweig (2007) examines migrant selectivity with data from the New Immigrant Survey. The NIS reports the wage an individual earned in his last job before coming to the United States, which Rosenzweig uses to estimate the marginal product of labor (per efficiency unit) by source country.²⁵ A country's overall emigration rate to the United States is decreasing in the marginal product of labor, suggesting countries with higher labor productivity send fewer migrants to the United States. Rosenzweig estimates that raising a country's marginal product of labor by 10% relative to the United States would reduce the number of emigrants obtaining US employment-based visas by 8.3%. The average schooling of emigrants to the United States is increasing in the marginal product of labor, indicating that in countries with higher labor productivity it is the more educated migrants who are most likely to leave. In a related work, Rosenzweig (2006) finds that the number of students who come to the United States for higher education and who stay in the United States after completing their education are each decreasing in the marginal product of labor in the source country, suggesting that low rewards to skill in a country induce students seeking university training to pursue their schooling abroad.

Any analysis of migrant selection based on observed characteristics leaves open the question of how migrants are selected on unobservables. McKenzie, Gibson, and Stillman (2006) examine this issue using data on Tonga, in which individuals may apply to a lottery to obtain a visa to move to New Zealand. Comparing visa applicants who lost the lottery (meaning they stayed in Tonga) with nonapplicants, they find that those desiring to migrate have higher earnings, controlling for observed characteristics, suggesting prospective migrants from Tonga are positively selected in terms of unobserved skill. McKenzie, Gibson, and Stillman find that failing to account for selection on unobservables leads to substantial overstatement of the gains to migration.

What does Eq. (5) imply about migration flows between countries? Suppose the random migration cost in Eq. (4) has an extreme value distribution, with the correlation in these costs across destination countries being equal to $0 \leq \sigma < 1$, in which case

we can apply results in Berry (1994) to write the log odds of migrating to destination country d for skill group z from source-country s as

$$\ln \frac{E_{sd}^z}{E_s^z} = [W_d(z) - W_s(z)] - \theta_{sd} z + \sigma \ln E_{sd/D}^z + \eta_{sh}^z, \quad (7)$$

where E_{sh}^z is the share of skill group z in s that migrates to d , E_s^z is the share of skill group z that remains in s , $E_{sd/D}^z$ is the fraction of skill-group z emigrants from s that choose destination d , and η_{sh}^z is a disturbance term. The bilateral migration flow is increasing in the level difference in wages between the source and destination, decreasing in bilateral migration costs, and increasing in the attractiveness of the destination relative to alternative destinations, as captured by $E_{sd/D}^z$. Equation (7) demonstrates the problem by estimating bilateral migration flows as a function of bilateral country characteristics only: the relative attractiveness of a destination is likely to be correlated with its own wages and migration costs, meaning that excluding $E_{sd/D}^z$ from the regression in Eq. (7) would introduce omitted variable bias into the specification.

One implication of Eq. (7) is that the higher the reward to skill in a destination, the more skilled will be the composition of its immigrants. Grogger and Hanson (2008) develop a fixed-effects estimator for Eq. (7) and, using data from Beine et al. (2006a), find that the bilateral flow of more educated migrants (relative to less educated migrants) is increasing in the destination-country earnings gap between high-income and low-income workers. These results can account for the observed pattern of emigrant sorting across destinations, seen in Table 4. The United States is by far and away the largest destination country for international migrants, with Canada being the second largest. In 2000, 53% of the foreign-born population in OECD countries resided in North America, while 36% resided in the European

Table 4 Share of OECD immigrants by destination region and education level, 2000

Destination region	Education group			
	All	Primary	Secondary	Tertiary
North America	0.514	0.352	0.540	0.655
Europe	0.384	0.560	0.349	0.236
Australia and Oceania	0.102	0.088	0.111	0.109
All OECD		0.355	0.292	0.353

Notes: This table shows the share of OECD immigrants by destination region and education group in 2000. Source: Grogger and Hanson (2008).

Union, and 10% resided in Australia and Oceania. The draw of United States and Canada is the strongest for the more educated. While North America attracts only 38% of emigrants with primary education, it attracts 66% of emigrants with tertiary education. In Europe, the shares are flipped, as it attracts 22% of emigrants with tertiary schooling and 53% of emigrants with primary schooling. This pattern of emigrant sorting is consistent with observed differences in the reward to skill. Among OECD destinations, the level of difference in income between high- and low-skill labor is largest in the United States, with Canada having the fourth-largest difference (the United Kingdom and Australia coming in at numbers two and three). Continental Europe, on the other hand, has a relatively low income gap between high- and low-skill labor. The consequence of these income differences appears to be that North America and Australia attract a more-skilled mix of immigrants, while Continental Europe attracts a less-skilled mix.

2.4 Networks and migration costs

Although the evidence in Table 3 points to growth in international migration, the global stock of emigrants remains small, at around 3% of the world population. This is surprising, given that the gains to international migration appear to be enormous. Hanson (2006) reports that in 2000 the average hourly wage for a male with nine years of education was \$2.40 in Mexico and \$8.70 for recent Mexican immigrants in the United States (in PPP-adjusted prices). At the average labor supply for adult male workers of 35 h per week for the United States, this would amount to an annual income gain of \$12,000. Using data from the New Immigrant Survey, Rosenzweig (2007) estimates an annual income gain to similarly educated legal Mexican immigrants of \$20,000.

One way to reconcile large and persistent cross-country income differences with small global labor movements is that receiving countries are successful in restricting labor inflows. While long queues for immigration visas in the United States and other countries do indicate that legal admission restrictions bind, rising levels of illegal immigration suggest that borders are porous. Further, observed costs of illegal entry are small in comparison to estimated income gains. In a sample of high migration-communities in Mexico during 2002–2004, Cornelius (2005) finds the average price paid by migrants for the service of being smuggled across the US border was \$1700.

Another explanation for small global labor flows is the existence of large unobserved migration costs associated with credit constraints in financing migration, uncertainty over economic opportunities abroad, the psychic cost of leaving home, or other factors. There is considerable academic interest in the role of migration networks in lowering such costs. Survey evidence suggests that transnational migration networks provide prospective migrants with information about economic conditions in destination countries, support in managing the immigration process, and help in obtaining housing and finding a job

(Massey & Espinosa, 1997; Massey, Goldring, & Durand, 1994). In the presence of network effects, labor outflows may accelerate over time (even as source-destination wage differences decline), due to a growing stock of migrants lowering moving costs for later migrants (Carrington, Detragiache, & Viswanath, 1996).

Much of the research on migration networks focuses on Mexico, for which there are individual-level data on migration behavior.²⁶ On the process of crossing the border, Orrenius and Zavodny (2005) report that among young males in Mexico the probability of migrating to the United States is higher for individuals whose fathers or siblings have emigrated. Gathmann (2004) documents that migrants with family members in the United States are less likely to hire the services of a professional smuggler, and, among those that do, likely to pay lower prices. And McKenzie and Rapoport (2006) find that average schooling is lower among migrants from communities in Mexico with a stronger US presence. These results are consistent with networks lowering migration costs.

One might be concerned that the presence of migration networks reflects unobserved characteristics of communities or families that are associated with a higher propensity to migrate, making the correlation between migration behavior and networks difficult to interpret. To address the issue of endogeneity in migration networks, Munshi (2003) instruments the size of the US population from a migrant's origin community in Mexico using lagged rainfall in the Mexican origin community. He finds that Mexican migrants in the United States are more likely to be employed and more likely to be employed in a higher-paying nonagricultural job than the larger US population of residents from their origin community in Mexico. These results suggest that having a larger network improves a migrant's ability to assimilate economically in the United States. Among nonagricultural workers, 78% received assistance in finding a US job, and among this group 47% received help from a relative and 47% received help from a friend or *paisano* (someone from their home region in Mexico).²⁷

While we still know little about the magnitude of migration costs, research on networks suggests that migrant flows are sensitive to changes in these costs. Other evidence on the sensitivity of migration to migration costs comes from illegal crossings at the Mexico-US border. For illegal migration, the intensity of border enforcement is an important determinant of entry costs, which take the form of fees paid to smugglers. Cornelius (2005) reports that smuggler prices to enter the United States illegally increased by 37% between 1996-1998 and 2002-2004, which spans the period during which the United States stepped up border enforcement efforts in response to the terrorist attacks of 9/11/01. Gathmann (2004) examines the consequences of expanded border enforcement for migration.²⁸ She identifies the correlates of smuggler prices paid by migrants from Mexico to the United States and estimates the impact of smuggler prices on migrant demand for smuggler services.²⁹ The price a migrant pays to a smuggler is higher in years when border

enforcement is higher, but the elasticity of smuggler prices with respect to enforcement is small, in the range of 0.2–0.5. During the sample period, a one-standard-deviation increase in enforcement would have led to an increase in smuggler prices of less than \$40. The demand for smuggler services and the probability of choosing to migrate to the United States are both quite responsive to changes in coyote prices. However, given the small enforcement elasticity of coyote prices, the observed increase in US border enforcement over 1986–1998 (during which time United States spending on border enforcement increased by four times in real terms) appeared to reduce the average migration probability in Mexico by only 10%.

In many destination countries, migrants reinforce networks by forming home-town associations that help members of their home communities make the transition to living in a new location. By creating links between the destination country and a specific community in the source country, these associations may lower migration costs for individuals linked by kinship or birthplace to migrants living abroad. Of 218 home-town associations formed by Mexican immigrants enumerated in a 2002 survey in California, 87% were associated with one of the nine central and western states in Mexico that have dominated migration to the United States since the early twentieth century (Cano, 2004), indicating that migrant networks in Mexico are organized along regional lines.³⁰

Regional variation in migration networks creates regional variation in migration dynamics. McKenzie and Rapoport (2007) show that in Mexican communities with historically weak migration networks, moderately more wealthy individuals are more likely to migrate, though very high wealthy individuals are not. Migrants are thus drawn from the middle of the wealth distribution, meaning that migration increases inequality. In communities with strong migration networks, however, lower wealthy individuals can afford to migrate, so that in these locations migration lowers inequality.

2.5 Discussion

Over the last decade and a half, migration flows from developing to developed countries have been increasing. The phenomenon is just beginning to be understood, as cross-country data on international migration have only recently become available. Bilateral migration flows are negatively affected by migration costs, as captured by geographic or linguistic distance between countries, the absence of migration networks, or the stringency of border enforcement against illegal entry. Emigration rates are highest for developing countries at middle income levels and with higher population densities. In most developing countries, it is the more educated who have the highest likelihood of emigrating. The positive selection of emigrants in terms of schooling is due in part to differences in the reward to skill across countries. High average labor productivity in the United States and other destinations more than compensates for these countries having relatively low percentage returns to additional years of education, giving more educated individuals

(or individuals seeking higher education) a relatively strong incentive to leave poor countries. Emigrants appear to sort themselves across destinations according to income-earning possibilities, with the countries that have the highest reward to skill—measured as the level difference in wages between high- and low-skilled labor—attracting the most educated mix of immigrants. Little is known about the precise magnitude of migration costs, the impact of skilled emigration on the incentive to acquire education in sending countries, or how receiving-country immigration policies affect the scale or composition of international migration.

3. IMPACT OF EMIGRATION ON SENDING COUNTRIES

Emigration changes a country's supply of labor, skill mix, and exposure to the global economy. These effects may have important consequences for a sending country's aggregate output, structure of wages, fiscal accounts, and trade and investment flows, among other outcomes. In this section, I discuss recent empirical research on the impact of emigration on developing economies.

3.1 Labor markets and fiscal accounts

To organize the discussion, it is useful to have a framework that identifies the channels through which emigration affects the well-being of individuals in an economy. Consider a country that produces a single output from two labor inputs, skilled labor (indexed by h) and unskilled labor (indexed by l), each of which is paid its marginal product. I assume there are H identical skilled workers and L identical unskilled workers, and that the two types of labor are complements in production. Let $V(Y_i)$ be the indirect utility enjoyed by a worker of type i , which depends on after-tax income available for consumption, Y_i . In turn, after-tax income depends on the wage, W_i , the income-tax rate, t_i , and government transfers, G_i , such that

$$Y_i = W_i(1 - t_i) + G_i. \quad (8)$$

The change in welfare that results from the emigration of ΔH skilled workers for a *nonemigrating* worker of type i can be written as

$$\frac{\Delta V_i}{\Delta H} = V'_i \left[\frac{\partial W_i}{\partial H} (1 - t_i) + \frac{\partial G_i}{\partial H} - W_i \frac{\partial t_i}{\partial H} \right], \quad (9)$$

where V'_i is the marginal utility of income. The first term in brackets in Eq. (9) captures the change in a worker's labor earnings; the second two terms capture the change in the net fiscal transfer that a worker receives from the government.

For an unskilled worker, the emigration of skilled labor, a complementary input, lowers his marginal product ($\partial W_l / \partial H < 0$), reducing his labor income. Assuming the government budget is balanced, taxes and transfers in a country would need to adjust to compensate for the loss in the net fiscal contribution of the emigrating workers. If skilled workers make a positive net fiscal contribution (i.e., the tax structure is progressive), skilled emigration would tend to reduce transfers ($\partial G_i / \partial H < 0$) and increases tax rates ($\partial t_i / \partial H > 0$) for all workers, ensuring that the posttax income of unskilled labor falls. All else equal, skilled emigration would reduce the welfare of unskilled workers.

For a nonemigrating skilled worker, the emigration of skilled labor raises his marginal product ($\partial W_h / \partial H > 0$), increasing his labor income. Following the above logic, skilled emigration would tend to reduce transfers received and increase taxes paid by nonemigrating skilled workers. Taking the positive change in pretax income together with the negative change in net fiscal transfer, the impact of skilled emigration on the posttax income and welfare of skilled workers is ambiguous.

Most research on the labor-market impacts of emigration focuses on Mexico. Mishra (2007), applying the regression framework in Borjas (2003), examines the correlation between emigration to the United States and decadal changes in wages for cohorts in Mexico defined by their years of schooling and labor-market experience. She estimates that over the period 1970–2000, the elasticity of wages with respect to emigration in Mexico is 0.40, implying that a 10% reduction in labor supply due to emigration would raise wages by 4%. Using a similar approach, Aydemir and Borjas (2007) estimate a wage elasticity for emigration in Mexico of 0.56.³¹

Wage elasticities of this magnitude suggest that emigration has had a substantial impact on Mexico's wage structure. Based on her estimation results and the fact that between 1970 and 2000, 13% of Mexico's labor force emigrated to the United States, Mishra (2007) calculates that emigration has raised average wages in the country by 8%.³² Upward wage pressure has been strongest for young adults with above-average education levels (those with 9–15 years of schooling), who in the 1990s were the individuals most likely to emigrate (Chiquiar & Hanson, 2005). Increased labor flows between Mexico and the United States appear to be one factor contributing to labor-market integration between the two countries.³³

In response to changes in labor supply associated with emigration, one might expect the supply of capital in Mexico to adjust, with the country becoming less attractive to inward foreign direct investment. Alternatively, higher wages could erode Mexico's comparative advantage in labor-intensive industries, reducing the net exports of labor services embodied in goods. Either change would tend to offset the effects of emigration on wages in the country. Since the estimation approaches proposed by Mishra (2007) and Aydemir and Borjas (2007) are reduced forms, they

capture the wage impact of emigration, *net* of these, and other adjustments. Their results suggest that any response of capital accumulation or trade to emigration is too slow or too small to undo the wage consequences of labor outflows, at least over 10-year time intervals.³⁴ Such a finding is not all that surprising. Factor-price differences between the United States and Mexico create an incentive for trade in goods, north-to-south flows of capital, and south-to-north flows of labor. Despite dramatic reductions in barriers to trade and investment between the two countries during the last two decades, US-Mexico wage differences remain large. Since trade and investment are insufficient to equalize factor prices within North America, theory would predict that migration from Mexico to the United States would affect wages in both countries, consistent with the evidence.³⁵

In many sending countries, the propensity to emigrate varies greatly across sub-national regions. In Mexico, central and western states have long had the highest labor flows abroad. The literature attributes regional variation in emigration to the emergence of migration networks, which grew out of the hiring practices of US agriculture. In the early 1900s, US labor contractors utilized Mexico's railroad network to recruit workers in the country's interior (Cardoso, 1980). Communities close to rail lines have had the highest emigration rates in the country since at least the 1920s.³⁶ With the advent of large-scale emigration from Mexico in the 1980s and 1990s, the historically high-migration states have had relatively large labor outflows. Between 1990 and 2000, the cohort of men in their 20s born in high-migration states declined by 33.4 log points, while the number of similarly aged men born in low-migration states dropped by only 9.4 log points. Since mortality rates are relatively low for this age group, the relative decline in the number of young men from high-migration states (of 24 log points) is most likely due to emigration. Hanson (2007) finds that over this time period, wages in high-migration states rose by 6-9% relative to wages in low-migration states, controlling for regional shocks associated with globalization.³⁷

The Mexican emigration experience differs from other countries in terms of the absence of positive selection, the high fraction of those leaving who enter the destination country as illegal migrants, and the sheer scale of the exodus. The positive selection of emigrants in most source countries raises the prospect of important fiscal impacts from international migration. In countries with progressive income taxes, the loss of skilled emigrants could adversely affect public budgets through a loss of future tax contributions. These lost contributions are, in part, the returns to public investments in the education of emigrating workers, which, after emigration, accrue to destination countries.

While there is a large body of theoretical literature on the taxation of skilled emigration (e.g., Bhagwati & Hamada, 1974; Bhagwati & Wilson, 1989; Docquier & Rapoport, 2007), empirical research on the subject is sparse. One recent contribution is by Desai, Kapur, and McHale (2003), who examine the fiscal effects of brain drain

from India.³⁸ In 2000, individuals with tertiary education made up 60.5% of Indian emigrants but just 4.5% of India's total population. Between 1990 and 2000, the emigration rate for the tertiary educated rose from 2.8% to 4.3%, compared to an increase of just 0.3% to 0.4% for the population as a whole. Desai et al. examine Indian emigration to the United States, which in 2000 was host to approximately 64.6% of India's skilled emigrants (and 48.9% of all Indian emigrants). They begin by producing a counterfactual income series that gives emigrants the income they would have earned in India based on their observed characteristics and the returns to these characteristics in India (using a Mincer wage regression). On the tax side, they calculate income tax losses by running the counterfactual income series through the Indian income tax schedule and indirect tax losses using estimates of indirect tax payments per unit of gross national income. On the spending side, they calculate expenditure savings by identifying categories for which savings would exist—which are most categories except interest payments and national defense—and then estimating savings per individual. The results suggest Indian emigration to the United States cost India net tax contributions of 0.24% of GDP in 2000, which are partially offset by the tax take on remittances of 0.1% of GDP. For India, the tax consequences of skilled emigration appear to be modest. Even tripling tertiary emigration, which would bring India in line with Mexico's emigration rate, labor outflows would still appear to have a small impact on the country's fiscal accounts.

The research discussed so far addresses the static consequences of emigration for an economy, ignoring dynamic considerations that may arise if skilled emigration raises the incentive of unskilled workers to acquire human capital. In theory, feedback effects from emigration to human-capital accumulation may change a country's rate of economic growth. Mountford (1997) shows that in the presence of human-capital externalities, an emigration-induced increase in the incentive to acquire skill can help an economy escape a poverty trap, characterized by low investment in education and low growth, and move to an equilibrium with high investment and high growth.³⁹ Yet, it is entirely possible for feedback effects to work in the opposite direction. Miyagiwa (1991) develops a model in which, because of human capital spillovers, the migration of skilled labor from a low-wage, skill-scarce economy to a high-wage, skill-abundant economy reinforces the incentive for brain drain, depleting the low-wage country of skilled labor. In Wong and Yip (1999), the negative effects of brain drain on the stock of human capital reduce the labor-exporting country's growth rate.

Given that plausible theoretical models offer very different predictions for the long-run consequences of skilled emigration, the effect of brain drain on an economy is ultimately an empirical question. As mentioned in Section 2.3, the literature on how emigration affects the incentive to acquire skill has yet to produce conclusive results, making it impossible to say whether the consequences of brain drain for

growth are likely to be positive or negative. Case-study evidence is similarly inconclusive. In China, India, and Taiwan, the migration of skilled labor to Silicon Valley in the United States—where Indian and Chinese immigrants account for one third of the engineering labor force—has been followed by increased trade with and investment from the United States, helping foster the creation of local high-technology industries (Saxenian, 2002). The recent rise in educational attainment in China, India, and Taiwan may be partly a result of the lure of working in the United States and the domestic expansion of sectors intensive in the use of skilled technicians.⁴⁰ In Africa, however, the exodus of skilled professionals, many of whom work in health care, may adversely affect living standards. Clemens (2007) reports that in 25 out of 53 African nations at least 40% of native-born individuals practicing as physicians were living and working abroad as of 2000.⁴¹ He finds a weak negative correlation between child mortality and the share of the stock physicians (or nurses) that has emigrated. Schiff (2006) offers further evidence that suggest pessimism about the prospects for a beneficial brain drain.

3.2 Remittances and return migration

In a static setting, where the only effect of international migration is to move labor from one country to another, welfare in the sending country would decline (Hamilton & Whalley, 1984). While the average incomes of migrants and destination-country natives would rise, average income (measured in terms of per capita GDP) in the sending country would fall (even though wages for labor would rise). Migrants, however, often remit a portion of their income to family members at home, possibly reversing the income loss in the sending country associated with the depletion of labor. In the last several years, there has been substantial academic and policy interest in the consequences of remittances for economic activity in sending countries.

Table 5 shows workers' remittances received from abroad as a share of GDP by geographic region. Remittances have increased markedly in East Asia and the Pacific, Latin America and the Caribbean, South Asia, and Sub-Saharan Africa. As of 2004, remittances exceeded official development assistance in all regions except Sub-Saharan Africa and were greater than 65% of foreign direct investment inflows in all regions except Europe and Central Asia. Among the smaller countries of Central America, the Caribbean, and the South Pacific, remittances account for a large share of national income, ranging from 10% to 17% of GDP in the Dominican Republic, Guatemala, El Salvador, Honduras, Jamaica, and Nicaragua, and representing an astounding 53% of GDP in Haiti (Acosta, Fajnzylber, & Lopez, 2007).

Reported remittances reflect those captured by the balance of payments, which Freund and Spatafora (2007) suggest may substantially understate the actual remittances. Formal remittance channels include banks and money transfer operators

Table 5 Workers' remittances and compensation of employees as % of GDP

Region	1990	1992	1994	1996	1998	2000	2001	2002	2003	2004	2005
East Asia and Pacific	0.50	0.56	0.66	0.71	0.93	1.00	1.10	1.47	1.56	1.48	1.50
Europe and Central Asia	—	—	1.17	1.02	1.45	1.42	1.31	1.27	1.24	1.28	1.44
Latin America and Caribbean	0.61	0.70	0.73	0.79	0.84	1.04	1.29	1.67	1.99	2.06	1.98
Middle East and North Africa	—	8.31	5.57	3.69	3.68	3.07	3.40	3.76	4.35	4.31	4.13
South Asia	1.41	1.76	2.24	2.42	2.47	2.85	3.10	3.72	4.09	3.57	3.53
Sub-Saharan Africa	0.72	0.76	0.94	1.04	1.47	1.49	1.55	1.67	1.49	1.60	1.57

(e.g., Western Union) for which service fees average 11% of the value of remittances. Informal remittances, which are moved by couriers, relatives, or migrants themselves, tend to have lower fees, but (presumably) higher risk. Formal remittances are negatively correlated with service charges, with a 10% increase in fees being associated with a 1.5% reduction in transfers. Fees are lower in economies that are dollarized and more developed financially (as measured by the ratio of bank deposits to GDP).

Theoretical literature on migration models remittances as the outcome of a dynamic contract between migrants and their families (e.g., Lucas & Stark, 1985). A family helps finance migration costs for one of its members in return for a share of future income gains associated with having moved to a higher wage location. Remittances are the return on investments the family has made in the migrant. The prediction is that remittances would rise following an increase in emigration and decline as existing emigrants age and pay off debts to their families. Obviously, emigration also means a loss in labor supply for the household in the sending country and may result in the separation of parents from children (particularly, in the case of temporary or guest worker migration), issues that are often left unexplored in empirical work.

Having migrants abroad may also provide insurance for a family. To the extent income shocks are imperfectly correlated across countries, migration helps families smooth consumption over time by keeping remittances high when sending-country income is low relative to the destination country, and low when sending-country income is relatively

high.⁴² Yang (2008) examines changes in remittances to households in the Philippines before and after the Asian financial crisis, which he uses as a natural experiment to examine the impact of remittances on household behavior. As of 1997, 6% of Philippine households had a member that had migrated abroad. Some had gone to countries in the Middle East, whose currencies appreciated sharply against the Philippine peso in 1997–1998, while others had gone to countries in East Asia, whose currencies appreciated less sharply or even depreciated. Consistent with consumption smoothing, remittances increased more for households whose migrants resided in countries that experienced stronger currency appreciation against the peso. Since income shocks associated with movements in exchange rates are largely transitory in nature, the response of remittances reveals the extent to which migrants share transitory income gains with family members at home. Yang finds that a 10% depreciation of the Philippine peso is associated with a 6% increase in remittances.

Contrary to Yang's results, remittances appear to be unresponsive to changes in government transfers. In Mexico (Teruel & Davis, 2000), Honduras and Nicaragua (Nielsen & Olinto, 2007) remittances are uncorrelated with changes in rural household receipts from conditional cash transfer programs, which were introduced into communities on a randomized basis, permitting the experimental analysis of their impact on household behavior. Were remittances a vehicle for consumption smoothing among rural households, one would expect them to decline for a sending-country household, following an exogenous increase in government income support. One possible difference between the Philippine migrants in Yang's sample and the Mexican and Central American migrants in Teruel and Davis' and Olinto's samples is that the large majority of Philippine migrants (95.6%) report they have gone abroad on temporary employment visas, meaning they are likely to return to the Philippines in the near future. Though the majority of migrants from Mexico and Central America may have gone abroad initially on an unauthorized basis (Hanson, 2006), many appear to remain in their destination country for a long period of time. These results suggest consumption smoothing may be more pronounced among temporary migrants.

There is some evidence that increases in remittances are associated with increased expenditure on education and health. Alejandra and Ureta (2003) find that in El Salvador households that receive remittances are more likely to allow children to stay in school, with the effect being stronger in rural areas. Why should remittances be correlated with school attendance? One possibility is that remittances allow credit-constrained households to increase investments in productive activities that capital-market imperfections prevent them from financing through borrowing. However, an equally plausible explanation is that households that receive remittances are less credit constrained to begin with and hence, are more likely to invest in education, suggesting that the correlation between remittances and educational investments may be the byproduct of their correlation with some omitted variable, such as unobserved wealth.

To identify the impact of remittances on education, Yang (2008) examines changes in household expenditure and labor supply in the Philippines before and after the Asian financial crisis. Households with migrants in countries experiencing stronger currency appreciation vis-à-vis the peso had larger increases in spending on child education, spending on durable goods (televisions and motor vehicles), children's school attendance, and entrepreneurship. In these households, the labor supply of 10–17 year old children fell by more, particularly for boys. Woodruff and Zenteno (2007) also find a positive correlation between migration and sending-country business formation, in their study for Mexico. For a sample of small-scale enterprises, capital investment and capital-output ratios are higher in firms where the owner was born in a state with higher rates of migration to the United States. Woodruff and Zenteno instrument for current migration rates using proximity to the railroads along which Mexico's initial migration networks became established (Durand, Massey, & Zenteno, 2001). Their results are consistent with two different mechanisms for business formation: remittances relax credit constraints on the creation of small enterprises, or return migrants—who may have accumulated valuable work experience in the United States—are more likely to launch new businesses upon returning to Mexico. Regarding the second mechanism, Dustmann and Kirchamp (2002) find that half of migrants returning to Turkey from Germany start a small business within 4 years of coming home, using labor income saved during their time as migrant workers.⁴³

Remittances indicate that migrants maintain contacts with family members at home. They may do so in part because they anticipate returning home in the future, in which case return migration may depend on their foreign earning opportunities. Yang (2006) finds that an exchange rate shock that raises the peso value of foreign earnings reduces the likelihood a Philippine emigrant returns home, with 10% real appreciation being associated with a 1-year return rate that is 1.4% lower.⁴⁴

The use of a clear empirical identification strategy in Yang (2006, 2008) and Woodruff and Zenteno (2007) is important, given the obvious concern that remittances and household expenditures are jointly determined. Many recent papers report a positive correlation between remittances and household spending on education, household spending on health, children's survival rates, or the likelihood a household is above the poverty line, among other outcomes.⁴⁵ With the absence of a natural experiment or valid instrument for remittances, such correlations are difficult to interpret. Less credit constrained households may be more likely to send migrants abroad *and* to invest in durable goods or services (Assuncao & Carvalho, 2007). Remittances are the return to households from having invested in sending a migrant abroad. Presumably, households invest in migration for the purpose of enjoying higher spending in the future, meaning remittances are evidence that a dynamic household contract has been fulfilled, not an independent causal force. One would hope that the recent enthusiasm among international financial institutions on the role

of remittances in economic development (e.g., Inter-American Development Bank, 2004) does not lead policy makers to ignore the economics of migration in recommending policies related to labor outflows.

3.3 Information and the flow of ideas

The emigration of labor creates linkages between a country and the rest of the world, which may help reduce international transaction costs. Casella and Rauch (2002) develop a model in which membership in a group—such as common ancestry or ethnicity—helps individuals in different countries reduce barriers to international trade associated with incomplete information. Relative to purely anonymous trade, the presence of group ties increases the volume of trade and GDP in the trading countries, though individuals lacking group ties are worse off (because they lose access to their more productive potential trading partners). Migration is an obvious mechanism through which cross-national group ties may be established.

The positive correlation between bilateral trade and migration has been interpreted as evidence of a “diaspora externality,” in which previous waves of migration create cross-national networks that facilitate exchange. Gould (1994) finds that the bilateral trade involving the United States is larger with countries that have larger immigrant populations in the United States. Head, Ries, and Swenson (1998) find that a 10% increase in Canada’s immigrant population from a particular country is associated with a 1% increase in bilateral Canadian exports and a 3% increase in bilateral Canadian imports, with more recent immigration having a stronger correlation with trade.⁴⁶ It is difficult to draw causal inferences from these results, since immigration may be correlated with unobserved factors that also affect trade, such as the trading partners’ cultural similarity or bilateral economic policies (e.g., preferential trade policies or investment treaties that raise the return to both migration and trade).

Pushing the analysis a step further, Rauch and Trindade (2002) focus specifically on networks associated with overseas Chinese populations. Successive waves of emigration from southeastern China have created communities of ethnic Chinese throughout Southeast Asia, as well as in South Asia and on the east coast of Africa. Rauch and Trindade find that bilateral trade is positively correlated with the interaction between the two countries’ Chinese populations (expressed as shares of the national population), similar to the findings by Gould and Head and Ries. More interestingly, the correlation between Chinese populations and trade is stronger for differentiated products than it is for homogenous goods. To the extent differentiated products are more subject to informational problems in exchange (Rauch, 1999), these are the goods one would expect to be most sensitive to the presence of business networks.

Still unclear is whether greater trade is the outcome of increased migration or a reflection of the types of individuals who select into migration. If more skilled and more able individuals are more likely to migrate abroad and more likely to exploit

opportunities for commercial exchange, then the correlation between trade and migration may be a byproduct of migrant self-selection. Subsequent policies to liberalize immigration in destination countries would not necessarily increase trade with sending countries, unless they allowed for the admission of individuals with a propensity to engage in trade. Head et al. (1998) find that immigrants admitted as refugees or on the basis of family ties with Canadian residents have a smaller effect on trade than immigrants admitted under a point system that values labor-market skills.

More controversial is the impact of emigration on political outcomes in sending countries. When individuals live and work in another country they are exposed to new political ideologies and alternative systems of government. This exposure may be most important for students who go abroad to obtain a university degree, as they are at an impressionable age and often travel on visas that require them to return to home after completing their studies. The US government, in part, justifies the Fulbright Program, through which it has funded 160,000 foreign students to study in the United States over the last several decades, on its contribution to spreading democracy abroad. Spilimbergo (2006) suggests there is an association between a country's democratic tendencies and the political systems of the countries under which its students did their university training. He finds a positive correlation between the democracy index in a sending country and the average democracy index in the countries in which a country's emigrant students are studying (lagged 5 years). Whether the political system of a sending country influences the types of countries in which its students choose to study is unknown. Kim (1998), for instance, finds that the bilateral flow of foreign students is larger between countries that share a common religion.

3.4 Discussion

In the short run, economic theory suggests that the exodus of labor from a country would put an upward pressure on wages. Evidence from Mexico indicates that emigration has increased wages for the skill groups and regions with the highest emigration rates. Still unknown is the extent to which trade and capital accumulation offset the labor-market impacts of emigration. The preponderance of relatively highly educated individuals among emigrants suggests labor outflows may have adverse consequences on sending countries' public finances. However, in the case of India the fiscal effects of skilled emigration appear to be small. Evidence for other countries is lacking.

In the last decade, a new theoretical literature has emerged which takes a more sanguine view of brain drain. While the idea that skilled emigration raises the incentive to acquire skill in a country is plausible, the literature is missing well identified econometric estimates of how human capital accumulation and economic growth respond to labor outflows. We do know that in most countries those emigrating tend to be more educated individuals, who are in relatively scarce supply. Standard economic models would suggest that their departure adversely affects the livelihoods of the poor majority in developing countries (Benhabib & Jovanovic, 2007). At least for now, there are no

compelling data to suggest this view to be overturned. There is some evidence that emigration may promote a country's foreign trade and democratic leanings.

While the outflow of labor associated with emigration reduces a country's GDP, migrant remittances may offset the loss in income. In Mexico, Mishra (2007) finds that remittances are larger than the reduction in GDP due to emigration. In some countries, data suggest households use remittances to raise spending, increase investment in business ventures and education, and smooth consumption over time. While remittances are positively correlated with many indicators of economic development, there are only a handful of studies in which this correlation has a meaningful econometric interpretation. Complicating inference about the development impacts of remittances is the fact that less credit constrained households are those most likely to send migrants abroad in the first place. Concluding that remittances *cause* these households to have higher spending, higher investment, or improved health outcomes for women and children may confound the effects of emigration with the effects of unobserved wealth that make emigration possible. Finding that remittances improve the livelihoods of the poor is certainly more exciting than saying wealthier households are more likely to enjoy higher standards of living, but it is not a result for which there is yet broad empirical support.

4. IMMIGRATION POLICY REGIMES

Distinct from other aspects of globalization, the policies that govern international migration are largely under the control of labor-importing countries. The closest source and destination countries have come to negotiating a multilateral deal on migration are discussions under Mode IV of the Doha Development Agenda of the World Trade Organization, which if adopted would permit the temporary movement of service providers across borders, addressing a limited set of international labor flows.⁴⁷ There is little meaningful dialogue between countries regarding the scale or composition of migrant flows, meaning that labor-importing countries set their immigration policies unconstrained by international agreements. By varying the restrictiveness of their admission policies, developed countries directly affect the livelihood of individuals in developing countries. To understand how international migration is regulated, one must examine how destination countries choose the number of immigrants to admit, the types of immigrants to admit, and the rights to grant these individuals.

4.1 Political economy of immigration policy

Why do countries restrict immigration? Without distortions, the first-best policy for a labor-importing country would be to have open borders. Yet, most developed countries are far from such a policy. The distributional impacts of immigration may have political consequences, which give politicians an incentive to restrict labor inflows from abroad. In an economy without distortions, those hurt by immigration would

include native workers that substitute for immigrant labor. In an economy with a progressive tax system and redistributive government transfers, some native taxpayers would also be hurt. In choosing an immigration policy, a government trades off political support from special interests against consumer welfare (which is enhanced by openness). In a context where the median voter's wages would be reduced by immigration, politicians may choose to restrict labor inflows in order to enhance their future electoral prospects (Benhabib, 1996; De Melo, Grether, & Müller, 2001).⁴⁸

Alternatively, active lobbying by special interests may influence immigration policy. Facchini and Willmann (2005) extend the Grossman-Helpman model of the political economy of trade policy to consider international factor mobility.⁴⁹ In their setup, governments restrict factor inflows from abroad through a per-factor unit tax or quota. They assume that the receiving-country government captures factor tax revenues or quota rents, and that individuals are organized according to their factor type and lobby the government on immigration policy. The first assumption appears to be counterfactual, as few governments collect significant payments from factor inflows. The second assumption has more empirical support. In the United States, periodic attempts to increase enforcement against illegal immigration are met with political opposition (Hanson & Spilimbergo, 2001). In equilibrium, each factor lobby offers the government campaign contributions to support stronger (weaker) restrictions on inflows of factors for which its members substitute (complement) in production.

For politicians to respond to pressure from voters regarding immigration policy, voters in destination countries must perceive that immigration affects their standard of living. In the United States, Scheve and Slaughter (2001a) find that opposition to immigration is stronger among less educated workers, which appear to be the group most hurt by labor inflows from abroad (Borjas, 2003). The opposition of the less educated is greater in regions where immigrant inflows have been larger. Less-skilled labor's skepticism about immigration mirrors its opposition to globalization more generally (Scheve & Slaughter, 2001b). Mayda (2006) obtains similar results for a cross-section of countries. In economies where immigrants are less skilled than natives, opposition to immigration is stronger among less-skilled residents.⁵⁰

Tax and transfer policies create a second motivation for a government to restrict immigration, even where the level of immigration is set by a social planner. If immigrants are primarily individuals with low income relative to natives (which may be true even if migrants are high skilled relative to nonmigrants in the source country), increased labor inflows may exacerbate distortions created by social-insurance programs or means-tested entitlement programs (Wellisch & Walz, 1998). Such policies may make a departure from free immigration the constrained social optimum.⁵¹

In the United States, the fiscal consequences of immigration appear to matter for immigration policy preferences. Low-skilled immigrants—who account for one-third of the US foreign-born population—tend to earn relatively low wages, pay relatively

little in taxes, and receive subsidized health care with relatively high frequency (Borjas & Hilton, 1996; Fix & Passel, 2002). Hanson, Scheve, and Slaughter (2007) find that US natives who are more exposed to immigrant fiscal pressures—are those living in states that have large immigrant populations and that provide immigrants access to generous public benefits—are more in favor of reducing immigration. This public-finance cleavage is strongest among natives with high earnings potential, who tend to be in higher tax brackets. Facchini and Mayda (2006) obtain similar results for Europe, where immigrants also appear to be a fiscal drain (Sinn, and Hans-Werner, 2004). More educated individuals, who are also likely to be high income earners, are more opposed to immigration in countries where immigrants are less skilled and governments are more generous in the benefits they provide.

Pay as you go pension systems create a further incentive for politicians to manipulate the timing and level of immigration (Poutvaara, 2005; Razin & Sadka, 1999; Scholten & Thum, 1996). Governments may choose to permit immigration of young workers, in order to smooth adjustment to demographic shocks, such as the aging of the baby boom generation (Auerbach & Oreopoulos, 1999; Storesletten, 2000). Given its graying population and unfunded pension liabilities, one might expect Europe to be opening itself more aggressively to foreign labor inflows (Boeri, McCormick, & Hanson, 2002). However, concerns over possible increases in expenditure on social insurance programs may temper the region's enthusiasm for using immigration to solve its pension problems (Boeri & Brücker, 2005; De Giorgi & Pellizzari, 2006).

Beyond the economic consequences of labor inflows, some argue that opposition to immigration is grounded in culture, with individuals preferring homogenous societies because they foster a stronger sense of national identity and civic purpose (Huntington, 2004). Consistent with this claim, the recent anti-immigration-based presidential campaigns of Pauline Hanson in Australia, Jean Marie Le Pen in France, and Tom Tancredo in the United States tout the negative cultural effects of foreign labor inflows. Using individual survey data, Dustmann and Preston (2004) suggest racist attitudes are an important component of opposition to immigration in the United Kingdom and Hainmueller and Hiscox (2004) claim that greater tolerance for immigration among the college educated reflects cosmopolitan attitudes rather than economic concerns.

4.2 The design of immigration policy regimes

Much of the academic literature treats immigration policy as though it were governed by a single instrument: the level of admissions. Yet, in practice policy makers use multiple instruments to manage entry from abroad.

Countries regulate legal immigration through a combination of numerical quotas, entry selection criteria, and restrictions on residency rights. While many countries have admission categories that allow unrestricted immigration, these are generally limited to immediate family members of citizens, as in the United States, or individuals from

countries within an economic union, as in the EU. Other legal immigrants are subject to quotas, whose number varies according to a nation's *ex ante* selection criteria. The United States allocates the majority of permanent residence visas to relatives of US citizens and legal residents; Australia and Canada favor legal immigrants that meet designated skill criteria; and many European countries reserve a large share of visas for refugees and asylees (OECD, 2006). Visas come with limited residency rights. Temporary visas specify a time limit for residence, the types of jobs a visa holder may hold, and the set of government benefits to which the holder has access. Permanent visas provide broader residency rights, such as mobility between employers and access to more government benefits, but do not always offer a clear path to citizenship.

Regarding illegal immigration, while countries do not explicitly set unauthorized labor inflows, they do implicitly determine the ease of illegal entry through their enforcement actions. By choosing the intensity with which they police national borders and monitor domestic worksites, governments influence the smuggling fee illegal immigrants pay to enter a country (Ethier, 1986; Gathmann, 2004). Enforcement also defines an *ex post* selection criterion for illegal immigrants: individuals who are able to evade capture by avoiding the police earn the right to stay in the country (Cox and Posner, 2007). The United States, for instance, concentrates enforcement on borders rather than in the interior, allowing most illegal immigrants who do not commit crimes or maintain a high public profile to remain on US soil (Davila, Pagan, & Grau, 1999). While illegal immigrants lack official residency rights, they are not devoid of legal protections. Again in the United States, illegal immigrants may report crimes, attend public schools, seek emergency medical services, obtain bank loans, or even acquire a driver's license, with minimal risk of deportation.

Cross-country differences in policy regimes do not affect the skill mix of immigrants as much as one might think. Antecol, Cobb-Clark, and Trejo (2003) find that excluding immigrants from Latin America—who benefit from close proximity to the United States—the education, English fluency, and income of immigrants in Australia, Canada, and the United States are relatively similar. This is true despite Australia's and Canada's use of a point system that favors skilled immigrants and the US reliance on family reunification, which takes no account of skill, for the majority of its admissions. Comparing immigrants admitted on employment-based visas in Australia and the United States, Jasso and Rosenzweig (2005) suggest that it is self-selection, rather than national screening mechanisms, which accounts for differences in immigrant skills.

Even with similarities between countries, there are differences within countries in how legal and illegal inflows are regulated. As discussed above, authorized entrants tend to be subject to quantity regulation and *ex ante* selection criteria and have either expansive residency rights (for permanent immigrants) or limited residency rights (for temporary immigrants); and unauthorized entrants tend to be subject to price regulation and *ex post* selection criteria and have minimal residency rights.

Why do countries permit both legal and illegal immigration? First, consider legal inflows. Quantity regulation allows a country to achieve specific goals in admissions, by assigning quotas to particular categories. The allocation of quotas may reflect a desire to maximize the immigration surplus (by admitting scarce labor types), political economy constraints on the level and type of immigrant inflows, or other objectives of government (e.g., national security, cultural homogeneity, humanitarian concerns). An *ex ante* screen has a cost in that the government foregoes the option to obtain information on an immigrant beyond observable characteristics, before offering admission (Cox and Posner, 2007). However, the cost of foregone information may be small for skilled immigrants whose abilities are verifiable in the form of educational degrees, professional awards, and past employment positions. The effective information cost may also be small where countries have strong preferences for specific types of entrants (e.g., family members), in which case any updating on immigrant quality after residence in the country would be unlikely to alter the admission decision.

Combining an *ex ante* screen with broad residency rights gives immigrants a strong incentive to assimilate. However, broad rights have a high fiscal cost, since they give immigrants access to government benefits. The cost of providing broad rights may be small for skilled immigrants, whose income-earning ability would make them net contributors to government coffers. For family-based immigrants, the perceived cost of broad rights may also be small since, as family members of residents, their well being may be an implicit component of national welfare. For refugees and asylees, a similar logic would not apply, perhaps accounting for why they tend to have narrow residency rights (Åslund, Edin, & Fredriksson, 2001; Hatton & Williamson, 2004).

Quotas do not imply as much inflexibility in immigration levels as it would seem, since countries often admit a mix of permanent and temporary entrants. Opponents to immigration may be unwilling to allow all entrants to be permanent. Temporary immigration quotas give politicians the power to rescind visas in the future, which may increase support for immigration. The cost of having temporary immigrants is a weak incentive to assimilate. Comparing the costs and benefits, we might expect the share of temporary immigrants in legal admissions to be higher when an economy is closer to a business cycle peak, at which point the option value of being able to expel current entrants in the future may be relatively high.

Constitutional rules governing citizenship may constrain legal immigration policy regimes. Countries allow individuals to acquire citizenship by birth, naturalization, or marriage. Under the *jus soli* principle, which is rooted in both civil and common law traditions, citizenship is acquired by place of birth, implying that the native-born child of an immigrant is a citizen. Under the *jus sanguinis* principle, citizenship is acquired by descent, such that the child of a citizen is also a citizen, regardless of birthplace. Current citizenship laws often embody both principles, though they tend to have emerged out of one tradition or the other. *Jus soli* was predominant in Europe through the eighteenth

century, given feudal traditions linking citizenship to land. The French adopted *jus sanguinis* in the early nineteenth century, which then spread throughout continental Europe and its colonies. The United Kingdom, however, preserved *jus soli*, which was adopted by the United States, Canada, and Australia (Bertocchi & Strozzi, 2006a, 2006b). Under a *jus sanguinis* tradition, a country may have difficulty in granting broad residency rights to immigrants whose parents were not citizens, as appears to be the case in France.

Source country policies may also affect which immigrants become naturalized in destination countries. During the 1990s, Brazil, Colombia, Costa Rica, the Dominican Republic, and Ecuador each enacted laws permitting dual citizenship. Mazzolari (2006) finds that between 1990 and 2000 US naturalization rates for eligible immigrants from these countries increased relative to immigrants from other countries, suggesting that not having to give up citizenship in the source may speed assimilation in the destination.

For illegal immigration, entry prices and selection criteria are defined implicitly through the intensity of border and interior enforcement (Either, 1986). Entry prices serve as selection device, since an individual must value migration to be willing to incur the cost of paying a smuggler. Entry fees thus select immigrants with relatively large perceived income gains (Orrenius & Zavodny, 2005), which would include those for whom immigration would yield large gains in either pretax income (due to a productivity gain from immigration) or posttax income (due to tax and transfer policies in the destination). While most destination countries would prefer to attract the first type of immigrant over the second, an entry price does not select between the two.

One way to encourage immigration of more productive illegal immigrants is through granting narrow residency rights. For instance, since 1996 noncitizens in the United States have been ineligible for most types of federally funded public assistance (Fix & Passel, 2002). A second way is through *ex post* screening. Interior enforcement helps screen illegal immigrants who commit crimes, try to obtain government benefits illicitly, or engage in other behavior deemed objectionable. Governments that choose *not* to monitor employers that hire illegal immigrants can ensure that illegals who come to work are able to remain in the country. In the United States, greater border enforcement does not appear to have strong deterrent effects on illegal entry (Davila, Pagan, & Soydemir, 2002) or to affect wages or employment in US border cities (Hanson, Robertson, & Spilimbergo, 2001), suggesting that the primary role of enforcement is not to disrupt US labor markets.

Combining price regulation, narrow residency rights, and an *ex post* screen helps countries attract productive and motivated illegal immigrants. This selection process may be particularly important for the low-skilled, whose observable characteristics may be uninformative about their productivity. In the United States, two-thirds of immigrants with less than a high school education appear to be in the country illegally (Passel, 2006), suggesting that the majority of the least skilled immigrants are unauthorized. Relative to similarly skilled natives, low-skill immigrants have high employment rates and low rates of participation in crime (Butcher & Piehl, 1998, 2006).

The United States and the EU have considered using expanded temporary immigration to absorb their illegal immigrant populations (Schiff, 2007; Walmsley & Winters, 2005). Large scale illegal entry in the United States began after the end of the Bracero Program (1942–1964), which admitted large numbers of seasonal laborers from Mexico and the Caribbean to work on US farms (Calavita, 1992). Could new guest worker programs end illegal inflows? Recent literature suggests that unless interior enforcement is highly effective at preventing employers from hiring illegals, a guest worker program that rationed entry would not curtail the employment of unauthorized labor but simply push these workers deeper into the underground economy (Djajic, 1999; Epstein, Hillman, & Weiss, 1999; Epstein & Weiss, 2001).

4.3 Discussion

In a neoclassical economy, the optimal immigration policy would be to allow the unfettered entry of labor from abroad. Yet, labor-importing countries tightly restrict labor inflows. Barriers to immigration in part reflect domestic political opposition to open borders, with those most opposed to labor inflows being the workers and taxpayers who are most exposed to the adverse consequences of immigration on labor markets and fiscal accounts. Immigration barriers may also represent a second-best policy that governments adopt in order not to exacerbate distortions associated with domestic social-insurance programs that they are unwilling to dismantle.

The structure of immigration policy regimes suggests that destination countries also use barriers to identify individuals who appear likely to be productive workers and/or have the desire to assimilate. Reserving immigration visas for skilled workers selects high ability foreigners in a transparent manner. Restricting the residency rights of immigrants helps screen those whose primary interest is in enjoying rich-country welfare benefits. Less transparently, barriers to illegal immigration also select the more productive and more motivated workers among the low-skilled, whose ability is hard to observe. The existence of informational problems in evaluating immigrants' abilities and motivations suggests that there may be gains from coordination between labor-exporting and labor-importing countries. Were labor-importing countries to have access to better information on the employment histories of low-skilled individuals in developing countries, they might be willing to accept them in larger numbers and require fewer of them to enter their economies as illegal immigrants.

For sending countries, the vast majority of which do not restrict emigration, the most relevant policies regarding labor outflows may pertain to education. Increasing secondary and tertiary educational opportunities in low income countries may increase the likelihood that its citizens will succeed in migrating abroad. If migration is temporary, such as to obtain an education or to complete a guest worker contract, the return to the poor country may be positive. The increase in earnings (and foreign trade and investment) created by emigration may more than compensate for the cost of

schooling. However, Rosenzweig (2006) suggests that emigration for the purpose of education from countries with very low labor productivity is unlikely to be temporary. Foreign students from countries with low skill prices are those most likely to remain in the United States following their schooling. In these contexts, subsidizing education may be tantamount to subsidizing permanent emigration. Of course, countries may choose specific educational programs that encourage the return of migrants or improve their chance of landing a guest worker visa (as the Philippines has done), which may yield a positive economic return. For a more detailed discussion of education policies in developing countries, see the chapter in this volume by Jere Behrman.

5. FINAL DISCUSSION

Despite recent advances in the theoretical and empirical analysis of international migration, there is still a great deal that we do not know about global labor movements. This is in part due to the lack of data. Only recently has information on the global stock of emigrants become available. Much of the individual level data on international migration covers Mexico and/or the United States, which are the subject of a large literature. As the largest sending and receiving country, there is still more to learn about the Mexico-US context. Yet, the highest payoff to research is likely to be in the many under-studied parts of the world. Since 1990, Central and Eastern Europe have become major sending regions; the Gulf States, Russia, and Spain have become an important receiving regions; and emigration from China, India, Indonesia, Pakistan, and the Philippines have accelerated, to name but a few of the recent developments in global labor flows. Women now account for a growing share of international migrants and migrants in general appear to have a presence in destination countries that is more permanent than in the past. None of these events is well understood. Combining data from population censuses in sending and receiving countries is one way to amass a large quantity of information on international migration from existing data sources, a strategy put to use recently but that is far from being fully exploited. Collecting panel data on the behavior of actual and potential migrants, which is a more costly but ultimately more illuminating approach, is essential for research on international migration to progress.

Given the magnitude of international wage differences, global migration is not as large as one would expect. We know little about the magnitude of international migration costs. What is the relative importance of uncertainty, credit constraints, and destination-country admission policies in keeping the poor from migrating to rich economies? Existing research is silent on this issue. While there is growing evidence that migration networks play an important role in reducing moving costs, the empirical dynamics of networks are poorly understood. Are there diminishing returns in the impact of network size on migration costs? Or does the existence of networks imply that spatial opportunities for emigration will only become more unequal over time?

There is abundant evidence that the more educated have the highest propensity to emigrate. While theoretical literature on brain drain is well developed, empirical work is limited. Given the importance of human capital in economic development, how skilled emigration affects a country's relative supply of skill is a question of first-order policy importance. As economists, we simply do not know what to tell developing countries about how changes in their education, tax, or other policies have affected skilled emigration, the domestic supply of skill, or remittances from skilled emigrants.

Over 10 year intervals, there is a positive correlation between emigration and wage changes, suggesting that labor outflows tend to put an upward pressure on wages. Largely unknown, at least empirically, is how emigration interacts with international trade or foreign direct investment. It appears that sending and receiving countries are still far from having equal factor prices, in which case we might expect to see trade, migration, and FDI to happen concurrently, even reinforcing one another. The literature provides insufficient guidance to developing countries about how the various mechanisms for globalization interact in different settings.

The inflow of remittances has been a welcome financial boon for many labor-exporting countries. While remittances may help deepen domestic financial markets, as households use banks or other intermediaries to manage lumpy income receipts from abroad, there should be no presumption that the primary motivation for remittances is to finance new investment. As the return on previous household investments in migration, most remittances may end up supporting consumption.

Destination country restrictions on labor inflows leave the world far from a state of open borders. While quotas on legal immigration appear to bind in all or nearly all labor importing countries, opportunities for illegal immigration makes labor inflows substantially more flexible than *de jure* policy regimes would suggest. Government choices over the components of immigration policy reflect the political organization of groups that would be hurt by immigration, national preferences over who merits inclusion as a citizen, and a tradeoff between providing incentives for assimilation and obtaining information about the ability of immigrants.

Within the development policy community, there are calls for rich countries to open their economies more widely to labor inflows from poor countries (e.g., Pritchett, 2006). Completely open borders are off the table politically. Were the developed world to propose an increase in immigration quotas, should developing countries take the offer? The literature suggests the answer depends on how destination countries structured the additional labor inflows. An increase in immigration quotas that targeted workers with higher levels of skill (relative to nonemigrants in source countries) could raise global income, even as it lowered welfare for the less-skilled majority in source countries. It could also lower global welfare, under an egalitarian social welfare function that gives each individual equal weight (Benhabib & Jovanovic, 2007). While

quotas targeted to less-skilled workers could raise global welfare (though in the presence of human capital externalities they would not maximize global income), the adoption of such a policy appears unlikely given the political opposition in destination countries to labor inflows that worsen distortions associated with social insurance and related programs.

The only feasible way to generate political support in destination countries for increased low-skilled migration would seem to require (a) insulating destination countries from immigration's fiscal effects, (b) allowing destination countries to capture more of the gains from global migration (Freeman, 2006), or (c) helping destination countries overcome informational problems in selecting less-skilled immigrants who are likely to be productive workers (Cox and Posner, 2007). So far, few democratic destination countries have been willing to try approach (a) or (b); approach (c) remains untested. Without policy experimentation, illegal entry may remain the primary means through which low-skilled workers in poor countries are able to migrate to rich ones.

End Notes

1. The more sizable migration flows into non-OECD countries are from the former Soviet Republics to Russia; Bangladesh to India; Egypt, India, Pakistan, and the Philippines to the Gulf States; Afghanistan to Iran; Iraq to Syria; other South African states to South Africa; Indonesia to Malaysia; Malaysia to Singapore; Guatemala to Mexico; and Nicaragua to Costa Rica (Ratha & Shaw, 2007).
2. The UN Universal Declaration of Human Rights (1948) states that "Everyone has a right to leave any country, including his own" (UN, 2002).
3. Topics I will not address include the age of mass migration (see, e.g., Bertocchi & Strozzi, 2006a; Hatton & Williamson, 1998), the impact of migration on receiving countries (see, e.g., Borjas, 1999b), and the emigration of retirees from rich countries (see, e.g., Williams, King, & Warnes, 1997). The first two topics have received much attention elsewhere; the third has received too little attention to merit discussion.
4. See Borjas (1999b) for a survey of the literature on immigration in the United States and the volumes in Borjas (2007) on the specific impacts of Mexican immigration.
5. See Docquier and Marfouk (2006) for further discussion of data sources on international migration.
6. Primary indicates 0-8 years of schooling, secondary indicates 9-12 years of schooling, and tertiary indicates 13 or more years of schooling.
7. OECD members in 2000 were Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.
8. In the United States, the undercounting problem does not appear to be too severe, with the US Census Bureau estimating that it misses only 5-10% of illegal immigrants (Passel, 2006).
9. Unless otherwise noted, tables and figures are based on calculations using raw data from Docquier and Marfouk (2006). In these data, the adult population is individuals aged 25 years and older.
10. As recently as 1990, the United Kingdom was the largest source country for immigrants in the OECD.
11. Other evidence suggests that Spain has also seen a large recent increase in its immigrant population.

12. See Hanson (2006) for a review on the literature of illegal migration from Mexico.
13. The countries with the highest emigration rates to the OECD are Guyana (42.1%) and Suriname (47.4%).
14. Three clear outliers in this relationship are Guyana and Suriname, which have low population densities and high emigration rates, and Singapore, which has a high density and low emigration.
15. One limitation of this study is that “immigrants admitted” are measured by the number of individuals who receive a US legal permanent residence visa, or green card, in a given year. A substantial fraction of green-card recipients are individuals already residing in the United States, either on a temporary immigration visa or as unauthorized immigrants. Over the period 1991–1999, for instance, 50.3% of US green card recipients were current US residents (U.S. Immigration and Naturalization Service, 2000), implying there is slippage between measured and actual inflows of immigrant labor.
16. Similar distance elasticity estimates can be found in the study by Karemra, Onguledo, and Davis (2000), for bilateral migration to the United States and Canada in the 1970s and 1980s, and Pedersen, Pytlakova, and Smith (2004), for bilateral migration to OECD countries in the 1990s.
17. One concern about the estimation results in Clark et al. is that they include both source-country dummy variables and the lagged stock of emigrants in the United States as regressors. Since the current stock is the sum of past flows, the specification is close to having fixed effects and a lagged dependent variable, which creates concerns about the consistency of the coefficient estimates.
18. See, for example, Borjas (1987), Karemra et al. (2000), Volger and Rotte (2000), Hatton and Williamson (2002), Pedersen et al. (2004); and Gallardo-Sejas, Pareja, Llorca-Vivero, and Martinez-Serrano (2006).
19. Borjas (2005) reports that half of foreign doctoral students in the United States remain in the country.
20. See Docquier and Rapoport (2007) for a survey of the theoretical literature on brain drain.
21. See Beine, Docquier, and Rapoport (2001) for related work.
22. See Dahl (2002) on the application of the Roy model to internal migration.
23. Feliciano (2001), Hanson and Chiquiar (2005), Orrenius and Zavodny (2005), McKenzie and Rapoport (2006), Cuecuecha (2005), and Rubalcava et al. (2006) find that emigrants from Mexico are drawn from the middle of the wage or schooling distribution; Ibarra and Lubotsky (2005) and Fernandez-Huertas (2006) find that Mexican emigrants are drawn from the lower middle of the wage or schooling distribution. No study finds that emigration rates decrease monotonically in skill, as predicted by Borjas (1987).
24. In 2000, the tertiary educated were 47.4% of emigrants and 16.5% of the population in Chile and 39.5% of emigrants and 12.4% of the population in Poland (where the population is residents plus emigrants).
25. First, he regresses source-country log wages on individual age, individual schooling, and source-country fixed effects, where the fixed effects reflect the marginal product of labor (under the assumption that the coefficients on age and schooling are constant across countries). Second, he corrects the estimated fixed effects for selectivity into migration by regressing them on source country GDP and average schooling and the inverse Mills ratio from a blocked Probit regression of migration to the United States (which takes source-country determinants of migration costs as arguments).
26. For research on networks outside of Mexico, see Du, Park, and Wang (2005) on China and Bauer and Gang (1998) on Egypt.
27. For other work on migration networks in Mexico, see Winters, de Janvry, and Sadoulet (2001).
28. For related work, see Reyes, Johnson, and Van (2002) and Angelucci (2006).
29. In the estimation of coyote prices, Gathmann (2004) instruments for border enforcement using the drug budget of the US Drug Enforcement Agency (DEA). In the estimation of the demand for coyote services, she includes both the smuggler price and the level of border enforcement as regressors, instrumenting the former with the average US prison term for smugglers (which rises over the sample period) and for the latter again with the DEA drug budget.

30. See Lapointe and Orozco (2004) also on Mexican hometown associations in the United States and Caglar (2006) on Turkish hometown associations in Germany.
31. To interpret the wage elasticities, let labor demand be $\ln W = \eta \ln(L - M)$, where W is the wage, $\eta < 0$ is the factor price elasticity (i.e., the inverse elasticity of labor demand), L is the native employment, and M is the number of workers lost to emigration; and let domestic labor supply be given by $\ln W = \kappa \ln L$, where $\kappa > 0$ is the inverse elasticity of labor supply. The resulting reduced-form expression for wages is $\ln W = \eta \rho m$, where $m = M/L$ is the emigration rate and $\rho = (\eta/\kappa - 1)^{-1} < 0$. The elasticity of wages with respect to emigration is $\eta\rho > 0$. Since $|\rho| < 1$, the emigration wage elasticity understates the factor price elasticity.
32. Both Mishra (2007) and Aydemir and Borjas (2007) treat emigration (in schooling and labor-market experience cells) as exogenous (after controlling for schooling, experience, and year fixed effects, and their interactions). Any endogeneity of emigration to wage shocks in Mexico would tend to bias the estimated emigration wage elasticity toward zero (since emigration would be negatively correlated with wage shocks), suggesting these results may underestimate the impact of emigration on wages in Mexico.
33. On US-Mexico labor-market integration, see also Robertson (2000).
34. See Borjas, Grogger, and Hanson (2007) on how to estimate the wage impacts of migration when capital accumulation is endogenous.
35. See Borjas and Katz (2007) on the wage impact of Mexican immigration in the United States.
36. From the 1920s to the 1960s, the nine west-central states accounted for 44.0–56.1% of Mexican migration to the United States, but only 27.1–31.5% of Mexico's total population (Durand, Massey, & Zenteno, 2001).
37. Hanson's (2007) results imply the elasticity of wages with respect to emigration is 0.7–0.8. This elasticity reflects both the direct effects of emigration on the labor supply *and* any indirect effects of historical emigration patterns on current regional wage growth, which may account for it being larger in magnitude than the estimates in Mishra (2007) and Aydemir and Borjas (2007).
38. See also Desai, Kapur, and McHale (2004).
39. A further beneficial effect of emigration is that it may increase the incentive to invest in productive skills—which are likely to have a positive return abroad—over rent-seeking activities—which are likely to have a low return abroad (Mariani, 2007).
40. Between 1990 and 2000, the share of the adult resident population (i.e., net of brain drain) with a tertiary education rose from 2.0% to 2.7% in China, 4.1% to 4.8% in India, and 12.2% to 19.1% in Taiwan.
41. These data underestimate emigration rates for African physicians because (a) the emigrant stock is calculated for just nine destinations (Australia, Belgium, Canada, France, Portugal, South Africa, Spain, the United Kingdom, the United States), and (b) only individuals listing their current occupation as medical doctor are counted as physicians.
42. See Rosenzweig and Stark (1989) on internal migration and consumption smoothing.
43. See Mesnard (2004) on return migration and self-employment in Tunisia.
44. Also on return migration, see Borjas and Bratsberg (1996) and Lacuesta (2006).
45. For a discussion of work in this literature, see Ozden and Schiff (2006) and Fajnzlber and Lopez (2007).
46. Trade links associated with immigration tend not to apply in the case of refugee flows.
47. Migration under Mode IV would result from a contract between a buyer in an importing country and a supplier in an exporting country, in circumstances where consummation of trade requires the presence of the supplier's employees in the buyer's location (e.g., trade in architectural services that requires the supplier to be present in the buyer's country in order to oversee construction of a building). Given the fixed costs involved in negotiating such contracts, they would likely be limited

- to skilled labor. Mode IV migration is distinct from migration under a guest worker program, in which an employer in an importing country directly hires a worker from an exporting country under a temporary contract.
48. Or, governments may restrict immigration because they weigh the welfare of different individuals unequally, for whatever reason favoring those opposed to immigration (Foreman-Peck, 1992).
 49. For related work, see Epstein and Nitzan (2006).
 50. See also Kessler (2001), Hatton and Williamson (2005), and O'Rourke and Sinnott (2006).
 51. In the long run, immigrants may affect voting outcomes directly through their participation in the political process (Ortega, 2004; Razin, Sadka, & Swagel, 2002).

References

- Acosta, P., Fajnzylber, P., & Lopez, H. (2007). How important are remittances in America? In P. Fajnzylber & H. Lopez (Eds.), *Remittances and Development: Lessons from Latin America* (pp. 21–50). Washington, DC: The World Bank.
- Adams, R. (2003). *International migration, remittances, and brain drain: A study of 24 labor exporting countries*. Working Paper No. 2972, World Bank Policy Research.
- Alessandra, C., & Rauch, J. E. (2002). Anonymous and group ties in international trade. *Journal of International Economics*, 58, 19–47.
- Alejandra & Manuelita Ureta (2003). *International migration, remittances and schooling: evidence from El Salvador*. *Journal of development economics*, 72(2), 429–461.
- Anderson, J. E., & van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature*, 42(3), 691–751.
- Angelucci, M. (2006). *U.S. Border Enforcement and the net inflow of illegal migration*. University of Arizona Mimeo.
- Antecol, H., Cobb-Clark, D. A., & Trejo, S. J. (2003). Immigration policy and the skills of immigrants to Australia, Canada, and the United States. *Journal of Human Resources*, 38(1), 192–218.
- Åslund, O., Edin, P. A., & Fredriksson, P. (2001). *Settlement policies and the economic success of immigrants*. IZA Discussion Paper No. 2730.
- Assuncao, J. J., & Carvalho, L. S. (2009). Financial constraints, migration, and inequality. *Journal of Development Economics*, (forthcoming).
- Auerbach, A. J., & Oreopoulos, P. (1999). Analyzing the fiscal impact of U.S. immigration. *American Economic Review Papers and Proceedings*, 89(2), 176–180.
- Aydemir, A., & Borjas, G. J. (2007, June). A comparative analysis of the labor market impact of international migration: Canada, Mexico, and the United States. *Journal of the European Economic Association*, 5(4), 663–708.
- Banerjee, A. V., & Newman, A. (1993). Occupational choice and the process of development. *Journal of Political Economy*, 101(2), 274–298.
- Barro, R. J., & Lee, J. W. (2000). *International data on educational attainment: Updates and implications*. Working Paper No. 42, Center for International Development.
- Bauer, T., & Gang, I. (1998). *Temporary migrants from Egypt: How long do they stay abroad?* CEPR Discussion Paper No. 2003.
- Beine, M., Docquier, F., & Rapoport, H. (2001). Brain drain and economic growth: Theory and evidence. *Journal of Development Economics*, 64(1), 275–289.
- Beine, M., Docquier, F., & Rapoport, H. (2006a). *Measuring international skilled migration: New estimates accounting for age of entry*. World Bank Research Report.
- Beine, M., Docquier, F., & Rapoport, H. (2006b). *Brain drain and human capital formation in developing countries: Winners and losers*. IRES Working Paper No. 2006-23.
- Benhabib, J. (1996). On the political economy of immigration. *European Economic Review*, 40, 1737–1743.
- Benhabib, J., & Jovanovic, B. (2007). *Optimal migration: A world perspective*. NBER Working Paper No. 12871.

- Berry, S. (1994). Estimating discrete choice models of product differentiation. *Rand Journal of Economics*, 25(2), 242–262.
- Bertocchi, G., & Strozzi, C. (2006a). *The age of mass migration: Economic and institutional determinants*. IZA Discussion Paper No. 2499.
- Bertocchi, G., & Strozzi, C. (2006b). *The evolution of citizenship: Economic and institutional determinants*. IZA Discussion Paper No. 2510.
- Bhagwati, J. N., & Wilson, J. (Eds.), (1989). *Income taxation and international mobility*. Cambridge: MIT Press.
- Bhagwati, J. N., & Hamada, K. (1974). The brain drain, international integration of markets for professionals, and unemployment. *Journal of Development Economics*, 1(1), 19–42.
- Bhagwati, J. N., & Rodriguez, C. (1975). Welfare theoretic analyses of the brain drain. *Journal of Development Economics*, 2(3), 195–221.
- Boeri, T., & Brücker, H. (2005). Migration, co-ordination failures, and EU enlargement. *Economic Policy*, 20(44), 629–703.
- Boeri, T., McCormick, B., & Hanson, G. H. (2002). *Immigration policy and the welfare state*. Oxford: Oxford University Press.
- Borjas, G. J. (Ed.), (2007). *Mexican immigration*. Chicago: University of Chicago Press and the National Bureau of Economic Research.
- Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77(4), 531–553.
- Borjas, G. J. (1991). Immigration and self-selection. In J. Abowd & R. Freeman (Eds.), *Immigration, trade, and the labor market* (pp. 29–76). Chicago: University of Chicago Press.
- Borjas, G. J. (1999a). *Heaven's door: Immigration policy and the American economy*. Princeton, NJ: Princeton University Press.
- Borjas, G. J. (1999b). The economic analysis of immigration. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1697–1760). Amsterdam: North-Holland.
- Borjas, G. J. (2003, November). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics*, 118, 1335–1374.
- Borjas, G. J. (2005, May). The labor market impact of high-skill immigration. *American Economic Review*, 95(2), 56–60.
- Borjas, G. J. (2006). *Labor outflows and labor inflows in Puerto Rico*. Harvard University Mimeo.
- Borjas, G. J., & Bratsberg, B. (1996). Who leaves? The out-migration of the foreign born. *Review of Economics and Statistics*, 78(1), 165–176.
- Borjas, G. J., & Hilton, L. (1996). Immigration and the welfare state: Immigrant participation in means-tested entitlement programs. *Quarterly Journal of Economics*, 111(2), 575–604.
- Borjas, G. J., Jeffrey G., and Gordon H. H. (2006). "Immigration and African-American Employment Opportunities: The Response of Wages, Employment, and Incarceration to Labor Supply Shocks." NBER Working Paper No. 12518.
- Butcher, K. F., & Piehl, A. M. (1998). Recent immigrants: Unexpected implications for crime and incarceration. *Industrial and Labor Relations Review*, 51, 654–679.
- Butcher, K. F., & Piehl, A. M. (2006). *Why are immigrants' incarceration rates so low? Evidence on selective immigration, deterrence, and deportation*. Wellesley College and Rutgers University. Mimeo.
- Caglar, A. (2006). Hometown associations, the rescaling of state spatiality and migrant grassroots transnationalism. *Global Networks*, 6(1), 1–22.
- Calavita, K. (1992). *Inside the state: The Bracero program, immigration, and the I.N.S.* New York, NY: Routledge.
- Cano, G. (2004). *Organizing immigrant communities in american cities: Is this transnationalism, or what?* Working Paper 103, Center for Comparative Immigration Studies, UCSD.
- Card, D. (2005). Is the new immigration really so bad? *Economic Journal*, 115(507), 300–323.
- Cardoso, L. (1980). *Mexican emigration to the United States, 1897-1931*. Tucson: University of Arizona Press.
- Carrington, W. J., & Detragiache, E. (1998). *How big is the brain drain?* IMF Working Paper 98/102.

- Carrington, W. J., Detragiache, E., & Viswanath, T. (1996, September). Migration with endogenous moving costs. *American Economic Review*, 86(4), 909–930.
- Chiquiar, D., & Hanson, G. (2005, April). International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2), 239–281.
- Clark, X., Hatton, T., & Williamson, J. (2007). Explaining U.S. immigration, 1971–1998. *Review of Economics and Statistics*, 89(2), 359–373.
- Clemens, M. (2007). *Do Visas Kill? Health Effects of African Health Professional Emigration*. Working Papers 114, Center for Global Development.
- Cornelius, W. A. (2005). *Impacts of U.S. immigration control policies on migratory behavior: The view from Mexican sending communities*. UCSD Mimeo.
- Cox, A. B., & Posner, E. A. (2007). The second order structure of U.S. immigration law. *Stanford Law Review*, 59(4), 809–854.
- Cuecuecha, A. (2005). *The immigration of educated Mexicans: The role of informal social insurance and migration costs*. ITAM Mimeo.
- Dahl, G. (2002). Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica*, 70(6), 2367–2420.
- Davila, A., Pagan, J. A., & Grau, M. V. (1999). Immigration reform, the INS, and the distribution of interior and border enforcement resources. *Public Choice*, 99(3-4), 327–345.
- Davila, A., Pagan, J. A., & Soydemir, G. (2002). The short-term and long-term deterrence effects of INS border and interior enforcement on undocumented immigration. *Journal of Economic Behavior and Organization*, 49(4), 459–472.
- De Giorgi, G., & Pellizzari, M. (2006). *Welfare migration in Europe and the cost of a harmonised social assistance*. IZA Discussion Paper No. 2094.
- De Melo, J., Grether, J. M., & Müller, T. (2001). *The political economy of international migration in a Ricardo-Viner model*. CEPR Discussion Paper No. 2714.
- Desai, M., Kapur, D., & McHale, J. (2003). *The fiscal impact of high skilled emigration: Flows of Indians to the U.S.* Weatherhead Center for International Affairs. Working Paper 03-01.
- Desai, M., Kapur, D., & McHale, J. (2004). Sharing the spoils: Taxing international human capital flows. *International Taxation and Public Finance*, 11, 663–1639.
- Djajic, S. (1999, February). Dynamics of immigration control. *Journal of Population Economics*, 12(1), 45–61.
- Docquier, F., & Marfouk, A. (2006). International migration by educational attainment, 1990–2000. In C. Ozden & M. Schiff (Eds.), *International migration, remittances, and the brain drain* (pp. 151–200). Washington, DC: The World Bank and Palgrave MacMillan.
- Docquier, F., & Rapoport, H. (2007). *Skilled migration: The perspective of developing countries*. Bar Ilan University Mimeo.
- Durand, J., Massey, D. S., & Zenteno, R. M. (2001). Mexican immigration in the United States. *Latin American Research Review*, 36(1), 107–127.
- Dustmann, C., & Kirchamp, O. (2002). The optimal migration duration and activity choice after re-migration. *Journal of Development Economics*, 67, 351–372.
- Dustmann, C., & Preston, I. (2004). *Racial and economic factors in attitudes to immigration*. University College London Mimeo.
- Du, Y., Park, A., & Wang, S. (2005). Migration and rural poverty in China. *Journal of Comparative Economics*, 33(4), 688–709.
- Epstein, G. S., Hillman, A. L., & Weiss, A. (1999). Creating illegal immigrants. *Journal of Population Economics*, 12, 3–21.
- Epstein, G. S., & Nitzan, S. (2006). The struggle over migration policy. *Journal of Population Economics*, 19 (4), 703–723.
- Epstein, G. S., & Weiss, A. (2001). *A theory of immigration amnesties*. IZA Discussion Paper No. 2830.
- Ethier, W. (1986, March). Illegal immigration: The host-country problem. *American Economic Review*, 76 (1), 56–71.
- Facchini, G., & Mayda, A. M. (2006). *Individual attitudes towards immigrants: Welfare-state determinants across countries*. CEPR Working Paper No. 5702.

- Facchini, G., & Willmann, G. (2005). The political economy of international factor mobility. *Journal of International Economics*, 67(1), 201–219.
- Fajnzylber, P., & Lopez, H. (Eds.) (2007). *Remittances and Development: Lessons from Latin America*. Washington, DC: The World Bank, 21–50.
- Feliciano, Z. (2001). The skill and economic performance of Mexican immigrants from 1910 to 1990. *Explorations in Economic History*, 38, 386–409.
- Fernandez-Huertas, J. (2006). *New evidence on emigration selection*. Columbia University Mimeo.
- Fix, M., & Passel, J. (2002). *The scope and impact of welfare reform's immigrant provisions*. Washington, DC: The Urban Institute.
- Foreman-Peck, J. (1992). A political economy of international migration. *The Manchester School*, 60, 359–376.
- Freeman, R. (2006). *People flows in globalization*. NBER Working Paper No. 12315.
- Freund, C., & Spatafora, N. (2007). *Informality and remittances*. The World Bank Mimeo.
- Gallardo-Sejas, H., Pareja, S. G., Llorca-Vivero, R., & Martinez-Serrano, J. A. (2006). Determinants of european immigration: A cross-country analysis. *Applied Economics Letters*, 13, 769–773.
- Gathmann, C. (2004). *The effects of enforcement on illegal markets: Evidence from migrant smuggling on the south-western border*. IZA Discussion Paper No. 1004.
- Gould, D. M. (1994). Immigration links to the home country: Empirical implications for U.S. bilateral trade flows. *Review of Economics and Statistics*, 76, 302–316.
- Grogger, J., & Hanson, G. H. (2008). *Income maximization and the selection and sorting of emigrants across destinations*. UCSD Mimeo.
- Grubel, H. G., & Scott, A. D. (1966). The international flow of human capital. *American Economic Review*, 56, 268–274.
- Hainmueller, J., & Hiscox, M. J. (2004). Educated preferences: Explaining attitudes toward immigration in Europe. Paper prepared for presentation at the *Annual Meeting of the American Political Science Association*.
- Hamilton, B., & Whalley, J. (1984). Efficiency and distributional implications of global restrictions on labor mobility. *Journal of Development Economics*, 14(1), 61–75.
- Hanson, G. H. (2006). Illegal migration from Mexico to the United States. *Journal of Economic Literature*, 44, 869–924.
- Hanson, G. H. (2007). Emigration, labor supply, and earnings in Mexico. In G. J. Borjas (Ed.), *Mexican immigration* (pp. 289–328). Chicago: University of Chicago and National Bureau of Economic Research.
- Hanson, G. H., Robertson, R., & Spilimbergo, A. (2001). Does border enforcement protect U.S. workers from illegal immigration? *The Review of Economics and Statistics*, 84(2002), 73–92.
- Hanson, G. H., Scheve, K., & Slaughter, M. J. (2007, March). Local public finance and individual preferences over globalization strategies. *Economics and Politics*, 19(1), 1–33.
- Hanson, G. H., & Spilimbergo, A. (1999). Illegal immigration, border enforcement and relative wages: Evidence from apprehensions at the U.S.–Mexico border. *American Economic Review*, 89, 1337–1357.
- Hanson, G. H., & Spilimbergo, A. (2001). Political economy, sectoral shocks, and border enforcement. *Canadian Journal of Economics*, 34(2001), 612–638.
- Hanushek, E. A., & Zhang, L. (2006). *Quality consistent estimates of international returns to skill*. NBER Working Paper No. 12664.
- Hatton, T. J., & Williamson, J. G. (1998). *The Age of Mass Migration: Causes and Economic Impact*. Oxford: Oxford University Press.
- Hatton, T. J., & Williamson, J. G. (2002). *What fundamentals drive world migration?*. NBER Working Paper No. 9159.
- Hatton, T. J., & Williamson, J. G. (2004). *Refugees, asylum seekers and policy in Europe*. NBER Working Paper No. 10680.
- Hatton, T. J., & Williamson, J. G. (2005). *A dual policy paradox: Why have trade and immigration policies always differed in labor-scarce economies*. NBER Working Paper No. 11866.
- Head, K., Ries, J., & Swenson, D. (1998). Immigration and trade creation: Econometric evidence from Canada. *Canadian Journal of Economics*, 31(1), 47–62.

- Huntington, S. (2004). *Who are we? The challenges to America's national identity*. New York: Simon and Schuster.
- Ibarrahan, P., & Lubotsky, D. (2005). The socioeconomic status of Mexican migrant families: New evidence from the 2000 Mexican census. In G. J. Borjas (Ed.), *Mexican immigration* (pp. 159–192). Chicago: University of Chicago and National Bureau of Economic Research.
- Inter-American Development Bank. (2004). *Sending money home: Remittance to Latin America and the Caribbean*. Washington, DC: IADB.
- Jasso, G., & Rosenzweig, M. (2005). *Selection criteria and the skill composition of immigrants: A comparative analysis of Australian and US employment immigration*. NYU and Yale University Mimeo.
- Karemera, D., Onguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics*, 32(13), 1745–1755.
- Kessler, A. (2001). *Immigration, economic insecurity, and the ambivalent American Public*. Working Paper No. 41, Center for Comparative Immigration Studies, UCSD.
- Kim, J. (1998). Economic analysis of foreign education and students abroad. *Journal of Development Economics*, 56(2), 337–365.
- Lacuesta, A. (2006). *Emigration and human capital: Who leaves, who comes back and what difference does it make?* Working Paper No. 620, Bank of Spain.
- Lapointe, M., & Orozco, M. (2004). Mexican hometown associations and development opportunities. *Journal of International Affairs*, 57(2), 1–21.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(3), 3–42.
- Lucas, R. E. B., & Stark, O. (1985). Motivations to remit: Evidence from Botswana. *Journal of Political Economy*, 93(5), 901–918.
- Mariani, F. (2007). Migration as an antidote to rent seeking. *Journal of Development Economics*, 84(2), 609–630.
- Massey, D. S., & Espinosa, K. E. (1997). What's driving Mexico-U.S. migration? A theoretical, empirical, and policy analysis. *American Journal of Sociology*, 102(4), 939–999.
- Massey, D. S., Goldring, L., & Durand, J. (1994). Continuities in transnational migration: An analysis of nineteen Mexican communities. *American Journal of Sociology*, 99(6), 1492–1533.
- Mayda, A. M. (2005). *International migration: A panel data analysis of economic and non-economic determinants*. IZA Discussion Paper No. 1590.
- Mayda, A. M. (2006). Who is against immigration? A cross-country investigation of individual attitudes toward immigrants. *Review of Economics and Statistics*, 87(3), 510–530.
- Mazzolari, F. (2006). *Determinants and effects of naturalization: Effects of dual citizenship laws*. UC Irvine Mimeo.
- McCulloch, R., & Yellen, J. L. (1977). Factor mobility, regional development, and the distribution of income. *The Journal of Political Economy*, 85(1), 79–96.
- McHale, J. (2007). *Taxation and skilled Indian migration to the United States: Revisiting the Bhagwati tax*. Queen's University. Mimeo.
- McKenzie, D., Gibson, J., & Stillman, S. (2006). *How important is selection? Experimental versus non-experimental measures of the income gains from migration*. World Bank. Mimeo.
- McKenzie, D., & Rapoport, H. (2006). *Self-selection patterns in Mexico-U.S. migration: The role of migration networks*. World Bank and Bar-Ilan University Mimeo.
- McKenzie, D., & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics*, 84(1), 1–24.
- Mesnard, A. (2004). Temporary migration and capital market imperfections. *Oxford Economic Papers*, 56, 242–262.
- Mishra, P. (2007). Emigration and wages in source countries: Evidence from Mexico. *Journal of Development Economics*, 82(1), 180–199.
- Miyagiwa, K. (1991). Scale economies in education and the brain drain problem. *International Economic Review*, 32(3), 743–759.
- Mountford, A. (1997). Can a brain drain be good for growth in the source economy? *Journal of Development Economics*, 53(2), 287–303.

- Munshi, K. (2003, May). Networks in the modern economy: Mexican migrants in the US labor market. *Quarterly Journal of Economics*, 118, 549–597.
- Nielsen, M. E., & Olinto, P. (2007). Do Conditional Cash Transfers Crowd Out Private Transfers? In Fajnzylber, P., & Lopez, H. (Eds.), *Remittances and Development: Lessons from Latin America* (pp. 253–298). Washington, DC: The World Bank.
- O'Rourke, K., & Sinnott, R. (2006). The determinants of individual attitudes toward immigration. *European Journal of Political Economy*, 22(4), 838–861.
- OECD. (2006). *International migration outlook*. OECD: Paris.
- Orrenius, P. M., & Zavodny, M. (2005, October). Self-selection among undocumented immigrants from Mexico. *Journal of Development Economics*, 78(1), 215–240.
- Ortega, F. (2004). *Immigration policy and the welfare state*. New York University Mimeo.
- Ozden, C., & Schiff, M. (Eds.). (2006). *International migration, remittances, and the brain drain*. Washington, DC: The World Bank and Palgrave McMillan.
- Passel, J. S. (2006). *Size and characteristics of the unauthorized population in the US: Estimates from the March 2005 current population survey*. Pew Hispanic Center.
- Pedersen, P. J., Pytlakova, M., & Smith, N. (2004). *Selection or network effects? Migration flows into 27 OECD countries, 1990-2000*. IZA Discussion Paper No. 1104.
- Poutvaara, P. (2005). *Social security incentives, human capital investment and mobility of labor*. IZA Discussion Paper No. 1729.
- Pritchett, L. (2006). *Let their people come: Breaking the Gridlock on global labor mobility*. Washington, DC: Center for Global Development.
- Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, 12(2), 111–134.
- Ramos, F. (1992). Out-migration and return migration of Puerto Ricans. In G. Borjas & R. Freeman (Eds.), *Immigration and the work force* (pp. 49–66). Chicago: University of Chicago Press and NBER.
- Rapoport, H. (2002). Migration, credit constraints, and self-employment: A simple model of occupational choice, inequality, and growth. *Economics Bulletin*, 15(7), 1–5.
- Ratha, D., & Shaw, W. (2007). *South-south migration and remittances*. The World Bank Mimeo.
- Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics*, 48 (1), 7–35.
- Rauch, J. E., & Trindade, V. (2002, February). Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1), 116–130.
- Razin, A., & Sadka, E. (1999). Migration and pensions with international capital mobility. *Journal of Public Economics*, 74, 141–150.
- Razin, A., Sadka, E., & Swagel, P. (2002). Tax burden and migration: Political economy theory and evidence. *Journal of Public Economics*, 85, 167–190.
- Reyes, B. I., Johnson, H., & Van, R. (2002). *Holding the line? The effect of recent border build-up on unauthorized immigration*. San Francisco: Public Policy Institute of California.
- Robertson, R. (2000). Wage shocks and North American labor market integration. *American Economic Review*, 90(4), 742–764.
- Rosenzweig, M. (2006). Global wage differences and international student flows. *Brookings Trade Forum*, 9, 57–86.
- Rosenzweig, M. (2007). *Education and migration: A global perspective*. Yale University. Mimeo.
- Rosenzweig, M., & Stark, O. (1989). Consumption smoothing, migration, and marriage: Evidence from rural India. *Journal of Political Economy*, 97(4), 905–926.
- Roy, A. D. (1951, June). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3, 135–146.
- Rubalcava, L., Teruel, G., Goldman, N., & Thomas, D. (2006). *Do healthier Mexicans migrate to the United States? New findings from the Mexican family life survey*. CIDE Mimeo.
- Saxenian, A. (1999). *Silicon Valley's new immigrant entrepreneurs*. San Francisco: Public Policy Institute of California.
- Saxenian, A. (2002). *Local and global networks of immigrant professionals in Silicon Valley*. Public Policy Institute of California Monograph.

- Scheve, K. F., & Slaughter, M. J. (2001a). Labor-market competition and individual preferences over immigration policy. *Review of Economics and Statistics*, 83(1), 133–145.
- Scheve, K. F., & Slaughter, M. J. (2001b). What determines individual trade-policy preferences. *Journal of International Economics*, 54(2), 267–292.
- Schiff, M. (2006). Brain gain: Claims about its size and impact on growth are greatly exaggerated. In C. Ozden & M. Schiff (Eds.), *International migration, remittances, and brain drain* (pp. 201–225). Washington, DC: The World Bank.
- Schiff, M. (2007). *Optimal immigration policy: Permanent, guest-worker, or mode IV?*. World Bank Mimeo.
- Scholten, U., & Thum, M. (1996). Public pensions and immigration policy in a democracy. *Public Choice*, 66, 347–361.
- Sinn, Hans-Werner, (2004). Migration, Social Standards and Replacement Incomes: How to Protect Low-income Workers in the Industrialized Countries Against the Forces of Globalization and Market Integration September. NBER Working Paper No. 10798.
- Spilimbergo, A. (2006). *Foreign students and democracy*. IMF Mimeo.
- Stark, O., Helmenstein, C., & Prskawetz, A. (1997). A brain gain with a brain drain. *Economics Letters*, 55, 227–234.
- Stark, O., & Wang, Y. Q. (2002). Inducing human capital formation: Migration as a substitute for subsidies. *Journal of Public Economics*, 86(1), 29–46.
- Storesletten, K. (2000). Sustaining fiscal policy through immigration. *Journal of Political Economy*, 108(2), 300–323.
- Teruel, G., & Davis, B. (2000). *Final report: An evaluation of the impact of PROGRESA cash payments on private inter-household transfers*. International Food Policy Research Institute.
- U.S. Immigration and Naturalization Service. (2000). *Statistical yearbook of the immigration and naturalization service, 1998*. Washington, DC: U.S. Government Printing Office.
- United Nations. (2002). The Universal Declaration of Human Rights. <http://www.un.org/en/documents/udhr/>
- United Nations. (2005). *Trends in total migrant stock: 2005 revision*. New York: Department of Economic and Social Affairs, United Nations.
- Volger, M., & Rotte, R. (2000). The effects of development on migration: Theoretical issues and new empirical evidence. *Journal of Population Economics*, 13(3), 485–508.
- Walmsley, T. L., & Alan Winters, L. (2005). Relaxing the restrictions on the temporary movement of natural persons: A simulation analysis. *Journal of Economic Integration*, 20(4), 688–726.
- Wellisch, D., & Walz, U. (1998). Why do rich countries prefer free trade over free immigration? The role of the modern welfare states. *European Economic Review*, 42, 1595–1612.
- Williams, A. M., King, R., & Warnes, T. (1997). A place in the Sun: International retirement migration from northern to southern Europe. *European Urban and Regional Studies*, 4(2), 115–134.
- Winters, P., de Janvry, A., & Sadoulet, E. (2001, Winter). Family and community networks in Mexico-U.S. migration. *The Journal of Human Resources*, 36(1), 159–184.
- Wong, K. Y., & Yip, C. K. (1999). Education, economic growth, and brain drain. *Journal of Economic Dynamics and Control*, 23, 699–726.
- Woodruff, C., & Zenteno, R. (2007, March). Migration networks and microenterprises in Mexico. *Journal of Development Economics*, 82(2), 509–528.
- Yang, D. (2006). Why do migrants return to poor countries? Evidence from Philippine migrants responses to exchange rate shocks. *Review of Economics and Statistics*, 88(4), 715–735.
- Yang, D. (2008, April). International migration, remittances, and household investment: Evidence from Philippine migrants' exchange rate shocks. *Economic Journal*, 118(528), 591–630.

Aid and Conditionality*

Jonathan R. W. Temple

Department of Economics, University of Bristol, Bristol BS8 1TN, United Kingdom

Contents

1. Introduction	4417
2. Key Concepts and Stylized Facts	4423
2.1 The Millennium Development Goals	4423
2.2 Recent trends in aid	4425
2.3 The forms of aid	4431
2.4 Fungibility	4432
2.5 Debt relief	4433
3. Aid and Growth in Theory	4435
3.1 Capital shortage and neoclassical growth models	4435
3.2 Poverty-trap models	4437
3.3 Virtuous circles and amplification effects	4439
3.4 Binding constraints	4439
3.5 Complementary inputs	4441
3.6 Growth versus basic needs	4442
3.7 Summary	4442
4. Has Aid Worked?	4443
4.1 Project evaluations	4443
4.2 Case studies	4444
4.3 Cross-country evidence	4445
4.4 Aid and social indicators	4447
4.5 Summary	4448
5. The View From Trade Theory	4449
5.1 The transfer problem	4449
5.2 The Dutch Disease	4450
5.3 Evidence on the Dutch Disease	4451
6. Aid, Fiscal Policy, and Volatility	4452
7. Aid and Governance	4454
7.1 History and the state	4456
7.2 Aid and political leadership	4457
7.3 Aid and the common-pool problem	4458
7.4 Aid and corruption	4459
7.5 Aid, political instability, and war	4461
7.6 Aid to fragile states	4462

8.	A New Partnership Model?	4463
8.1	Should democracies get the most aid?	4464
8.2	Democracy as a necessary condition	4465
8.3	Democracy as a sufficient condition	4466
8.4	Summary	4467
9.	Conditionality in Theory and Practice	4468
9.1	Policy conditionality	4469
9.2	Commitment problems	4470
9.3	Heroes and villains	4471
9.4	Conditionality versus ownership	4472
9.5	Governance conditionality	4474
10.	Selective Aid Allocation	4475
10.1	Developing rules for aid allocation	4477
10.2	Dynamic aid allocation	4477
11.	The New Conditionality	4478
11.1	The meaning of ownership	4479
11.2	Poverty Reduction Strategy Papers	4479
11.3	Social participation	4480
11.4	New forms of conditionality	4481
11.5	The future of conditionality	4483
12.	Donor Policies	4483
12.1	Proliferation of projects and programs	4484
12.2	Learning and accountability	4485
12.3	Randomized trials and evidence-based aid policies	4486
12.4	Evaluate, evaluate, evaluate	4489
13.	Meeting Basic Needs	4490
13.1	Public service delivery	4491
13.2	Agriculture and rural development	4492
14.	Aid for the Twenty-First Century	4493
14.1	New directions for foreign aid	4495
15.	Conclusions	4499
Appendix		4501
End Notes		4506
References		4511

Abstract

This chapter examines the conditions under which foreign aid will be effective in raising growth, reducing poverty, and meeting basic needs in areas such as education and health. The primary aim is not to draw policy conclusions, but to highlight the main questions that arise, the contributions of the academic literature in addressing them, and the areas where much remains unknown. After describing some key concepts and trends in aid, the chapter examines the circumstances under which aid might transform productivity, and when it can achieve things that private capital flows cannot. The chapter reviews the relevant theory and evidence. Next, it turns to some of the other considerations that might form part of a structural model linking outcomes to aid. These include Dutch Disease effects, the fiscal response to aid, and the important connections between aid and governance, both positive and negative.

The second half of the chapter examines when donors should attach conditions to aid. It reviews the debates on traditional policy conditionality, and potential alternatives, including the ideas underpinning the new "partnership" model. This model gives greater emphasis to a combination of autonomy and accountability, for countries where governance is strong. In other cases, donors may seek to attach conditions based on governance reform, and introduce new versions of traditional policy conditionality. The chapter also discusses controversies over the appropriate role of country ownership of aid programs. It goes on to discuss some donor failings, the future roles of randomized trials and evaluation, and the scope for aid to meet basic needs. The chapter ends with a discussion of some of the most innovative ideas for the reform of aid, and a summary of the main conclusions.

JEL classifications: F35, O10, O40

Keywords

foreign aid
conditionality
governance

1. INTRODUCTION

In January 1949, US President Harry S. Truman addressed the American people. He emphasized the urgent need for economic development around the world:

More than half the people of the world are living in conditions approaching misery. Their food is inadequate. They are victims of disease. Their economic life is primitive and stagnant. Their poverty is a handicap and a threat both to them and to more prosperous areas. For the first time in history, humanity possesses the knowledge and the skill to relieve the suffering of these people.

Sixty years later, that urgent need remains. It is estimated that more than one billion people live on less than a dollar a day, in conditions of absolute poverty and deprivation. Although the problems of underdevelopment have sometimes disappeared from view in the Western media, they currently have a higher profile than ever before. As Easterly (2001) noted, recent debt relief campaigns have included roles for such disparate figures as Bono, Jeffrey Sachs, the late Pope John Paul II, and the Dalai Lama. The glossy magazine *Vanity Fair* devoted its entire July 2007 issue to Africa, with a role for Bono as guest editor. Exaggerating only slightly, movie stars are now expected to combine their views on personal growth with a careful analysis of the latest debt relief initiatives.

Greater awareness of world poverty leads many to ask whether aid is an effective remedy. Advocates of foreign aid have to battle the perception that it does not work, and find it hard to establish definitive success stories. In fact, it is not difficult to write a selective history of foreign aid as a chapter of disasters, in which noble intentions sometimes end in tragedy and farce. These accounts often draw attention to the aid given to

dictators such as Zaire's Mobutu, despite an abysmal policy record. They often feature examples of misguided projects, such as the British ground nuts fiasco, the Nigerian state-owned Ajaokuta steel mill that cost billions and has yet to produce a single tonne of steel (Easterly, 2006b) and the Morogoro shoe factory in Tanzania, built with World Bank money, which never produced at more than 4% of its installed capacity (Easterly, 2001).

To some extent, these examples can be seen as anomalies, the expensive follies that will always arise when large sums of money are spent in policy environments of varying quality. As Bhagwati (1972) pointed out, risky investments are a routine part of economic activity in developed countries, and projects in the private sector sometimes fail. Perhaps more worrying are the accounts of public service delivery in some developing countries, in which various kinds of failure seem routine, and undermine the ability of public expenditure to meet basic needs.

Another concern is the distance that can loom between intentions and consequences. To illustrate this, and the complexity of what can happen when one society encounters another, Cohen (2006, pp. 2-3) provides an account of a mountain village in Algeria under colonial rule. The French colonialists used DDT to combat malaria and typhoid, and built a road to address the region's isolation. These actions, intended to bring the benefits of Western civilization, then set in train a complex chain of events including a population explosion, more intensive farming, and deterioration in the land needed to support the villagers' livestock. Some people were ruined, others made rich; inequality increased, and the equilibrium of a traditional society was irreversibly altered. Cohen's sharply drawn vignette is a reminder that even well-intentioned actions have consequences that unfold, for good and bad, in ways that are difficult to predict. It is much easier to change the world than it is to improve it, and any history of foreign aid is partly a history of the world's stubborn resistance to improvement.

Considerations like these invite skepticism about the effectiveness of foreign aid, but this is a topic on which there is always more to be said. Even some of the most trenchant critics of existing aid policies, notably Easterly (2006b), acknowledge the benefits of some of the public health measures that have been undertaken around the world, partly on the basis of aid flows and external technical assistance. Sachs (2005) provides a number of examples of aid success stories, including the eradication of small-pox, the drive to eradicate polio, the control of African river blindness, and the targeted science of the Green Revolution in agriculture. The Campaign for Child Survival, begun by UNICEF in 1982, is estimated to have saved around 12 million lives by the end of the 1980s. The belief that aid is inevitably wasted begins to look like a leap of faith, or lack of faith. For some of the most thoughtful critics, it is the potential for aid to do so much good that makes its past and current failings so controversial.

Such impressionistic accounts can take us only so far. Some of the academic research on aid, in aiming at more rigorous answers, circles around what seems the

most natural question: does aid work? The problem with this question can be seen by analogy. Nobody asks whether medical science works, because the question invites an answer—yes and no—that is correct but uninformative. A more interesting question is: *when* does aid work? This chapter aims to summarize what is currently known about when aid works and when it doesn't; and identify the subset of things we need to know, that there is some hope of finding out.

But the “does aid work?” question never quite disappears from view, which testifies to the wider skepticism about whether aid can ever achieve its goals. In this view of the world, it remains a meaningful question, just as it would have been sensible to ask eighteenth-century doctors whether they knew what they were doing. Most people can agree on the goals, but views differ on whether we have the knowledge and means to achieve them. We cannot rule out the possibility that aid, like early medicine, sometimes makes things worse. The analogy with early medicine also hints at something else: failure today need not imply failure tomorrow, if there is a readiness to learn from experience.

But the medical metaphor may ultimately be more interesting for its limitations. Most obviously, it assumes too much of the knowledge and skills of outsiders. Many are wary of the idea that aid can ever be a well-directed surgical intervention that makes all the difference. Instead, the aim is to influence an ongoing process that is not well understood, and it is not clear where to intervene, or how. The process by which underdevelopment persists must depend on complex social, political, and economic interactions, and the appropriate response will vary from case to case. This suggests the need for intellectual caution. After all, in building the necessary knowledge, aid agencies have fewer cases to extrapolate from than a doctor treating a patient. In such a context, learning will be slow and unsteady, with reversals and errors along the way. Meanwhile, aid recipients will have knowledge of their circumstances that outsiders do not, and may resist any suggestion of passive acquiescence in external advice.

The medical metaphor breaks down altogether by implying that the patient is a single actor. Some of the most important debates in this literature concern the effects of aid when governments have objectives that conflict with those of donors, and with the interests of their own private sector and the poor; when governments are divided between reformers and nonreformers; when distinct social groups seek control of the apparatus of the state and the government budget; and when governments, democratic or otherwise, lack the means to deliver effective public services. Considering these points in more general terms, the analysis of aid should be based on an equilibrium in which political and economic outcomes are jointly determined, and in which the government is not a unified whole but characterized by a wide array of incentive and management problems.

This means that an understanding of aid effectiveness has to go beyond the traditional concerns of early development and growth economics. In some ways, the

discussion of aid has mirrored wider intellectual developments in economics as a discipline. The early literature tended to see aid as a cash injection into a simple hydraulic system, in the style of Keynesian macroeconomics. The literature then moved toward the evaluation of specific projects using cost-benefit analysis, adapted to the context of developing countries. More recently, when researchers talk about aid, they talk about political economy and institutions, the theory of incentives, the statistical evaluation of health and education interventions, and the administration and reform of public services.

With so many relevant considerations, any review of the aid literature faces a core structural problem. It does not make sense to discuss aid effectiveness outside the wider context of underdevelopment, and without paying attention to general equilibrium effects. The ideal starting point would be a structural model that related outcomes to aid policies, based on well-understood mechanisms. But this prospect is distant, and the existing literature has to be reviewed on different lines. The chapter will be organized in roughly two halves. The first will take a general equilibrium perspective, discussing some aggregate models, theory and evidence on the Dutch Disease, and the joint determination of an economic and political equilibrium. It will also discuss the cross-country evidence that, for better or worse, underpins much of the academic literature on foreign aid. In effect, aid will be seen through the lenses of growth economics, trade theory, and political economics, in that order. The chapter will sketch some of the considerations and intermediating mechanisms that would be needed in a full structural model.

The second half of the chapter will be organized more explicitly around current debates among donors. An important part of the background is the emergence of a new “partnership” model for donors and aid recipients, summarized in, for example, World Bank (2004) and Bourguignon and Leipziger (2006). This model draws together some of the trends in aid policies that emerged in the 1990s. These include a shift toward greater targeting of aid toward countries with well-regarded institutional and policy environments; a greater commitment on the part of donors to achieving concrete progress on growth and poverty; more emphasis on governance, institutions, and local “ownership” of reforms; the streamlining of conditions attached to loans; and greater reliance on direct support for the budgets of recipient governments, relative to sponsorship of multiple projects. The chapter will review these changes in donor policy, drawing attention to areas where the model is incomplete and differences in opinion remain.

The aims and structure of the chapter can be contrasted with Collier (2007) and Sachs (2005). Their books advocate a range of measures that could assist poorer countries, each located within a particular diagnosis of the constraints on development and poverty alleviation. Part of their contribution is to locate aid explicitly within a wider context. For example, the approach of Collier (2007) is to discuss obstacles to

development—conflict, resource rents, adverse geography, and weak governance—and then consider the role of various policy instruments, in which financial transfers are just one possible form of assistance. The aims of this chapter are necessarily less ambitious, since space does not allow a lengthy discussion of the wider context, the key constraints on growth and development, and alternatives to financial transfers. All this means that drawing practical conclusions is risky. The chapter rarely attempts to contribute to the policy debate directly. Instead, it aims to review and evaluate the academic literature that should inform that debate.

There is at least one point of agreement with the analyses of Collier and Sachs, however. They both emphasize the need for differential diagnosis: a policy response that is based on a flexible typology of developing countries. This typology might differentiate between the main obstacles to growth, and specify different forms of conditionality for different countries. The desirability of this is hardly earth-shattering news, and the real challenge is translating it into something concrete enough to be operational. If only implicitly, much of the chapter will be about this translation.

An operational approach requires consideration of the objectives of developing-country governments, and the appropriate response. The naïve view, which is nevertheless popular with some NGOs, is that attaching conditions to aid is always inappropriate, because governments should be trusted to act in the interests of their populations. At the other extreme, the orthodoxies of the Washington Consensus were founded on wariness of the aims and capacities of some developing-country governments. Instead, the leading institutions promoted reforms that were designed to expand the scope of markets and limit the scope for discretionary government action. These goals were sought partly by persuasion and partly by explicit conditions on loans. In retrospect, however, the loan conditions imposed in the 1980s and 1990s turned out to be an inadequate substitute for local commitment to reform.

If there is a consensus emerging in development policy, it is that neither extreme is sustainable, and that conditionality should be balanced against autonomy in different ways for different countries. This links to another theme of the chapter, the consequences of the spread of democracy, a remarkable feature of the 1990s. It may be optimistic to talk of a new politics in developing countries, but improvements in governance have helped to drive the new partnership model. The obvious criticism is that the model, in emphasizing autonomy and accountability, only applies to countries where the problems facing donors are least serious to start with. For countries where governments assign a low priority to development and poverty alleviation, or lack state capacity, the path is less clear. A lot of faith is being placed in conditionality attached to governance and the policy-making process, and the chapter will return to this idea several times.

Another current debate surrounds the prospects for scaling-up aid successfully. The discussion will often draw on the parallel between aid flows and resource rents.

The much-discussed “resource curse” thesis is that high and unstable revenues from natural resources have led to weak performance. This is a warning that other forms of transfer to developing countries could be ineffective or even harmful. The prospect of scaling-up reinforces the need to analyze Dutch Disease effects, the consequences for political economy and governance, and difficult but fundamental questions about long-run accountability and the risks of aid dependence.

There is also a long-standing controversy over how much aid should be seeking to achieve, and the terms in which it should be framed. Easterly (2009b) contrasts the “transformational” project of some recent commentary with a more cautious “marginal” approach based on incremental changes and experimentation; the distinction is about habits of mind rather than specific proposals. But there are really three positions here, rather than two: not only transformation versus gradualism, but also the competing idea of transformation through small changes at a large number of margins. And there are important connections to an older debate: whether to make basic needs the immediate priority, or promote long-term growth, so that poorer countries generate the capacity to help themselves.

In discussing these issues, the chapter will try to avoid a common weakness of the academic and policy literature. The intellectual fashions in the study of foreign aid sometimes congregate around abstract, high-level concepts like decentralization, governance, and reform.¹ The literature occasionally degenerates into bland prose in which capacities are built and institutions strengthened, without any concrete detail on how these outcomes can be achieved. Policy recommendations made on this basis will inevitably lack credibility and, without concrete detail, attempts to remake the world in the name of abstract concepts are doomed to failure. It is not an accident that a significant fraction of the literature on foreign aid is dedicated to unpicking the conceptual simplicity of what went before.

There are at least two underlying reasons for this state of affairs, one bad, and one more defensible. The bad reason is that filling out practical detail can be an onerous task. The more defensible reason is that many of these debates are complicated and hard to resolve: aid thinking is often difficult thinking. Fidelity to the world’s complexity may be an admirable quality in, say, a novelist, but as an intellectual strategy it has clear problems. In the study of foreign aid, the most important questions are often difficult or impossible to answer. Attempts to derive simple formulas and unifying concepts are an almost inevitable counterpart of intellectual progress.

While searching for a path through these ideas and concepts, much of the chapter will stay close to the academic literature in its concerns. The advantage is an in-built bias toward the research questions where progress is most likely to be feasible. At the same time, it will sometimes be important to highlight areas which have resisted understanding, yet may be fundamental to aid in practice. Even 60 years after Truman’s inaugural address, not all observers are equally confident that humanity has the knowledge, the skills, or the political will needed to address world poverty.

2. KEY CONCEPTS AND STYLIZED FACTS

This section will briefly set out some of the historical background, stylized facts, and key concepts. To some extent, the beginnings of modern aid programs are visible in measures undertaken in the early twentieth century by colonial powers. In the 1930s, the US government began to lend to countries in Latin America, as part of Franklin D. Roosevelt's "Good Neighbors" initiative. But the modern, large-scale aid effort is more immediately recognizable in the institutions and programs established after 1944. In particular, the USA provided billions of dollars of economic and technical assistance toward the postwar reconstruction of Western Europe, under the auspices of the Marshall Plan. The IBRD arm of the World Bank Group came into formal existence in 1945, and the IDA arm 15 years later. Simultaneously with the development of these and other multilateral institutions, national development ministries and aid agencies began to provide significant resources to poorer countries on a bilateral basis. Table 1 gives a schematic overview of the main developments since the 1960s, many of which will be discussed in the following sections.²

2.1 The Millennium Development Goals

The goals of aid have varied over time, sometimes with growth in the ascendancy, and sometimes basic needs. In the 1990s, international conferences and summits began to develop an explicit set of humanitarian development goals. In September 2000, the member states of the United Nations unanimously adopted the Millennium Declaration, leading to the adoption of the Millennium Development Goals. They range widely, with 2015 as the main target date. Compared to 1990, the aim is to halve the proportion of people whose income is less than \$1 a day, and the proportion of people who suffer from hunger; achieve universal primary education; empower women; reduce under-five mortality by two-thirds; reduce maternal mortality by three-quarters; combat HIV/AIDS, malaria, and other diseases; halve the proportion of people without sustainable access to safe drinking water and basic sanitation; and achieve, by 2020, a significant improvement in the lives of at least 100 million slum dwellers.³

This is an ambitious agenda. Academic estimates of the costs, in terms of the additional money needed each year, range as high as \$75 billion (Clemens, Kenny, & Moss, 2007). In terms of the achievements so far, the usual summary is that performance is weakest in sub-Saharan Africa, which may not reach any of the goals by 2015. Easterly (2009a) argues that this reflects the flawed design of the goals, which specified an extent of progress for Africa that had few historical precedents elsewhere, and was never likely to be achieved. Later sections of the chapter will address the capacity of aid to meet basic needs, in Africa and elsewhere, and the continued relevance of other goals, primarily economic growth.

Table 1 The history of aid: A schematic presentation

	Context	Development orthodoxy	Donor focus	New trends/rising forms of aid
1960s	Major bilateral aid programs established	Role for state in planning and investment, perhaps also in production	Productivity Infrastructure	Technical assistance Project aid
1970s	Expansion of multilaterals (IMF, World Bank)	As in 1960s	Agriculture Basic needs	Greater focus on low-income countries
1980s	Debt crisis Fading of central planning Rise of NGOs	Washington Consensus	Macroeconomic reform Growth	Program aid Structural adjustment lending Debt relief
1990s	HIPC debt problems Aid fatigue Transition after Cold War	Market-friendly institutions Export and FDI promotion	Poverty alleviation Governance Investment climate	Sector-wide support HIPC initiative
2000s	Antiglobalization movement High media profile of world poverty “War on terror”	Uncertainty	Millennium Development Goals Governance Health/HIV-AIDS	Poverty Reduction Strategy Papers (PRSPs) Selective aid allocation Budget support Postconflict aid

Intended as broadly indicative only. The form of the presentation draws on a similar table in Hjertholm and White (2000).

2.2 Recent trends in aid

This section describes some recent trends in aid, using data from the World Bank's *World Development Indicators* (April 2008) and the standard (and well designed) online database maintained by the Development Assistance Committee (DAC). The DAC is the grouping within the OECD that deals with issues related to developing countries. DAC-affiliated donors have traditionally accounted for the majority of the world's aid to developing countries, although Middle Eastern countries were significant donors in the late 1970s and early 1980s, and aid from China has increased in recent years.

The DAC defines net overseas development assistance (ODA) as financial flows to developing countries from the official sector (essentially, governments and multilateral institutions) that have the promotion of economic development and welfare as their main objective, and that are on concessional financial terms: either grants, or loans at interest rates substantially below a reference rate, typically 10%. To qualify as concessional, loans must have a grant element of at least 25%, in present-value terms.⁴ The net ODA figures, although widely used, combine grants and concessional loans in a way that has potential dangers. As a capital flow concept, the figures for net ODA count loans at face value, and are net of principal repayments, but not of interest repayments on old loans. As Bhagwati (1972) cautioned, this means that net ODA can overstate net resource transfers.⁵ But, at least in present value terms, most aid is currently in the form of grants, as we will see in the following section.

Figure 1 shows the growth in the real value of aid over time, measured in terms of 2006 US dollars, and divided between the major donors. Interesting patterns include the decline in aid in the mid-1990s, the rapid recent growth, and the increase in the share of aid coming from the European members of the DAC. The relative importance of multilateral aid has also increased since the 1960s. But it should be noted that the comparison of aid flows over time raises significant problems, especially when the aim is to determine the real value of the aid to recipient countries. The extent of tied aid has varied over time, and dollar-denominated price levels will vary within recipient countries (Bhagwati, 1970, 1972).

Figure 2 provides some evidence on where the aid has been allocated: it shows a long-term decline in Asia's share of net ODA, an increase in the share of Africa in the 1970s, which has since stabilized, and a fairly stable share for the Americas. The remainder (not shown) is divided between smaller groups, such as Oceania and the East European transition countries. One long-term change is that India accounted for as much as 10% of net ODA in the mid-1970s, but that figure is now closer to 2% (Gupta, Pattillo, & Wagh, 2006).

The relative merits of grants and loans have been much debated. Figure 3 plots the ratio of grants to net ODA (both in current prices), for the DAC donors (left-hand axis) and the multilaterals (right-hand axis). The relative importance of grants fell in

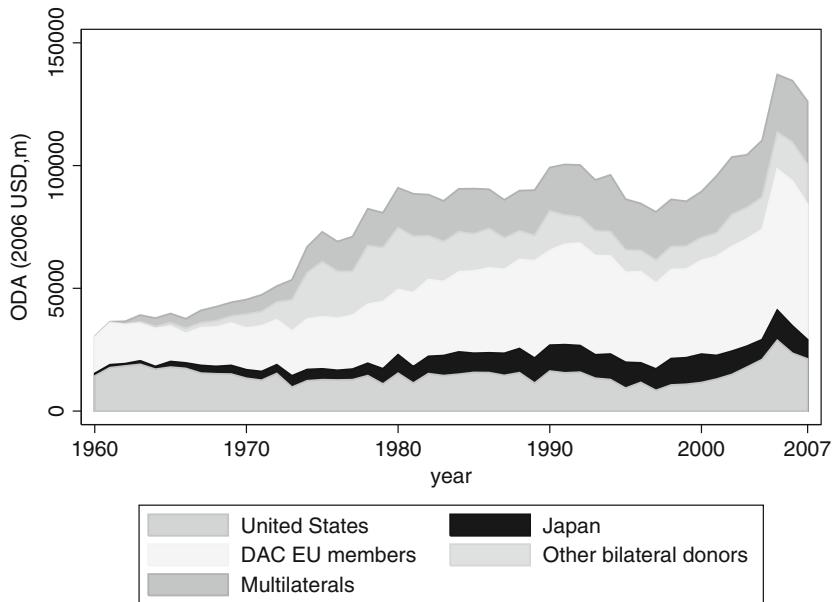


Figure 1 Net ODA in real terms, 1960–2007. Source: Online DAC database (December 2008).

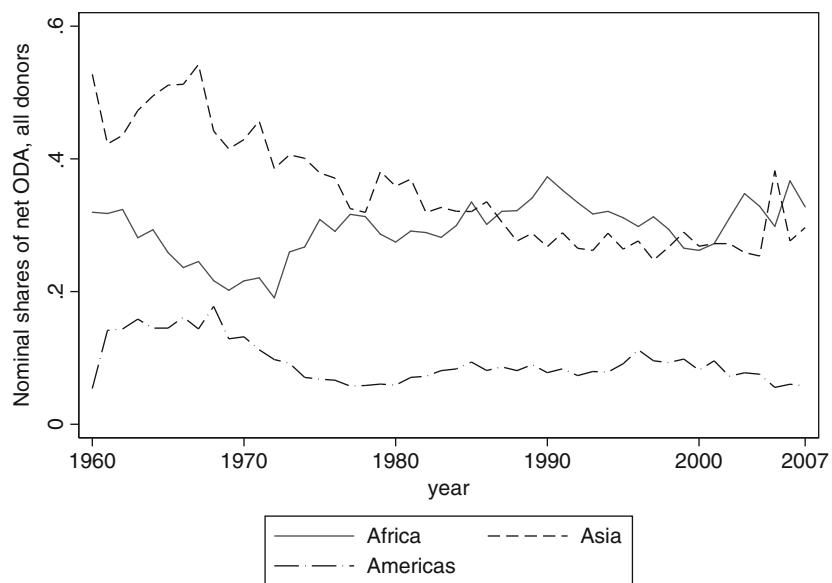


Figure 2 Nominal shares of net ODA for three regions. Source: Online DAC database (December 2008). Not all regions shown.

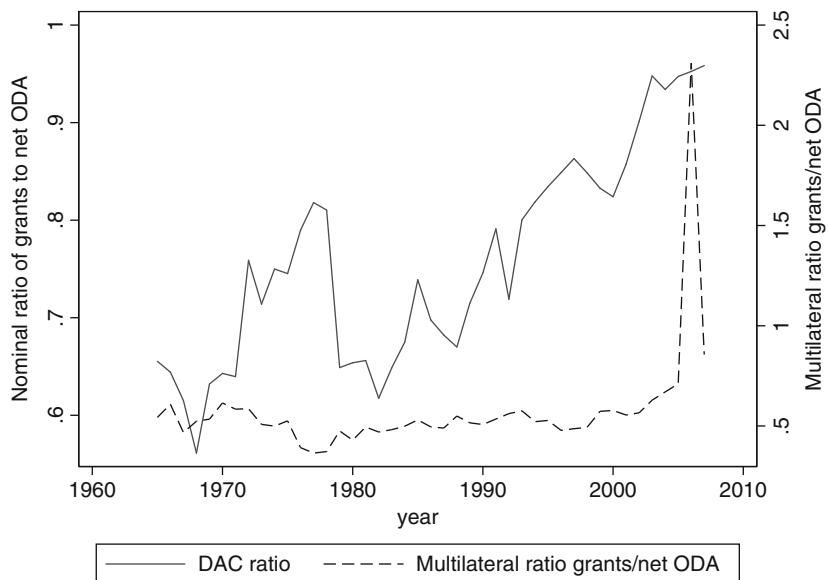


Figure 3 Ratio of grants to net ODA, 1965–2006. Source: Online DAC database (December 2008). The left-hand axis shows the figures for DAC-affiliated donors and the right-hand axis those for the multilateral donors. The spike in 2005 for the multilaterals reflects the unusually high debt relief of that year.

the 1980s, when structural adjustment lending was introduced, but has since climbed steeply. The large spike in the 2005 ratio for the multilaterals reflects the unusually high debt relief of that year. In terms of other changes in the nature of aid, the early 2000s have seen a greater role for payments directly into government budgets (“budget support”) and other forms of program aid, such as the pooling of resources from multiple donors for sector-wide support in areas such as health and education (“SWAs”). Several donors are committed to expanding program aid further, relative to more traditional project-based aid. This is part of a longer term shift away from project aid, although the financing of discrete projects continues to play a major role. For more details on these trends, see Gupta et al. (2006) and Riddell (2007).

The allocation of aid across countries is highly skewed. Aid as a share of national income tends to be highest for the poorest countries, as in the first panel of Figure 4, using data from 2006. The dashed line is based on a nonlinear regression fit, and shows that poorer countries are more likely to receive high aid relative to their income. Bourguignon, Levin, and Rosenblatt (2009) estimate that, if aid were equally shared within each recipient country, about 40% of global aid would be allocated to the poorest decile of the world’s population, and about 25% to the second decile. One reason these figures are not higher is the “small-country bias” whereby aid intensity tends to be greater for countries with small populations. This is shown in the second panel of

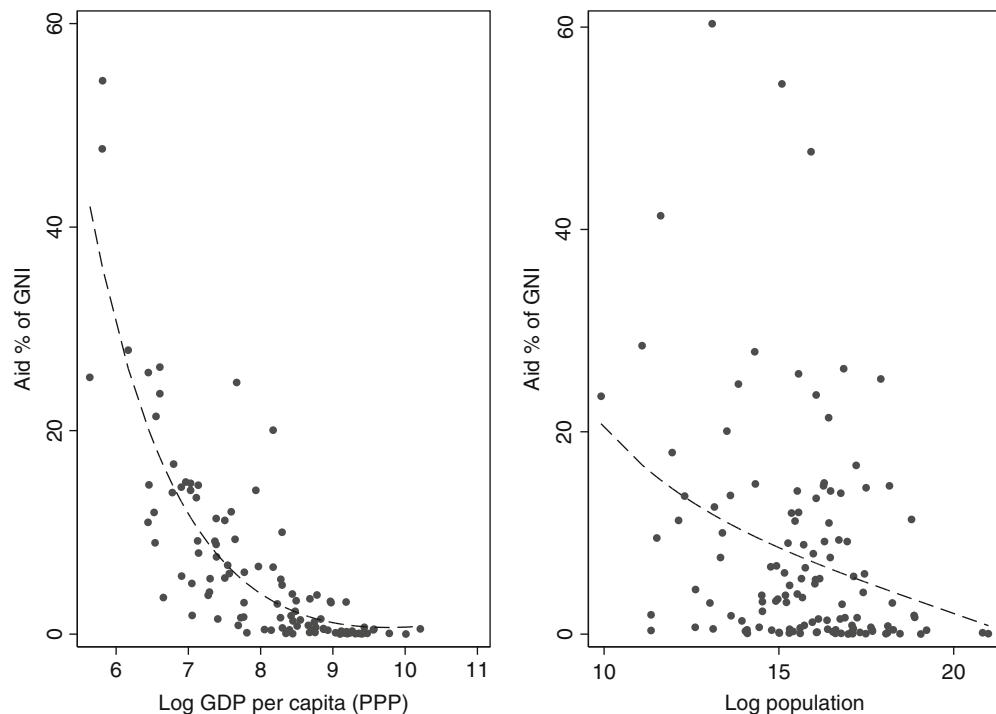


Figure 4 Aid intensity by *GDP per capita* and population size, 2006. Source: World Development Indicators (April 2008 edition). The first panel shows the correlation between aid intensity (aid as a ratio of gross national income) and log *GDP per capita* in PPP terms. The second panel shows the correlation between aid intensity and log population. Both figures are drawn for 2006. The dashed line shows a nonlinear regression fit.

Figure 4, again for 2006 (but including more countries) and where the dashed line is a nonlinear regression fit.

There are some other well-known features of aid allocation. Some donor countries are especially likely to be relatively generous toward their former colonies. More generally, it is widely acknowledged that aid has been used with political and strategic ends in mind (see, e.g., Alesina & Dollar, 2000). The US has traditionally given substantial assistance to Egypt and Israel. During the Cold War, aid was sometimes directed toward regimes allied with the West, to stop countries going over to “the other side” (Kanbur, 2006). In one of the most infamous examples, Zaire under Mobutu received substantial aid, despite skepticism that the money would be used effectively. In the early 2000s, the strategic aims of aid have reconfigured around the “war on terror,” with greatly increased aid to Afghanistan and Pakistan.

Some criticisms of aid become most relevant when it represents a high share of the recipient’s income, and of government expenditure. Figure 5 shows how the share of

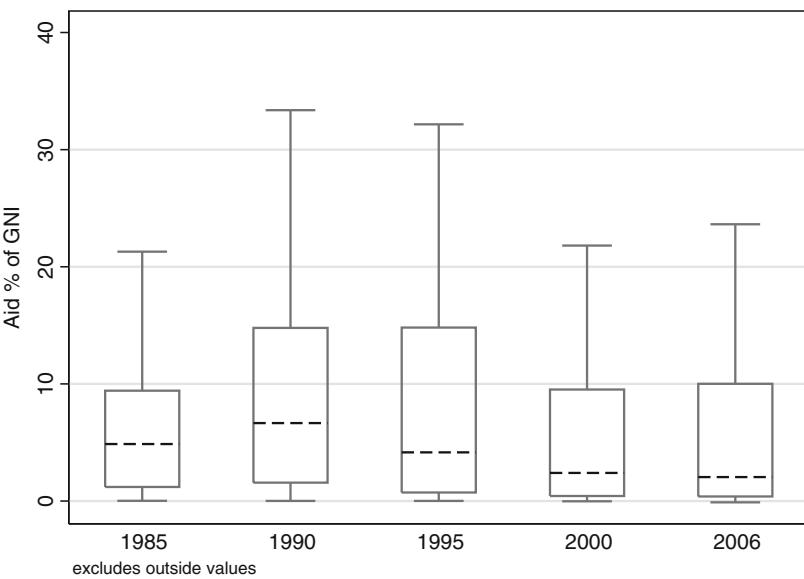


Figure 5 Distribution of aid intensity for 78 countries, 1985–2006. Source: World Development Indicators (April 2008 edition). The box-plots show the distributions of aid intensity for a consistent sample of developing countries with available data, all with populations greater than 250,000 in 1960. The top of each enclosed box corresponds to the 75th percentile for that year, and the bottom of each box to the 25th percentile, while the dashed line indicates the median.

aid in the gross national income of recipient countries has evolved, using Tukey box-plots; the lower and upper ends of each box signify the 25th and 75th percentiles for that year, while the dashed line corresponds to the median. Note that aid is more than 10% of national income for a substantial number of countries.⁶ Table 2 lists the top 20 aid recipients, not in terms of total aid received, but in terms of aid intensity (aid as a share of gross national income).

It is also interesting to consider the generosity of aid from developed countries, relative to their incomes. The long-term aid target for developed countries has been 0.7% of GDP, but most of the large donors (in population terms) are some way short of this, and donors vary in their current targets. Many are committed to increasing the generosity of aid, but past commitments have not always been met. The general pattern is that smaller donors have led the way in terms of generosity. In 2004, the US government allocated 17 cents per US citizen per day to foreign aid, the UK 31 cents, France 33 cents, Denmark 92 cents, and Norway \$1.18 (Gupta et al., 2006). As for the 0.7% target, this has always been somewhat arbitrary, and perhaps no more than a political device through which pressure can be placed on governments.⁷ In this it seems to have failed, and was perhaps never likely to capture the public imagination. The Millennium Development Goals are likely to form a more effective way of mobilizing public support.

Table 2 Average aid intensity, 1997-2006

Country	Aid/GNI
Timor-Leste	0.455
Guinea-Bissau	0.376
Afghanistan	0.322
Liberia	0.316
Eritrea	0.315
Sierra Leone	0.275
Burundi	0.275
Mozambique	0.266
Congo, Dem. Rep.	0.220
Malawi	0.211
Rwanda	0.203
Nicaragua	0.183
Guyana	0.178
Mongolia	0.172
Zambia	0.168
Mauritania	0.165
Lao PDR	0.159
Madagascar	0.144
Mali	0.143
Bhutan	0.142

Source: World Development Indicators (April 2008 edition). The table shows the top 20 aid recipients by aid intensity, based on the average ratio of aid to gross national income, over 1997–2006. The smallest countries, defined as those with populations smaller than 250,000 in 1960, have been excluded.

Uninformed observers might guess from these tables and charts that aid provides a substantial share of the income of the developing world. This conclusion is mistaken, partly because aid flows are often directed toward the poorest countries, and those with the smallest populations. Once we aggregate aid and income across developing countries, the share of aid flows in the income of the developing world is relatively

low. A reasonable rough estimate would be that aid currently accounts for 1% of the income of the developing world. The share is substantially higher for Africa, with its disproportionate number of poor countries with small populations. It is also worth noting several reasons why aid intensity might increase over time (Collier, 1999). First, income has been growing in the donor countries. Second, as some countries grow richer, they will join the list of donors. Third, as the number of aid recipients falls, the ratio of aid to recipient income is likely to increase. These effects will be partially offset by relatively fast population growth in the developing world, however. And although much recent commentary has anticipated a scaling-up of aid flows, these plans may be undermined by the severe financial crisis of 2008, and the associated deterioration in public finances around the world.

Official aid, the main focus of this chapter, is supplemented by private giving, through charities. One estimate for 2001 indicates the charitable contribution from OECD member countries was about \$17 billion, compared to ODA of \$52 billion (see Atkinson, 2005a). And a major change of the last 25 years has been the increasing importance of NGOs, partly funded by private donors and partly by aid agencies and governments. Riddell (2007) estimated that, by 2004, the total value of NGO-funded activities was equivalent to around 30% of ODA. As examples, he cites estimates that NGOs and other nonstate organizations are responsible for around 40% of healthcare and education services in Ghana, Kenya, and Zimbabwe, and points out that one NGO, the Bangladesh Rural Advancement Committee, provides basic health services that cover the majority of its country's population. The study by Arndt, Jones, and Tarp (2007) of one highly aid-dependent country, Mozambique, estimated NGO aid at around 2% of GDP. The study of NGO provision is hampered by a lack of data and the heterogeneity of the sector. Yet its growing importance has the potential to influence long-standing debates on issues such as fungibility and public sector provision.

2.3 The forms of aid

Aid has often been tied to the purchase of goods and services from the donor country. This is most obvious for technical assistance, which is partly used to purchase the services of external consultants. Other instances arise when aid is used as a means of export promotion; for example, much food aid is purchased from the farmers of the donor. Tied aid is generally seen as inefficient, and one widely cited estimate is that the tying of aid reduces its real value by 15–30%, and even more in the case of food aid. The bilateral donors vary widely in the extent to which their aid is tied. Some are publicly committed to reducing tied aid, or have already done so, but the evidence on the overall trend depends on the treatment of technical assistance; see Riddell (2007, pp. 98–101) for discussion and references. One important practical question is the extent to which aid tying makes it easier to gain political support for aid programs:

for example, a high proportion of US aid is tied, but the supposed benefits for the US economy have often been used in making a political case for external assistance. Without tying, US aid might be lower.

Aid in kind, especially food aid, may have distinctive effects. The usual criticism is that food aid, by lowering food prices in the aid recipient, will undermine its domestic agriculture. But the effect on households will vary, depending on whether they are net buyers or sellers of the type of food in question. In a “revisionist” investigation of Ethiopia, at the household level, Levinsohn and McMillan (2006) find that net buyers of wheat are poorer than net sellers, there are more net buyers at all levels of income, and poorer households benefit proportionately more from a lower price of wheat. This implies that many households, and especially the poorest, would benefit from food aid.⁸ In practice, the relative importance of food aid has declined, so that it now represents no more than 2% of ODA (Gupta et al., 2006).

Technical assistance continues to account for a substantial fraction of aid spending; Easterly and Pfutze (2008) report that, across bilateral donors, the mean share of technical assistance in their aid is 24%, with a standard deviation of 18%. This form of aid remains especially controversial. In its favor, compared to other forms of aid, it should be largely immune from the harmful effects associated with Dutch Disease and political economy considerations. But purchasing advice from overseas consultants diverts aid budgets back to the citizens of rich countries, does little to promote domestic capacity, and represents a form of aid that is mostly supply-driven. White (1998, p. 86) notes the irony in the lack of a market-based approach. Similarly, Eaton (1989) observes that technical assistance raises the same issue as other transfers in kind: why not give money, so that recipient governments can decide what assistance to purchase at world prices? He notes that transfers in kind might be justified if the donor has more expertise in identifying and meeting the technical needs, if the donor has stronger incentives than private consultants to transfer relevant skills, and if the quality of external consultants is not readily observable to recipient governments. Arndt (2000) provides another detailed discussion. At a minimum, the criticisms suggest that cross-country analyses should disaggregate aid, and treat technical assistance as a separate category.

2.4 Fungibility

A central issue in the analysis of aid is that flows may be “fungible,” in the sense that they can be folded into the wider plans of recipient governments. When analyzing aid that is notionally earmarked for a specific use, the government’s other activity cannot be taken as given, since any given unit of funding can be substituted for any other. Governments may well adjust their plans, and the net effects of the project in achieving the donor’s objectives may differ from those of the project analyzed in isolation. At one extreme, the net effect of aid earmarked for health and education projects might be to finance government spending on arms, or Mercedes limousines, or transfers to a

favored social group, leaving total health and education spending unchanged. There are limits to the potential for such reallocation, especially once aid represents a large share of the government budget. But the idea that aid is fungible calls into question the wisdom of relying solely on discrete, donor-financed projects, for the implementation of aid and the evaluation of its effectiveness. It has been one motivation for the recent shift from project aid to program aid and direct budget support.

If aid is fungible, the scope for using aid to good effect may be constrained, especially in countries where leaders are primarily committed to goals other than development. But studies of individual countries and projects vary in their conclusions. Pack and Pack (1993) suggest that fungibility may have thwarted donor intentions in the Dominican Republic, but the same authors find no such evidence for Indonesia, where aid generally appeared to have the net effects that donors intended (Pack & Pack, 1990). A recent detailed study of a road rehabilitation project in Vietnam indicated fungibility within the sector—some project aid may have been spent on new roads, rather than rehabilitation of existing roads—but a significant net effect of the project on kilometers of road repaired or built, indicating that aid did indeed “stick” to the sector (van de Walle & Mu, 2007).

Using a cross-country approach, Mishra and Newhouse (2007) find a significant effect of health aid on total health spending, which suggests that aid earmarked for health is not fully fungible. Dreher, Nunnenkamp, and Thiele (2008) find parallel effects of aid on primary schooling. But the best-known empirical analysis of fungibility is that of Feyzioglu, Swaroop, and Zhu (1998), using cross-country panel data on aid flows and government expenditure composition for 1971–1990. They concluded that concessionary loans to the agricultural, education, and energy sectors were substantially offset by reductions in government spending in those areas, so that aid did not “stick.” Although this study has been widely cited, the cross-section dimension is small (38 countries) and the robustness of the findings has rarely been investigated by other researchers. There is clearly scope for more research, especially given that the extent and seriousness of fungibility may vary across recipient governments, and with the extent of aid.⁹ In extreme cases, where a recipient government is initially spending little on social sectors, fungibility is much less of a concern. The model of Cordella and Dell’Ariccia (2007) indicates that project aid is preferable to budget support when the aid program is substantial relative to the recipient government’s resources.

2.5 Debt relief

One form of aid that deserves special consideration is debt relief. Among the general public, the case for debt relief is often regarded as self-evident: poor countries should not be making payments to rich countries or international institutions, and the granting of debt relief should not be classed as aid. An economist’s perspective is that debt relief, like an outright grant, represents an increase in the permanent income of the recipient

and involves an opportunity cost for the donor, with consequences for other aid recipients. From this perspective, debt relief implicitly penalizes those countries which have repaid their debts successfully, and may also increase the likelihood that countries will build up unsustainable debts at some point in the future, in the expectation of future debt relief (the “moral hazard” problem). Yet debt relief remains a significant component of total aid: Bourguignon et al. (2009) estimate that debt forgiveness grants and rescheduled debt in 2002 were equivalent to 10% of that year’s net ODA.

One answer to the above concerns is that a debt relief initiative, such as that announced at the 2005 Gleneagles Summit of the G8, may involve some degree of “additionality,” raising the present value of aid transfers compared to a counterfactual. But to expect full additionality seems naive, unless it is much harder to mobilize public and donor government support for other forms of aid. This is conceivable, but perhaps only if there is widespread misperception of opportunity costs. In practice, the G8’s multilateral debt relief initiative (MDRI) will have consequences for donor budgets until the middle of this century, and the actual extent of additionality must be regarded as highly uncertain. Gunter, Rahman, and Wodon (2008) discuss the issues further. They also note that the significance of debt relief for achieving wider aid goals can be exaggerated: for example, the MDRI implies average debt-service savings of \$1.2 billion a year over 2006–2015, compared to net ODA to Africa of around \$23 billion in 2003.

One set of arguments for debt relief is that eliminating “debt overhang” will raise investment, and improve the incentives to choose policies with long-term benefits. Imbs and Ranciere (2007) provide a recent discussion and references. They present evidence that debt overhang matters in practice, but that debt relief has not been effective in moving countries out of overhang zones. Similarly, Depetris, Chauvin, and Kraay (2007) find that debt relief is relatively responsive to country policies and institutions, but not especially responsive to the debt burden, which suggests that reducing debt overhang is not the primary motivation in practice. More generally, the many historical instances of debt relief raise the fundamental question of when lending is the right approach. Debt relief effectively converts earlier loans into grants, and the status quo has sometimes risked an unsatisfactory hybrid, in which relative transfers are determined in a haphazard and somewhat perverse way. Radelet (2006) argues that aid to the poorest countries, in particular, should be exclusively in the form of grants. Jayachandran and Kremer (2006) provide a recent analysis of loan sanctions designed to prevent dictators from borrowing internationally, both from international agencies and private banks, as in the legal idea that “odious” debt should not be respected by successor regimes.

Some of these issues lead naturally to considering the roles and policies of the multilateral institutions: the World Bank, IMF, the UN agencies, and the regional development banks. In order to limit its coverage to manageable dimensions, this survey will not discuss the roles and policies of specific institutions in detail, or various

proposals for reform. Eaton (1989) and Moss (2007) include brief overviews of the history and roles of the World Bank and IMF, and Moss gives suggestions for further reading. Vreeland (2007) reviews various debates on the development role of the IMF, while Rodrik (1996) provides a conceptual discussion of the role of multilateral lending. One specific debate, associated with a recommendation of the Meltzer Commission, concerns whether the World Bank should cease lending to middle-income countries. For a defense of this lending, see de Ferranti (2006).

3. AID AND GROWTH IN THEORY

One argument for aid is that it will hasten the economic transformation of developing countries. Are there plausible circumstances in which aid can achieve this? This section will review the implications of formal growth models. This will be foreign aid seen through the lens of growth economics, with the strengths and weaknesses that implies. But before pursuing this in more detail, it is worth remembering that aid can be justified even in the absence of productivity improvements. A simple utilitarian perspective suggests that taxing individuals in rich countries, in order to raise consumption in poorer countries, could be welfare-improving even when the effect is confined to raising consumption, and even if some aid is wasted. Indeed, for countries that are structurally disadvantaged, perhaps by being remote from world markets, this direct effect of redistribution on steady-state consumption may be the most that aid can achieve.¹⁰

3.1 Capital shortage and neoclassical growth models

The idea that aid can remedy a capital shortage is almost as old as the idea of aid itself. It has often motivated the use of financial programming models to quantify a gap between domestic saving and investment needs. It is obvious that such an approach is too mechanical, but a useful way to expose some of the flaws is to begin with the Ramsey growth model. This is not because of any descriptive accuracy, but because it tells us how an optimizing social planner *should* respond to a permanent resource transfer.¹¹ The social planner chooses the balance of consumption and investment at each date, to maximize the discounted lifetime utility from consumption, using income from production and aid inflows, but without access to other international capital flows. Obstfeld (1999) uses this model to investigate how aid transfers should be divided between consumption and investment. If the economy is already on its steady-state path, the transfers should be used to finance additional consumption, rather than used to increase the capital stock. This is because the long-run marginal product of capital is determined by structural parameters such as the discount rate and the depreciation rate, and these are invariant to the size of the transfer. But if the economy is currently below the steady state, using part of the transfer for investment will increase the speed of convergence to the steady state, and raise welfare.

These results are sometimes misinterpreted as saying that aid is largely ineffective in neoclassical growth models, or even that aid which finances consumption is wasted. This is a misunderstanding, because the Ramsey model tells us how a social planner should respond optimally to transfers. In this model, the benefit of aid must be either to raise consumption now, or to raise it later. Hence, transfers will often be used to finance consumption today, rather than invested in order to raise consumption in the future. In practice, higher consumption might be associated with outcomes such as better nutrition and health status. The tendency to undervalue such gains arises when aid is expected to be transformational, a case that is not exactly bolstered by the model's invariance of long-run productivity to aid flows.

Although transfers raise the speed of convergence, the welfare benefits seem modest. Obstfeld (1999) simulates a Ramsey-type model with some liquidity-constrained consumers and some that can optimize. The growth effect of an unanticipated, permanent increase in aid is small, and in his experiments, a high proportion of the aid is consumed rather than invested. Arellano, Bulíř, Lane, and Lipschitz (2009) obtain similar results in a model with stochastic productivity and transfers.¹²

These analyses are stylized, but any case for aid that is based on raising investment faces more general problems. As many have noted, the presence of diminishing returns means that investment rates cannot readily account for the vast international disparities in living standards. For example, with a Cobb-Douglas production function, the elasticity of steady-state output with respect to the investment rate is $\alpha/(1 - \alpha)$, where α is the capital share. With a capital share of one-third, steady-state output is proportional to the square root of the investment rate. Aid-financed increases in investment will quickly encounter diminishing returns, unless it is legitimate to define capital more broadly. The impact of an increase in investment, financed by aid, will be relatively modest: worth having, but not enough for economic transformation.

But there are deeper questions that need to be asked. How should the flows of aid enter the model, and how should outcomes be linked to aid? In the earlier discussion, aid transfers have been treated as if dropped from a helicopter, as in Milton Friedman's thought experiment about an exogenous shock to the money supply. A more satisfactory account of aid would need to assign stronger roles to the main actors—governments, households, and firms—since it will be their decisions and capacities that ultimately determine the effects of aid. For example, the existing models rarely have much of interest to say on the analytical issues grouped as the “private sector response” to aid. Related to this point, the programming and twin-gaps approaches never clarified why investment is too low in the first place. This opens up the important criticism of aid programs advanced by Bauer:

If all conditions for development other than capital are present, capital will soon be generated locally, or will be available to the government or to private businesses on commercial terms from abroad, the capital to be

serviced out of higher tax revenues or from the profits of enterprise. If, however, the conditions for development are not present, then aid—which in these circumstances will be the only source of external capital—will be necessarily unproductive and therefore ineffective. (Bauer, 1971, pp. 97–98)

This raises a fundamental quandary for believers in capital shortage. If all the conditions for successful capital accumulation are in place, why is capital not already forthcoming from domestic and foreign investors? This argument has become more important, not less, over time. International capital flows have increased substantially, and Cohen and Soto (2002) and Caselli and Feyrer (2007) even suggest that the return to capital has been broadly equalized around the world. And observed capital flows are often in the opposite directions to those predicted by the textbooks: see Prasad, Rajan, and Subramanian (2007). This is not the pattern that would be expected, if investment opportunities were plentiful in poorer countries.

Any growth economist discussing aid needs to address Bauer's paradox. There are at least four interesting scenarios in which aid could achieve things that private capital cannot. These arise from models of poverty traps; the possibility of virtuous circles; growth that is stymied by one or more binding constraints; and the role of complementary inputs, such as infrastructure or institutions. Although these possibilities are sometimes conflated, it is worth keeping them analytically separate, and the remainder of this section will discuss each in turn.

3.2 Poverty-trap models

In models of poverty traps, low productivity is one possible outcome, but a better equilibrium is also attainable, given the underlying fundamentals. In principle, aid inflows might advance a recipient economy beyond a “tipping point,” either switching the outcome from one equilibrium to another, or eliminating the low-income equilibrium altogether. Supporters of large increases in foreign aid sometimes appeal to arguments of exactly this kind, as in Sachs et al. (2004) and Sachs (2005). There are a number of mechanisms by which multiple equilibria can arise, and they are at least as old as the Solow model: both Nelson (1956) and Solow (1956) contain early discussions. If we take the Solow model as a starting point, multiplicity can arise when population growth or the saving rate vary with income *per capita*. Other models emphasize threshold effects and production externalities, and the literature now includes many other examples of poverty traps. As in some other areas of economics, the surfeit of candidate models hints at a fundamental difficulty: how many have testable or falsifiable implications, and how can we discriminate among them? In practice, many assessments of poverty-trap models confine themselves to assessing the plausibility of key assumptions.

The existence of a poverty trap is not sufficient to justify aid. Whether or not aid is the right response will depend on the origins of the trap. Some poverty traps might arise through weak governance, or because opportunities for rent-seeking have led to

a powerful constituency against reform. These models may ultimately prove more insightful than somewhat mechanical variations of the Solow model. Yet within this alternative class of models, aid may be ineffective as a means of attaining a better equilibrium, and could even make things worse.

Another open question is whether even the standard poverty-trap mechanisms can justify scaling-up aid, once explored in more quantitative terms. To illustrate with the simple case of endogenous fertility rates, it is important to know the sensitivity of fertility rates to income *per capita*, the contrast between equilibria, in terms of the differentials in income *per capita*, and the magnitude of the aid needed to move the economy out of the poverty-trap equilibrium. There has been remarkably little work of this kind, but the analysis in Kraay and Raddatz (2007) demonstrates that it could be productive. They find relatively little support for two specific poverty-trap mechanisms, namely endogenous saving rates and threshold effects on productivity. A possible extension would be to study alternative poverty-trap models, as in Graham and Temple (2006). There might be an especially strong case for investigating traps that arise through political economy considerations, to quantify the extent to which aid then has negative effects.

Easterly (2006c) argues that the broad stylized facts of development are inconsistent with poverty-trap models. Even the countries in the lowest quintile of GDP *per capita* in 1950 had seen, by 2001, their GDP *per capita* increase by 2.25 times on average. There are few countries where GDP *per capita* has genuinely stagnated, and the economic history of most developing countries is a more complex story of unsteady growth, short-lived booms, and major collapses in output (Little, Cooper, Corden, & Rajapatirana, 1993; Pritchett, 2000). But it may be that testing for outright stagnation takes the idea of a poverty trap too literally, as Easterly acknowledges. Consider what happens if multiple equilibria arise within the Solow model. The low-income equilibrium can display positive growth, if total factor productivity grows over time. This could eventually move an economy across a threshold, or even eliminate the low-income equilibrium. But the process is likely to be slow, and hence a justification for aid would remain.

Choosing a starting point later than 1950 provides more support for poverty-trap accounts, reflecting the many instances of slow growth in the 1980s and 1990s. Recent experience might suggest that the world's economies are sorting into distinct sets of winners and losers, where the latter are trapped in poverty. This view of the world is emphasized in Collier (2007). If his arguments are correct, the set of countries most in need will gradually change in composition: the low-income countries without major structural disadvantages will grow out of poverty, leaving behind a group where the obstacles to development are most fundamental and hardest to overcome. Some of the existing challenges and dilemmas faced by donors will only get worse.

3.3 Virtuous circles and amplification effects

In poverty-trap models, the transition from one equilibrium to another would often give rise to sudden growth. For this reason, the poverty-trap ideas are sometimes conflated with a separate hypothesis, that of a virtuous circle. Even if the equilibrium at any point in time is unique, it is possible that growth can itself promote and reinforce the conditions needed for further growth. Aid flows that initiated such a process would have a more dramatic effect than in the neoclassical model. For example, the overall impact of higher investment can be amplified by the endogenous accumulation of human capital, as in Mankiw, Romer, and Weil (1992). They show that the elasticity of steady-state output with respect to the investment rate may be higher than in the standard Solow model, because increased investment ultimately leads to higher educational attainment. More generally, all that is needed is that rising income induces changes which have benefits for income. Other amplification effects could arise through endogenous institutional change, improvements in the quality of policy, or diversification that lessens exposure to shocks. When these effects are incorporated in a neoclassical growth model, anything that raises the capital stock, including aid inflows, will have a larger effect on steady-state productivity than in the textbook version of the neoclassical growth model.

From a conceptual point of view, these mechanisms also answer Bauer's objection. Most amplification effects can be represented as intertemporal externalities. It may be that higher productivity leads to better institutions in the future, but this is irrelevant to the investment decisions of individual firms or potential entrepreneurs, and they will not internalize these wider benefits when making their decisions. From society's point of view, there will be underinvestment.

Thinking about aid and growth in terms of virtuous circles continues to hold some sway over development commentary. It is otherwise hard to explain the longevity of Rostow's metaphor of "take-off" into sustained growth, with its implication that getting off the ground is harder than staying in the air. But the empirical relevance of amplification effects is underresearched. DeLong (1997) is a rare example of a paper that takes them seriously, and indicates how progress might be made. Compared to models of poverty traps, a scientific approach seems easier in this case. Since amplification effects are consistent with a unique equilibrium, they can be examined using standard empirical methods, as in Mankiw et al. (1992). A convincing demonstration of their relevance would strengthen the case for aid.

3.4 Binding constraints

The ideas discussed thus far are somewhat abstract, and bear little relation to donor policies in practice, or the controversies over the Washington Consensus. The policy orthodoxy of the 1980s and 1990s came under attack from many directions. At least

as a growth strategy, that orthodoxy was sustained partly by the lack of a credible alternative. Since economic policy abhors a vacuum, there has been a search for better ways to conceive growth strategies and policy advice. The leading candidate, now widely discussed within donors, frames the growth problem in terms of addressing binding constraints (Hausmann, Rodrik, & Velasco, 2008; Rodrik, 2007a). This has interesting implications for aid policies, which are worth discussing in detail.

The approach begins by criticizing reform efforts which proceed on too many fronts at once. Rather, the aim should be to identify precisely what is holding growth back, and target reforms where they will have the greatest marginal benefit. Since the marginal effects of policies are not independent, one or more constraints on growth may be “binding” in the sense that reforms elsewhere will have limited benefits, at least until the key constraints have been addressed. To the extent that donors can identify binding constraints, and then direct aid inflows and technical assistance at relaxing them, the effect of aid may be powerful. Again, aid may be able to achieve things which private capital flows cannot.

This has some echoes of the early literature on foreign aid. Influenced by Europe’s experience under the Marshall Plan, development economists identified low saving rates and a lack of foreign exchange as the key binding constraints on economic growth. By relaxing these constraints, aid could achieve rapid development. But if the intervening decades have taught growth economists anything, it is that development and its constraints have to be understood in more sophisticated terms.¹³

In principle, the binding constraints perspective could improve aid effectiveness in two main ways. First, within donors and recipients, it should concentrate efforts on those areas of the economy where aid and external advice can make a difference to growth. Second, it may become clear that some binding constraints cannot be addressed by aid, with implications for the allocation of aid across countries. For example, the recent growth literature often identifies institutional weaknesses as a key constraint. It is not clear that aid and conditionality can be used to build better institutions. In that case, the appropriate response would be to redirect resources toward countries with good institutions and binding constraints that are easier to address. The characterization of growth in terms of binding constraints may turn out to be a natural partner for selective aid allocation, rather than ensuring that aid is effective for every possible recipient. In countries where binding constraints cannot be relaxed, aid should be directed at humanitarian goals rather than seeking economic transformation.

The new approach is especially attractive in providing concrete advice on the diagnosis of constraints. Hausmann, Rodrik and Velasco (2008) provide a decision tree which frames the relevant questions, and patterns of symptoms. In developing the implications for aid, one consideration is that different constraints will apply at different horizons. An economy might currently be growing slowly because of macroeconomic disarray, when the key medium-run constraint is inadequate infrastructure or a lack of

human capital. It might be easier and more appropriate for donors to address these medium-run constraints, rather than attempting to influence macroeconomic policy.

Less obviously, the consequences of misdiagnosis are also relevant. It is not certain that a binding constraint should immediately be targeted. One constraint on growth might be that firms find it difficult to access credit, but financial liberalization brings risks of its own, especially in some institutional environments. The need for robustness implies that it will sometimes be preferable to address constraints other than the one most obviously binding. But the approach accommodates this problem more easily than some of its precursors, and seems less risky than wholesale reform based on generic recommendations, partly because it allows scope for gradualism and experimentation.

It is easy to see why this perspective has quickly become popular.¹⁴ Its greatest strength is its practicality, and it will stand or fall on the success of the associated policy advice, rather than abstract debates about alternative conceptions of development. Given the importance of such advice, and the implications for conditionality and aid allocation, the investigation of these ideas should be a research priority. Statistical work may help to establish the most relevant constraints, together with other evidence, such as surveys of firms.¹⁵ One approach would be to look at transitions between different growth regimes within countries, as in Hausmann, Pritchett, and Rodrik (2005). It is too early to say whether statistical work will be reliable, and it may quickly hit the limits of the data. Hausmann, Klinger, and Wagner (2008) offer some interesting suggestions and insights on other ways to identify specific constraints.

3.5 Complementary inputs

Another answer to the Bauer paradox is that some inputs relevant to the private sector will not be adequately provided by the market. The most obvious example is certain kinds of infrastructure, an area where governments have usually taken the lead due to a variety of market failures. Inadequate infrastructure is an obvious candidate for a binding constraint, and for the diagnostic methods recommended by Hausmann, Klinger and Wagner (2008): a lack of reliable electricity supplies can be a significant constraint on the manufacturing sector, something that could be assessed for a given country by examining the incidence of private generators, and whether sectors and techniques intensive in electricity use are less well represented than elsewhere. For formal models of growth with a role for aid-financed public investment or infrastructure, see Agénor and Yilmaz (2008), Chatterjee, Sakoulis, and Turnovsky (2003), and Chatterjee and Turnovsky (2004, 2005).

It is also possible to conceive of the institutional environment as a particular form of infrastructure, and one that will have consequences for entry, production, and investment decisions. Put differently, institutions are a complementary input for the private sector. This would be a powerful answer to Bauer's paradox, if only more was known

about building better institutions. One theme of the later discussion will be the connections between aid and governance, both positive and negative.

3.6 Growth versus basic needs

So far, this section has sought to establish circumstances in which aid might transform the productivity of developing countries. A range of answers to the Bauer paradox are possible, and still others can be advanced (see Eaton, 1989). For some, however, the emphasis on transformation is the wrong place to start. It asks too much of our current understanding of growth, and of the capacities of donors and recipient governments. Perhaps it would be better to emphasize the humanitarian role of aid, and concentrate on practical steps to improve health and educational attainment (e.g., Banerjee, 2008). The debate here is largely about expected payoffs. Randomized trials can be used to identify interventions that have a high probability of success. But some of these interventions will have a modest expected payoff compared to a change in growth policy, because the benefits of achieving growth are large enough to outweigh the low probability of success.

Assessing these probabilities has to be rather subjective, so positions in this debate are largely a matter of intuition and gut instinct. Views may differ on whether the conventional wisdom on growth is sometimes actively harmful, for example. Not enough is known about the trade-off between addressing basic needs and promoting economic growth, and the extent to which they are mutually reinforcing. The pursuit of immediate humanitarian goals, such as better health and education, might ultimately promote growth and beneficial forms of social change. The available evidence suggests that education and cognitive skills have growth benefits (Hanushek & Wößmann, 2008). There is less evidence that increased life expectancy promotes growth (Acemoglu & Johnson, 2007). Later sections of this chapter will discuss policies to address basic needs, together with the potential roles of agriculture and rural development.

3.7 Summary

According to neoclassical growth models, the goals of aid should not be defined in overly ambitious terms. Although it is not difficult to construct models in which aid could have more powerful effects, evaluating their relevance is harder. To some extent, the discussion reaffirms the earlier conclusion of Eaton (1989, p. 1374) that “there may be too many rather than too few methodologies for analyzing this issue.” It also mirrors the widespread view of growth economics, that the field is theory-rich and data-poor. To borrow an idea from Malthus, the number of models continues to grow exponentially, while the data needed to discriminate between them are increasing only arithmetically. The solution may lie in growth models with more detailed microeconomic foundations and direct links to micro data.

In contrast, the existing models are usually narrow in scope. The analyses are often framed in the context of one-sector models, when multisector models, as in the dual

economy tradition, might have different implications. The mechanisms linking aid and productivity are somewhat mechanical, and this approach cannot say a great deal about benefits for health, education, poverty, and food security. Until recently, the aggregate models have typically omitted considerations such as governance and political economy that are central to debates about aid effectiveness. With all this in mind, the remainder of the chapter will largely abandon the perspective of aggregate growth economics in favor of these wider questions.

4. HAS AID WORKED?

This section will review some of the evidence on the effects of aid. As noted in the introduction, it is hard to establish definitive success stories. The Marshall Plan is sometimes cited, but as Bauer (1969) noted, that was largely about rebuilding something that had previously existed, at least in economic and institutional terms. Today, aid agencies face more complex problems and dilemmas, and must work on a wider variety of fronts. Accordingly, the assessment of aid effectiveness has become ever more complicated. Partly because of this, there are some who regard this effort as a distraction from the urgent business of helping others lead better lives. Yet an important message of the aid literature is that the unexamined gift is sometimes not worth giving.

If there is a single piece of evidence which threatens to overwhelm all others, it is the observation that ratios of aid to GDP are often highest for African countries, without a clear payoff in terms of sustained growth. But even this simple picture becomes more complex on closer inspection. Many African countries have recorded improvements in infant mortality, school enrolment and, before HIV-AIDS, in life expectancy. How much of this can be attributed to aid is highly uncertain, but the fact that the question can be posed at all is important. It is a reminder that productivity is not the only relevant consideration. When aid is spent on antimalaria bed nets, medicines, or cleaner water supplies, it may have benefits for living standards that are never captured in GDP statistics, or that are reflected in higher productivity only many years later.

4.1 Project evaluations

For a long time, the most widely discussed evidence came from project evaluations, often carried out by aid agencies, and based primarily on cost-benefit analyses and reports on completed projects. Cassen and Associates (1994) provide a useful review of the main findings. Internal evaluations by the World Bank for the 1960s and 1970s often found relatively high rates of return, but typically lower in Africa than in Asia. The more recent summary of Riddell (2007, pp. 180–185) emphasizes the high success rates indicated by project evaluations, with some evidence of improvement over time. The obvious questions are whether the criteria for success are defined precisely and objectively enough for these evaluations to be fully credible, and whether

longer term assessments would be equally favourable. The positive results from the evaluation of individual projects seem to conflict with the more ambiguous evidence that has emerged from aggregate aid-and-growth regressions, and this led Mosley (1986, 1987) to define a “micro-macro” paradox.

In recent years, assessments of aid effectiveness have given less emphasis to project evaluations. This reflects a partial shift from project-based aid to program aid; doubts about the sustainability of projects, once support has finished; and greater awareness of the limitations of a partial equilibrium approach. Mosley’s micro-macro paradox might be explained by a general equilibrium response to aid, such as the Dutch Disease, or political economy mechanisms. If aid is at least partially fungible, the limitations of individual project evaluations become even more apparent. Yet, despite all these points, something valuable may have been lost. Even though some individual assessments of success or failure will be unreliable, the project evidence as a whole can be informative, especially in discriminating between different types of project, and effectiveness in different contexts. The kind of detailed review carried out by Cassen and Associates (1994) has largely disappeared from the academic literature, and this is a significant loss. As in the Edgar Allan Poe story “The Purloined Letter,” where a letter is hidden by leaving it openly on view, the project evidence is sometimes ignored, precisely because it is the obvious place to look.

One concern about project-based evaluation, whether *ex ante* or *ex post*, is that of sustainability. Donors often seek projects that will continue to have benefits even after financial support has been withdrawn or scaled down. The record in practice has been mixed. Classic examples are roads that are financed by aid, but then fall into disrepair, and health facilities that are built but not staffed. Pritchett and Woolcock (2008) argue that rural water supply projects have not always achieved lasting benefits, because of poor maintenance of standpipes, and the lack of institutions that could later substitute for donor involvement. Kremer and Miguel (2007) study various ways of promoting sustainability in a deworming program, including school health education and cost-sharing payments from those taking the drugs, but find these alternative interventions to be ineffective. They argue that financial sustainability may often be hard to achieve. The implication is that sustainability should be considered within the cost-benefit framework, not outside it, and caution is needed whenever a favorable evaluation relies on optimistic assumptions about sustainability.

4.2 Case studies

While thinking about aid, it is often suggested that much could be learnt from detailed case studies. This can be motivated in terms of the “thick description” needed to do justice to the context and the many relevant factors that cannot easily be quantified. The risk is that a case study becomes exclusively descriptive, and does little to establish the underlying structure, draw conclusions, or suggest potential generalizations.

To avoid this, one approach is to examine a specific issue: for example, case studies of the relationship between aid and conflict, on a country-by-country basis and then compared, with a similar structure to the collection of Collier and Sambanis (2005) on civil war. The ideal might be country studies coauthored by country experts, economists, and political scientists. But, however well done, one weakness of case studies is that much rests in the eye of the beholder. This can be illustrated using the studies of aid and reform in African countries in Devarajan, Dollar, and Holmgren (2001). The introduction to this volume drew on the individual country studies to reinforce the arguments of the World Bank's 1998 report *Assessing Aid*. Yet, Tarp (2001) takes issue with several of the offered interpretations of the country studies. This hints at the difficulty that arises in synthesizing case study evidence. Different researchers, proceeding in good faith and with a similar intellectual background, may arrive at very different readings of the evidence.

These arguments notwithstanding, case studies are surely undervalued by most economists. It could be seen as a major failing of the discipline that it attaches too little prestige to the development and use of historical skills, and the unhappy outcome is that much of the academic literature on foreign aid relies on cross-country statistical evidence, when analytical and theoretically informed case studies would often be more illuminating.

4.3 Cross-country evidence

No comprehensive discussion of foreign aid should entirely ignore the cross-country evidence. The remainder of this section will examine this evidence and its major weaknesses. Many of the points are relevant beyond the aid-and-growth regressions that have gained most attention. Cross-country empirical studies underpin some influential contributions to the policy debate, and have been used to analyze the effects of aid on conflict, governance, institutions, public expenditure, taxation, and social indicators. It is, therefore, important to understand the limitations of this evidence. Since the evaluation of aid effectiveness is formidably difficult, it is perhaps not surprising that a lot of weight is placed on statistical analyses, but this is sometimes closer to evading problems rather than avoiding them. For reasons that will be discussed in the following section, the available evidence is generally too fragile to bear much weight, even though it has improved over time.

The early statistical research on aid effectiveness focused on the associations between aid and rates of saving and investment. Hansen and Tarp (2000) provide a review of this literature, and conclude that aid and investment are positively associated.¹⁶ In recent years, interest in this association has waned, partly because of recognition that capital shortage theories are too simplistic, and because the quality of investment is likely to vary widely across different policy environments.¹⁷ If the allocation of investment is critical, a more disaggregated analysis is needed. Some of the most

interesting evidence comes from microeconomic data on public and private investment projects: the returns have typically been higher in countries with better macroeconomic policy outcomes, and greater civil liberties (Isham & Kaufmann, 1999; Isham, Kaufmann, & Pritchett, 1997).

But the dominant approach in the literature has been to estimate cross-country regressions that relate growth (or the level of a steady-state growth path) to the ratio of aid flows to GDP, and control variables. Among some analysts, these studies are viewed with a skepticism that has only increased over time. In his extensive discussion of foreign aid, Riddell (2007) devotes just 3 of his 505 pages to growth regressions. His argument is not simply based on the fragility of the evidence. Instead, the framing of the problem is said to distort the terms of the debate from the start. These studies

... fuel the false notion that aid's benign or adverse effects are automatic, divorced from the context in which it is provided, the forms in which it is given, and the conditions attached to its provision. On the contrary, the country-based evidence tells us that ownership, commitment, context and capacity all matter a great deal: indeed they provide the only reliable backdrop against which to judge whether aid works or not. (Riddell, 2007, p. 225)

These criticisms are important. The estimated relationships between aid and growth are reduced-forms, and the reduction eliminates most of the considerations needed for well-informed policy. Another concern is that aid can be an unhelpful analytical category: perhaps we should be thinking about specific programs, in areas such as education, health, and infrastructure, rather than studying the effects of a somewhat artificial aggregate. But it remains worth tempering some of the more dismissive criticisms, and trying to establish where cross-country empirical work might have a role. There are at least two reasons for this. First, starting from data and evidence is preferable, for all its flaws, to working backward from a set of desired conclusions, not a rare event in such a passionate debate.¹⁸ Second, some of the debates on aid effectiveness turn on combinations of macroeconomic or general equilibrium effects that could be hard to detect, other than in cross-country data. Dutch Disease effects are the most obvious candidate: even if aid influences the real exchange rate, the long-run effect on productivity growth is partly an empirical question. Similarly, long-run effects of aid on institutional development and the domestic political process might be hard to detect in other ways. This perspective suggests that the hypotheses in cross-country work should be sharpened, moving beyond reduced-forms to look at intermediate channels and mechanisms. Some recent work has made progress in this direction.

Moreover, the alternative approaches to studying general equilibrium effects are not without problems of their own. The leading alternative, advocated by Obstfeld (1999) among others, is the use of medium-scale structural models. These can be used to simulate dynamic responses to various kinds of transfers, while allowing forward-looking

behavior. What divides opinion here is not the desirability of a structural model, but its current feasibility, and the extent to which measurement should be constrained by theory. Until we know more about the key influences on aid effectiveness, relying exclusively on structural models risks underestimating both our ignorance about what matters most, and the value of reduced-form relationships in addressing that ignorance.

Yet even considered on limited terms, as an imperfect complement to structural models, the current studies have significant weaknesses. Ideally, papers on aid and growth would take great care over the data; would address the endogeneity issues that arise immediately when aid is not randomly assigned; would almost certainly use a panel data approach, to eliminate the effects of country-specific unobservables; and would allow a complex dynamic response to aid, and various forms of heterogeneity. The empirical literature has steadily moved in some of these directions, and is substantially better than earlier work. But few studies satisfy all these requirements simultaneously, especially when examined closely. As elsewhere in the growth literature, the reliability of the findings is limited by the small number of countries available to researchers. The appendix to this chapter discusses the econometric issues in more detail.¹⁹

One point bears especial emphasis. Given the many possible effects of aid, a careful empirical approach requires a sharp distinction between short-run and long-run effects.²⁰ As Collier (2006a) notes in discussing the effects of resource rents, extra time has sometimes revealed additional costs rather than additional benefits. It is a strength of panel data approaches that they can accommodate a response to aid that evolves over time. In principle, a model with sufficiently rich dynamics can allow the long-run response to differ from the short-run response even in sign. Yet, few papers have taken a thoughtful approach to modeling a dynamic response to aid inflows.

4.4 Aid and social indicators

The recent literature has started to follow Boone (1996) in looking at the relationship between aid and social indicators of direct interest. The previously cited studies by Dreher et al. (2008) and Mishra and Newhouse (2007) are examples of this approach, focusing on primary enrolment and infant mortality respectively. Masud and Yontcheva (2005) relate infant mortality and illiteracy rates to bilateral aid and some NGO aid. Gomanee, Morrissey, Mosley, and Verschoor (2005) consider the relationships between aid, infant mortality, and the Human Development Index. Most of these studies use fixed-effect estimators and find that, within countries, increased aid is associated with improvements in social indicators. The problems arising from nonrandom assignment of aid will be less severe than in the growth case, if aid allocation is largely independent of the unobservables most relevant to health or education outcomes. The possible exception is government effort, or commitment to development, which could play an important role and is hard to observe directly. And in other respects, this

form of research faces similar problems to growth regressions. Again, the literature does not distinguish clearly between short-run and long-run effects. Although some problems can be overcome, others are less tractable. Social indicators are typically rather persistent, implying that the amount of information in the time-series data is limited. A further problem is data availability: the data on aid commitments disaggregated by purpose, and on government expenditure by sector, are sufficiently incomplete to be a major constraint on studies of this kind.

4.5 Summary

The cross-country evidence on the effects of aid must be regarded as a work-in-progress. Judged against the criteria set out above, there is not a single paper which succeeds in all dimensions, and it is not difficult to extend the list of problems further. It is also easy to conclude that the empirical literature on aid and growth is less than the sum of its parts. Recent conclusions vary widely, from relatively optimistic (Clemens, Radelet, & Bhavnani, 2004; Dalgaard, Hansen, & Tarp, 2004; Hansen & Tarp, 2000; Minoiu & Reddy, 2007) through to ambivalent (Roodman, 2007a) and skeptical and pessimistic (Doucouliagos & Paldam, 2008; Easterly, Levine, & Roodman, 2004; Rajan & Subramanian, 2008).

Some of the variation arises from modeling choices: positive effects of aid on growth are more commonly found in panel data than in cross sections. Some of the panel data studies allow for fixed effects, and so partially address the nonrandom assignment of aid. But the other identifying assumptions often seem flawed, given the absence of plausible time-varying instruments. The use of instruments is usually more plausible in the cross-section studies, which often follow Boone (1996) in emphasizing strategic links, and measures of country size or influence. Rajan and Subramanian (2008) is one of the best examples of this approach. They sometimes find that aid has a significantly negative effect, but this result is fragile, and aid is insignificant in most of their regressions.

The possibility of a negative effect is worth discussion, given the arguments that aid can be actively harmful. It seems especially important to ask how often panel data studies indicate a negative effect of aid. Rajan and Subramanian (2008) find a significantly negative effect in two of their eight main panel data regressions (their Tables 9 and 10). In other panel data studies, the findings tend to vary between a positive and significant effect, and insignificance. An insignificant coefficient should usually be seen as absence of evidence, not evidence of absence, at least until the economic implications of a confidence interval have been explored.²¹ For now, the panel data evidence rarely supports the claim that aid is actively harmful. It is not clear that we can go much further, however. Although the attempt to quantify the effect of aid should be central to this literature, it is hard to draw conclusions when the distinction between short-run and long-run effects is so rarely made.

Easterly (2008, p. 18) is probably right to say that “the regression wars on aid and growth show no sign of ending any time soon.” The different sides have worked on their firepower, but there is a lack of consensus even on basic issues of specification and control variables, and this makes the whole enterprise appear less than scientific. The estimated models vary across different papers, but researchers rarely seek to establish why their findings differ from previous studies. There are signs that this is beginning to change. One of the strengths of Easterly et al. (2004) and Roodman (2007a) is that they explicitly build on previous work and examine the sensitivity of previous findings to specific modeling choices. In the long run, this will increase the coherence and reliability of the evidence, provided there is also scope for the occasional “mutations” that are needed to introduce good ideas or eliminate bad ones.

As noted earlier, one research priority should be to explore channels and intermediating relationships, and the effects of aid on a wider range of outcomes: not only social indicators, but also governance, conflict, public expenditure, taxation, sectoral structure, and political outcomes.²² Although many of these have already been studied, the quality of the statistical work is somewhat variable. Without reliable evidence on these questions, it will be hard to understand the conditions needed for effective aid, or think carefully about how aid should be allocated across countries. The various intermediating relationships will be the focus of the next three sections of the chapter.

5. THE VIEW FROM TRADE THEORY

Does the uncertain association between aid and growth say more about the limitations of the data, or the limitations of aid? It may point to something systematic: perhaps economic and political forces limit the benefits of transfers. The focus of this section will be the general equilibrium effects associated with resource transfers, and in particular the notion of a Dutch Disease. The relevant ideas will be discussed at some length. This is partly because the issues are sometimes misunderstood, and partly because they form an important dividing line between those who advocate increased aid, and those who regard that strategy as risky.

5.1 The transfer problem

The early general equilibrium analyses of transfers emphasized their effect on the international terms of trade. These effects arise because transfers alter the world division of expenditure between countries, and so alter world prices. Induced changes in the terms of trade could either reinforce the beneficial effect of the transfer on the recipient's welfare, or partially offset it. Trade theorists have established conditions under which a transfer of purchasing power could even enrich the donor and leave the recipient worse off. This striking result can potentially arise when a transfer takes place between two countries that also trade with a third; or where a transfer aggravates the effects of

an existing distortion, such as a subsidy, tax, or tariff.²³ It is often argued, however, that the conditions necessary for the transfer paradox are restrictive. And precisely because the “classical” version of the transfer problem relies on endogenous changes in the terms of trade, the effect seems likely to be modest, given that aid flows are small relative to trade flows. In reviewing the literature on the transfer problem, Eaton (1989) suggests that research attention to the issue has been disproportionate. The recent debate on aid has mainly focused elsewhere.

5.2 The Dutch Disease

In the context of foreign aid, Dutch Disease arises when a transfer of purchasing power to a developing country leads its traded goods sector to contract. The simplest analysis of this effect considers a small open economy with two sectors, traded and nontraded.²⁴ The price of the composite traded good is determined by the world market, and is therefore exogenous. In contrast, the price of the nontraded good is determined by domestic supply and demand. This means that an aid transfer, by increasing national income, will drive up the relative price of nontraded goods and increase the returns to factors in the nontraded sector. The typical result will be a real exchange rate appreciation, and a fall in the output of the traded goods sector, as factors of production move into the nontraded sector until equilibrium is restored.

This analysis characterizes the general equilibrium response in real terms. The adjustment to the transfer also has implications for short-run macroeconomic policy. If the nominal exchange rate is fixed, the required increase in the relative price of nontraded goods will have to be brought about by an increase in their nominal price; and hence the combination of a fixed exchange rate and increased aid flows will be inflationary.²⁵ But great care is needed before labeling any of this a disease, or talking about exchange rate “overvaluation.” The effect described is simply an efficient general equilibrium response to a transfer. In the absence of distortions, the aid transfer will be associated with higher welfare. Other forms of transfers, such as remittances from overseas workers to their families, will lead to a similar general equilibrium adjustment.

The usual argument that transfers can be damaging, or at least partially offset by general equilibrium effects, relies on externalities specific to the traded sector. These externalities could be static, as in contemporaneous spillovers within the traded sector. They could also be dynamic, with the possibility of learning-by-doing as a leading candidate. Rodrik (2008b) develops a newer argument, in which the traded sector is disproportionately affected by the institutional and market failures of poorer countries. Given any of these possibilities, an aid transfer could reduce productivity growth, and its welfare effect becomes ambiguous. The initial benefits of the increase in national income could be more than offset by dynamic losses. This implies a link between Dutch Disease arguments and wider debates over growth strategies. It is hard to argue that the Dutch Disease is damaging, without also acknowledging that

specialization according to static comparative advantage may be costly in the long run. The relevant evidence overlaps closely with the evidence needed to justify interventionist trade and industrial policies. On this view, one can simultaneously believe in a Dutch Disease and noninterventionist policies, but only when the state is believed to lack the capacity for effective intervention.

Dutch Disease effects can be offset by other policy instruments. The state may be able to influence productivity, lower the costs of the traded goods sector, or provide direct subsidies. In principle, the effect of aid on the size of the traded goods sector can be offset by appropriate subsidies to that sector. The existence of the distortion is already enough to justify subsidies, and aid transfers will simply increase the level of the appropriate subsidy. But this scheme would demand greater state capacity, and more insulation from pressures for misuse, than seems likely for most aid-dependent countries.

A more promising approach recognizes that the extent of the Dutch Disease depends on how the transfers are spent. Most obviously, the real exchange rate appreciation will be less when the increased national income is spent partly on imports, because this limits the increase in demand for nontraded goods that is the heart of the problem. Technical assistance is a good example of aid that will not greatly affect demand for nontraded goods. Aid could also be used to lower the costs of the traded goods sector, by measures such as investments in roads, ports, and more reliable power supplies.²⁶ Once we distinguish between the uses of aid, generalizations about the Dutch Disease seem risky. And even without this distinction, not all models imply that transfers will be associated with deindustrialization (Corden & Neary, 1982). The effect is likely to depend on country characteristics. In an economy of the Lewis (1954) type, in which a traditional labor-intensive sector ensures that wage levels are relatively insensitive to labor demand, aid might have little effect on the labor costs of the traded sector.

5.3 Evidence on the Dutch Disease

Empirical researchers have searched for the Dutch Disease in various ways, corresponding to different links in the argument. Some results on aid dependence and the relative price of nontraded goods, or the share of manufacturing exports in total exports, can be found in Arellano et al. (2009) and Yano and Nugent (1999). The cross-country study by López, Molina, and Bussolo (2008) finds that a surge in worker remittances does indeed cause an appreciation in the real exchange rate. But for aid analysts, the most interesting papers attack the problem more directly, either using differences in outcomes at the sector level, or investigating the relationship between real exchange rates and productivity growth.

As Neary and van Wijnbergen (1986) note, the sectors most likely to be affected by the Dutch Disease are those exposed to foreign competition and with little

price-setting power. In many poor countries, the leading candidates will be producers of labor-intensive traded goods. These are the sectors most affected by real wage increases when labor demand in the nontraded sector increases. With this in mind, Rajan and Subramanian (2006) look for Dutch Disease effects by examining the relative growth rates of industrial sectors within each country. We should expect to see that the difference between growth rates in labour-intensive sectors and other industrial sectors is largest in the countries receiving large aid inflows. They find support for this hypothesis, using UNIDO data on industrial value added for 28 sectors, and data on 33 countries for the 1980s and 15 countries for the 1990s.

The use of sectoral data allows relatively clean identification, and is among the most persuasive evidence for the general equilibrium response predicted by standard models. But none of these findings confirm the existence of adverse effects on overall welfare. That question is harder to answer, for reasons similar to those in the separate debates about appropriate trade and industrial policy.²⁷ The most relevant evidence here is the panel data study by Rodrik (2008b), which indicates that real exchange rate undervaluations have significant growth benefits, especially for industrial growth. A corollary is that the Dutch Disease may be a genuine disease.

6. AID, FISCAL POLICY, AND VOLATILITY

Large and variable aid inflows pose significant challenges for fiscal policy. Svensson (2000a) reports that across the 50 most aid-dependent countries over 1975–1995, the mean ratio of aid to central government expenditure was around 50%. This means that political economy and fiscal effects will be central to understanding aid effectiveness in countries where aid intensity is high. This section will discuss the fiscal effects, and the next section sketches the possible consequences for political economy and accountability.

It is a long-standing criticism of aid that it could weaken the incentives for recipient governments to secure revenues through taxation (Kaldor, 1963). This argument is not as clear-cut as it may seem. Under the standard principles of public finance, the optimal response to an aid transfer will be to raise expenditure by less than the transfer, and use the remainder to reduce taxes and the government budget deficit (Collier, 1999; Kimbrough, 1986). Put bluntly, raising taxes in developed countries in order to reduce taxes in developing countries could easily be a legitimate strategy. It is a little odd that conservative critics of foreign aid, in particular, are willing to emphasize the economic benefits of low taxation for developed countries but not for aid recipients. This argument is reinforced by the historical importance of trade taxes to many low-income countries.

To see reduced fiscal effort as pathological requires a more complex story, perhaps one in which domestic institutional development is treated as endogenous. Azam, Devarajan, and O'Connell (1999) construct a model in the spirit of Kaldor (1963), in

which there is learning-by-doing in tax collection. In this sense, aid can weaken domestic institutional development, and the optimal response of the donor may be to defer aid while institutions develop.

The evidence on the relationship between aid and tax revenues is mixed. The early literature on fiscal response models has often been criticized as too simplistic (e.g., White 1998, pp. 38–39). The recent panel data study by Remmer (2004) found that aid is associated with reduced tax effort but a higher share of government spending in GDP, implying that aid is not fully offset by lower tax effort. Similarly, Burnside and Dollar (2000) included some evidence that the share of government purchases in GDP is positively associated with aid. Gupta, Clements, Pivovarsky, and Tiongson (2004) find that tax effort is raised by loans and lowered by grants, where the latter effect is strongest in countries with a high level of corruption. McGillivray and Morrissey (2004) and Moss, Pettersson, and van de Walle (2008) provide more detailed reviews of this literature.

Another consequence of high aid intensity is that when aid flows are volatile or unpredictable, they will destabilize the government budget. The modern literature on aid volatility starts with Lensink and Morrissey (2000), Pallage and Robe (2001), and Bulíř and Hamann (2003, 2008). The latter found that aid is often more volatile than fiscal revenues, and rarely stabilizing: instead, shortfalls in aid and revenue tend to go together. There is a danger of overinterpreting these results, however. The central questions are not whether aid to a recipient economy varies over time, but whether the variation acts to increase or reduce overall volatility, and whether the variation is a corollary of optimal policies. Hudson and Mosley (2008) argue that volatility “is often an inescapable by-product of characteristics of aid frequently seen as positive” and show that volatility is higher for “reactive” forms of aid such as food aid, emergency assistance, and budget support, and for countries with weaker records on meeting conditions on IMF and World Bank loans. Trying to isolate more damaging, and avoidable, forms of volatility will pose a major empirical challenge.

None of this denies that aid commitments and disbursements sometimes vary for reasons that have little to do with optimal donor policies, and the implications for recipients are worth discussing. It is often argued that countries anticipating an aid windfall should smooth consumption, and accumulate the equivalent of a sovereign wealth fund, but van der Ploeg and Venables (2008) highlight circumstances where windfalls should be used to increase current consumption and investment.

In practice, volatile or unpredictable aid threatens to destabilize government budgets. Eifert and Gelb (2008) cite evidence on budget instability in Africa, with a majority of countries experiencing annual deviations of spending from projections greater than 10%. Celasun and Walliser (2008) similarly emphasize the budgetary complications of unpredictable aid. The parallel literature on natural resource revenues had earlier highlighted the dangers of unstable transfers. Case studies, such as those of oil windfalls in Gelb and Associates (1988), show that revenue volatility destabilizes

investment planning and the wider government budget. More empirical research is needed on the fiscal response to variable aid flows, and on whether the much-discussed “resource curse” arises mainly from the level of rents, or their volatility. Existing analyses of volatility have tended to concentrate on year-to-year variation, and the patterns and effects of instability at different horizons are not well understood.

In terms of policy responses for aid recipients, Eifert and Gelb (2008) suggest the use of flexible precommitment rules and a reserve buffer, as a way to reduce volatility even when aid is sensitive to changes in policy. But for donors, a deeper set of arguments relates to the potential role of aid as an insurance mechanism. The welfare gains from successful insurance could be substantial. Arellano et al. (2009) consider a dynamic representative-agent model in which productivity and aid transfers are stochastic. When they calibrate the model to data for the Côte d’Ivoire, they find that eliminating aid volatility would have benefits equivalent to 8% of total aid, while using aid to insure against productivity shocks would have benefits equivalent to 64% of total aid. The model is too stylized to take these numbers as more than indicative, but the effects are large enough to be interesting, even without modeling effects on the stability of the government budget, political stability, and the risk of conflict. Similarly, Pallage, Robe, and Bérubé (2006) show that reallocating aid flows across time could lead to substantial welfare gains for countries experiencing macroeconomic shocks.

One obvious case for insurance stands out. Some of the world’s poorest countries are dependent on a narrow range of primary exports, and therefore vulnerable to fluctuations in world prices.²⁸ A natural implication is that aid inflows should be especially directed to commodity exporters, to form a more stable source of revenue, and one that could be varied over time to offset changes in the terms of trade. Collier and Goderis (2008) find little evidence that, historically, aid has been directed to countries unusually exposed to external shocks. The extent to which this is desirable may depend on how resource rents interact with governance: perhaps countries dependent on primary commodity exports are often less well placed to use aid effectively.

7. AID AND GOVERNANCE

Commentary on foreign aid often takes the view that, if aid is sometimes a failure, it is a necessary failure. Donors must “fail again, fail better.” But while outwardly modest, even this position risks complacency. It tends to assume that the worst possible outcome is the absence of improvement, whereas some of the failures of foreign aid will be actively harmful. And nowhere is that more obvious than in considering the effects of aid on governance. It is possible that aid could undermine accountability, weaken the development of state capacity, worsen corruption, and even lead to conflict. This section will review these arguments in detail.

The analysis represents a significant departure from the earlier discussion, which has mainly considered economic mechanisms. It is now a commonplace of the literature that a true general equilibrium analysis should include the joint determination of a political equilibrium. Even more ambitiously, institutions should be treated as endogenous in the long run. This recognition is central to some of the most fundamental debates on aid effectiveness. Few of them are easy to resolve empirically, but the parallel between aid transfers and natural resource revenues can be informative. Many accounts of the resource curse emphasize similar mechanisms to those covered below.

Critics of aid often highlight the potentially corrosive effects of aid dependence. This term is rarely given a precise definition, and inherits some of its rhetorical force from its echoes of welfare dependence: perhaps recipients of large transfers risk becoming passive, and less willing to help themselves.²⁹ In seeking to make these ideas more precise, Bräutigam and Knack (2004) describe several ways in which countries dependent on aid might see their institutions deteriorate over time. Work has begun on testing such ideas, which is not straightforward. Cross-section studies that relate institutions to aid face many of the same objections as aid-and-growth regressions. As for panel data studies, measuring institutional quality and governance is an imprecise art, so that much of the time-series variation in the available measures is likely to be noise. Current panel data findings vary: Coviello and Islam (2006) cannot detect an effect of aid on institutional quality, but Djankov, Montalvo, and Reynal-Querol (2008) find that aid is associated with deterioration in democratic institutions and in the extent of checks and balances on the executive. Rajan and Subramanian (2007) introduce a new approach, similar to their work on Dutch Disease: their paper examines the differential effects of aid on industrial sectors, depending on the varying sensitivity of different industries to weak governance. They find that, in countries which receive more aid, governance-sensitive industries grow more slowly than others. Overall, the empirical literature has developed rapidly, but the robustness of these findings remains to be determined.

A common practitioner view is that the institutional capacity of aid recipients is overwhelmed by the demands of multiple donors (e.g., Kanbur, 2000, 2006; Morss, 1984). The lack of coordination among donors means that programs and projects are endlessly multiplied, placing strains on domestic ministries, and diverting attention from domestic constituencies. This risks undermining the long-term development of both institutional capacity, especially in the civil service, and the domestic political process. Like much else that is wrong with foreign aid, it could be alleviated by better coordination among donors. Yet if there is a deeper question here, it is why the donor-prompted demands for state capacity are not met with a domestic supply response. Some of the discussion that follows will suggest possible answers.

7.1 History and the state

It is sometimes argued that aid has been especially ineffective in sub-Saharan Africa. Herbst (2000) attributes some of Africa's problems to weak states which find it difficult to maintain order or provide public goods, for long-standing reasons, some of which predate the colonial period. Africa's particular historical path and structural characteristics, such as low population densities and disconnected peripheries, are said to have limited the development of effective states, in contrast to the path taken by Europe and the countries heavily influenced by European institutions. This does not show the European path to be inherently more desirable: European institutions have partly been shaped by centuries of wars with immediate neighbors. It does mean that African countries may face a particular problem in terms of state capacity, and their ability to use aid effectively. In particular, some African states have not developed efficient institutions for taxation and public expenditure. Robinson (2002, p. 513) summarizes Herbst's view of the consequences for African countries:

... they have no tax bases and instead have to resort to highly distortionary methods of raising taxes (such as taxing trade) or redistributing income (for example via employment in parastatals)... States never had to make political concessions to their citizens, hence the lack of functioning domestic political institutions such as parliaments and the completely unconsolidated nature of democracy in Africa. Moreover, the lack of these institutions can help explain the venality and state corruption in Africa since these institutions provide key checks on such abuses. Finally, this set of institutions has been further encouraged in the last fifty years by foreign aid and development assistance.

If aid has perpetuated the weaknesses of African states, it is partly by allowing governments to retain power without developing effective fiscal institutions. Researchers in political science such as Moore (2004) view effective governance as the outcome of a historical bargaining process between the state and taxpayers. This is associated with the development of a "tax state" with a broad tax base, and mechanisms for the oversight of public expenditure. To the extent that developing countries can rely on aid to fund much of their expenditure, the bargaining process may be undermined, and with it the long-term scope for accountability. Many of the problems that could arise are similar to those associated with rentier states, notably oil producers; governments with access to large nontax revenue streams can exercise a degree of autonomy that is unhealthy for long-term institutional development. Moss et al. (2008) set out the relevant arguments in more detail.

The idea that aid could undermine accountability is historically informed and has to be taken seriously. It is, therefore, worth examining some of its weaknesses. It is not clear that the incentives for individual taxpayers, or the media, to monitor the uses of taxation are much stronger than their incentives to monitor the uses of aid revenues.

More importantly, although European countries may have developed effective states through a historical bargaining process, it does not follow that the same route is optimal or even feasible for low-income countries today, especially if the process would take decades. The development of European state capacity was partly driven by the need to finance wars, which complicates any analysis in terms of an implicit bargain.³⁰ So does the historical importance of trade taxes, rather than income taxes, to many of the most aid-dependent economies. The contributors to Bräutigam, Feldstadt, and Moore (2008) discuss these questions in more detail.

Despite these criticisms, the idea that aid can undermine the “social contract” surely defines an important constraint on aid effectiveness, especially for countries where the ratio of aid to government expenditure is high. A situation where recipient governments see themselves as primarily accountable to donors, rather than their own citizens, is clearly problematic. Not least given plans to increase aid, there is a clear need for more research on the relationships between aid, accountability, and tax effort, and on the parallel effects of resource rents.

7.2 Aid and political leadership

Critics of aid often argue that it is misused by local elites. One class of political economy models takes the identity of the elite as given, and assumes that its objectives differ from those of donors. In particular, Boone (1996), Adam and O’Connell (1999), and Lahiri and Raimondos-Møller (2004) assume that the state has been captured by an interest group that represents a subset of the population and can make transfers to that subset. Unconditional aid will be at least partially diverted toward these transfers, and its benefits unequally shared.³¹ Perhaps not surprisingly, empirical tests of this idea have to be somewhat indirect. Angeles and Neanidis (2009) relate growth to aid in a panel, and find the marginal benefit of aid to be lower in countries with substantial settlement by Europeans, arguing that European settlement has been associated with powerful elites with little concern for the poor.

Other models recognize that growth-promoting policies may not be in the interests of government leaders, and that, to secure their rule, leaders may need to implement policies that have adverse side effects for growth. Again, it is useful to view this with Africa in mind. The political process in African states has often been characterized as one of “personal rule,” in which presidents are unconstrained by effective checks and balances, parties have a limited role, and clientelism is deeply embedded both in politics and the wider society. Bates (1981) argued that inefficient policies in some African countries not only allowed transfers to rulers, but were also used to ensure their political survival. With similar ideas in mind, Acemoglu, Robinson, and Verdier (2004) present a formal model of personal rule. In such an environment, unconditional aid could make it easier for a leader to respond to threats to their position, by using

transfers to buy off pivotal groups. Aid might be used in this way only off the equilibrium path, so aid could be perpetuating the rule of a leader even while it appears well-spent. The problem might be alleviated by effective conditionality or selective aid allocation, and there is another open question about the extent to which aid has genuinely sustained leaders in power. Acemoglu et al. cite a scholarly consensus that aid played a key role in Mobutu's hold on power in Zaire.

Leaders may be wary of growth, when it threatens their hold on power. If growth brings socioeconomic development, it could also bring countervailing centers of power, lower costs of coordination for political opponents, and greater restraints on predation and political corruption, such as a free press and an effective judiciary. As a result, leaders may find it in their interests to stifle economic growth (Adam & O'Connell, 1999). Aid may be ineffective, and used primarily to reduce political competition rather than spur growth. And since unconditional aid will be one source of income for political incumbents, it will increase their incentive to block development. But the practical relevance of these arguments remains an open question, and some numbers make this point. A leadership that sustained a 4% growth rate over 20 years would see GDP rise by 120%. Set against the much greater output disparities between rich and poor countries, this seems unlikely to generate the extensive societal changes that would be a genuine threat to an incumbent regime. Moreover, economic growth should increase the popular support for rulers, and an autocratic regime could benefit directly from growth and use the proceeds to buy off opponents (Overland, Simons, & Spagat, 2005). These considerations imply that the trade-off between growth and contestability will vary with the alternative sources of revenue available to a regime, including resource rents.

7.3 Aid and the common-pool problem

Other models examine the interaction between aid inflows and competition among distinct social groups. As with resource rents, unconditional aid transfers create "rents to sovereignty" that risk greater rent-seeking and conflict. Bhagwati, Brecher, and Hatta (1985) gave a stark example in which a transfer would be exhausted by the resources used in lobbying: the benefits of foreign aid for national income are completely offset by the value of the fall in domestic output, and the recipient country is no better off than in the absence of aid. The more general point is that, as with resource rents, aid could increase the returns to rent-seeking relative to those from production. Aslaksen and Torvik (2006) survey the recent literature, and present a model in which rents influence the benefits of participation in a democratic process, relative to seeking power through civil conflict.

If a resource windfall or aid transfer is temporary, the economically rational policy will usually be to smooth expenditure. In practice, governments in developing countries seem to respond disproportionately to temporary transfers, sometimes even

raising spending more than one-for-one. Tornell and Lane (1999) and Svensson (2000a) show that such a response may nevertheless be politically rational. Their analyses are based on the common-pool problem, in which social groups compete for allocations from common resources, such as the government budget. In the paper by Svensson (2000a), this process is modeled as a repeated game between social groups. One difference from standard common-pool problems is that the choices of the donor will also influence outcomes. If aid is sensitive to country needs, the donor response will improve outcomes in the event of noncooperation, and this makes cooperation harder to sustain.³²

The direction of change in available resources, rather than their absolute level, may also be relevant. Rodrik (1999) argues that fast-growing countries are distinguished partly by their capacity to adjust to negative shocks, such as adverse changes in the terms of trade. In countries with weaker institutions and greater social polarization, a negative shock risks a period of social conflict and instability, with damaging consequences for economic growth. In this view of the world, aid should be directed toward countries experiencing major external shocks, as a way of heading off the social conflict that would otherwise follow diminished resources. This reinforces the point made earlier: stabilizing aid flows is a less sensible policy target than using aid to reduce overall volatility.

7.4 Aid and corruption

In the Western media, the coverage of foreign aid gives especial prominence to political corruption. A populist criticism of aid is that it has simply lined the pockets of corrupt dictators. At first glance, there is not much to debate here: where a leader's power is unchecked, government-to-government aid will only serve to increase the personal fortune of leaders and their allies, and help them to resist political competition. It is surely a mistake to see this as the central dilemma for foreign aid. The policy recommendation is obvious: don't fund the likes of Mobutu. And such transfers are less likely now that the strategic imperatives of the Cold War, real or perceived, no longer apply. The harder dilemmas arise when an autocratic leader is corrupt, but has at least some commitment to development, or at least some incentive to act in the general interest.

Away from the most glaring instances of autocratic corruption, the issues quickly become more complicated. Political corruption can take other forms, embedded in patron-client relationships and vote-buying, politicization of some economic activities, repression of the media, and complex regulations that become pretexts for bribes and political favors (see Pande, 2008 for a review). There are also forms of bureaucratic corruption that arise through the everyday administration of the state, such as bribes extracted by low-level officials, or theft from public services. These different forms of corruption will vary in their implications for aid effectiveness and the allocation of aid across countries.³³ For donors, three questions become especially relevant. The first

is whether aid risks the promotion of rent-seeking and corruption, as suggested by the earlier discussion of common-pool problems. The second is the extent to which different forms of corruption become an effective tax on aid, again limiting its benefits. The third is how donors should respond to corruption, and especially the forms that do most to render aid ineffective.

The cross-country literature has partly used measures of corruption in order to proxy for rent-seeking activity. This has interest partly because rent-seeking can otherwise seem an uncomfortably abstract, featureless concept. Svensson (2000a) uses indices of corruption and regulation, and the black-market exchange rate premium, to infer the effect of aid on rent-seeking. He finds that aid increases corruption, and this effect is stronger in countries that are ethnically diverse. The review of the determinants of corruption by Treisman (2007) does not consider the role of aid, but suggests that resource exporters tend to be more corrupt: this is consistent with an effect of nontax revenues on rent-seeking. The panel data study included in Alesina and Weder (2002) finds that aid increases corruption. But the findings in the wider literature remain too disparate to be confident about policy conclusions. For example, using an instrumental variable approach, Tavares (2003) finds that the effect of aid is to decrease corruption. Dalgaard and Olsson (2008) provide a survey of this and related work. The variety of findings in the literature might indicate that more research is needed, but the precedent of aid-and-growth regressions does not augur well.

Arguably, empirical research has more reliable things to say about the possible consequences of corruption, with implications for aid allocation. Mauro (1998) found that the ratio of government education spending to GDP tends to be lower in countries that are rated as corrupt. Keefer and Knack (2007) show that the share of public investment in GDP, and in total investment, are both higher in countries with weak governance and limited constraints on government, even conditional on the level of GDP *per capita*. A natural interpretation is that public investment provides an especially lucrative means for government leaders and officials to extract rents and bribes.

Certain forms of bureaucratic corruption, notably theft from public service providers, impose an effective tax on aid. In the course of his skeptical reviews of the potential of aid, Easterly (2008, 2009b) cites studies of Guinea, Cameroon, Uganda, and Tanzania which estimated that 30–70% of government-purchased drugs disappeared before reaching any patients. Perhaps the first point to make is that there is a world of difference between 70% of drugs reaching patients, and 30%. And it is inevitable that some fraction of aid transfers will be wasted or misappropriated.³⁴ Ultimately, a point will be reached where the marginal social cost of reducing fraud will outweigh the benefits. The socially optimal level of corruption is not zero, any more than the socially optimal levels of crime or pollution are zero. Aidt (2003) reviews the relevant literature, and the arguments that a benevolent government may find it optimal for some degree of corruption to persist.

The costs of corruption may go far beyond an implicit tax, however. A deeper analysis of corruption emphasizes its interaction with domestic politics and institutions. One of the best summaries can be found in van de Walle (2005), who sketches how corruption, weak institutions, and flawed forms of politics can reinforce each other, because corruption means that too many actors have a vested interest in current institutions, and will resist any development of institutional capacity, or forms of political competition that move beyond clientelism.³⁵ This vicious circle may be difficult for donors to break, but achieving this would have high returns: not simply less corruption, but also stronger institutions and better political outcomes. One open question is the extent to which governance conditionality can be used to break this circle at a different point. With African countries especially in mind, van de Walle argues that aid should be conditional on the enforcement of presidential term limits.

It is not clear whether donors can exert much leverage on corruption. One strategy has been to withhold aid where corruption is apparent, and Pande (2008) lists some recent examples. The direct effect on incentives may be weak: the resources foregone by a regime will count for little, relative to the domestic political economy forces and social norms that render corruption an equilibrium outcome. The indirect effects of withholding aid on the perceptions of leaders, both among the domestic population and foreign investors, could have more powerful effects. For corruption at lower levels of government, some of the most interesting evidence on remedial measures comes from randomized trials. Olken (2007) studied corruption in 600 village road projects in Indonesia, and compared a top-down preventive measure (more central government audits) with monitoring based on grassroots participation. He found that the top-down measure was relatively effective, despite the prevalence of corruption in Indonesia.

Although addressing corruption does not seem easy, some of the new evidence is informative, and might allow real progress. But if corruption is sometimes hard to address, a pressing question for donors is how to map the disparate forms of corruption into explicit or implicit rules for the allocation and delivery of aid. In practice, donors have been criticized for failing to direct aid away from corrupt countries (Alesina and Weder, 2002). In an ideal world, allocation rules would recognize that the forms of corruption, and its consequences, differ across societies. Sections of the empirical literature on corruption risk brushing these distinctions aside, but this does not mean that donors should follow suit.

7.5 Aid, political instability, and war

While thinking about the nature and effects of corruption, and in reading the aid literature more generally, it is tempting to posit a law: the most important aid questions are also the hardest to answer. The possibilities that aid may lead to social conflict, political instability, coups, and civil war cannot be ruled out, especially when aid is given to countries that have recently experienced conflict. In practice, donors vary widely in

their response to conflict, which hints at the continuing uncertainty in this area (Balla and Reinhardt, 2008). The early 2000s have seen some donors giving increased amounts to postconflict reconstruction, partly with a view to reducing political instability and the risks of future conflict.

The links between aid and violent conflict are not easy to investigate. Much of the evidence comes from cross-country empirical studies, with all their limitations. For sub-Saharan Africa, de Ree and Nillesen (2009) find that aid is not associated with the risk of onset of civil war, but is associated with shorter durations when conflict does break out. Collier (2007) argues that aid does not increase the risk of civil war, and to the extent that it raises living standards, may in fact reduce it. He is more circumspect about the effects of aid on the risk of a coup, and explains the possible difference with civil war in terms of timing.³⁶ Collier and Rohner (2008) argue that democracy in low-income countries could raise the risk of political violence, because it constrains government repression of rebellions. Yet such hypotheses are complex and subtle, and place many demands on limited data. Easterly (2009b) is right to warn that this form of evidence cannot always bear the weight that is placed on it. Work along these lines is perhaps better viewed as a way of beginning to think systematically about some important hypotheses. Taking this further is likely to require detailed case studies of aid and conflict, informed by a combination of theory and history, and perhaps more use of regional data on conflict within countries; Besley and Persson (2008b) provide some references.

When aid is fungible, it may increase the risk of regional arms races and external war. Collier and Hoeffler (2007) examine these effects, and conclude that aid may have doubled military spending in Africa. This raises further questions, including the extent to which military spending will limit external conflict and internal rebellions through deterrence effects. Collier and Hoeffler find little evidence for such offsetting effects, in which case an aid-financed arms race will lead to military spending that is inefficiently high.

7.6 Aid to fragile states

Further degrees of complexity are introduced by the consideration of failed or fragile states: how should donors proceed when the state is fragile, and largely unable to enforce the rule of law, exert control over its territory, or deliver basic services? This question has gained particular urgency, partly because the category of fragile states increasingly overlaps with that of the poorest countries, and partly because of arguments that aid could help to forestall more serious problems, and reduce the risk of future conflict. Yet, aid in these circumstances involves fundamental ethical dilemmas, which will sometimes require a choice between evils. For example, attempting to use aid to stabilize a fragile state might effectively mean funding internal repression, or enhancing its capacity to fight an external war. It is not clear that economists are well

suited to analyzing these dilemmas, which require broader expertise in history, politics, and philosophy, and detailed knowledge of a specific country's situation. The dilemmas for outsiders become even more complex and profound when the structural problems of a nation state may render it unsustainable. Herbst (2000) argues that some African countries have reached this position, so that the idea of a nation state is effectively a fiction. Somalia is not a nation in any meaningful sense, but only a geographic area, or site of conflict. Perhaps it would be better to recognize this, with redrawn borders, or arrangements that go beyond traditional states, such as power-sharing or UN-backed "trusteeship." Others see many dangers in such ideas, and might argue that conflict is largely a function of poverty. There is scope for sins of omission, but also those of commission. These questions are daunting, and require a much broader and historically informed approach than an economist can provide, especially when the costs of a mistaken analysis are potentially so vast.

On this basis, the wider questions must remain beyond the scope of this survey, but it is worth briefly discussing the delivery of aid within failed or failing states. By definition, these are environments in which transfers to leaders are likely to be ineffective or harmful, as Easterly (2006b) illustrates with numerous examples. His position is that donors should abandon the pretence that much can be done. It is close to the realism of a doctor carrying out advanced triage in the wake of a disaster: it is important to recognize that some patients cannot be saved, because treating them will divert time and resources from those with more chance of survival. There is a bleak logic to this position. The leading alternative is the bypass model: route money via NGOs, perhaps mainly for humanitarian relief, or fund the operation of independent service authorities that carry out basic provision of education and health care, operating outside what is left of the state. One argument for this approach is that, in societies where little is being done for the poor, the returns to successful interventions will be high. But van de Walle (2005) argues that the past record of independent provision does not bode well. And although the bypass model is sometimes recommended in recent commentary, it may not be feasible on a large scale, especially given the considerable personal risks faced by aid workers and service providers in fragile states.

8. A NEW PARTNERSHIP MODEL?

So far, the chapter has reviewed aid effectiveness through the lenses of growth economics, trade theory, and political economics. It has typically stayed close to the academic literature in its concerns, trying to draw together some of the mechanisms needed for a structural model. But some of these issues may seem remote from the immediate practical questions faced by donors. The remainder of the chapter will begin to engage with some of the policy debates more directly. The discussion will be arranged around the new "partnership" model, which draws together some of the most

important changes in donor policies since the mid-1990s. Relative to older approaches, the partnership model gives more emphasis to selective aid allocation; to the achievement of concrete progress in meeting humanitarian goals; to governance, institutions, and (at least rhetorically) local ownership of reforms; and greater reliance on budget support rather than project financing and *ex ante* policy conditionality.

One reason for speaking of a new model is that several of these policies are mutually reinforcing. They share an emphasis on the autonomy of recipient governments, balanced by the need for accountability. The constituent policies also share a historical background, and reflect similar developments and pressures. These include shifts in the conventional wisdom on the binding constraints on growth and aid effectiveness; the diminished role for political and strategic motivations for aid, following the end of the Cold War; and political reform in some developing countries. Some of these influences are plausibly exogenous, but another impetus for change has been self-examination on the part of aid agencies and policy makers. This includes a widespread view that policy conditionality in its traditional 1980s form, whatever its merits in principle, was a failure in practice. A landmark in the development of the new partnership model was the 1998 publication of the *Assessing Aid* report by the World Bank, which questioned the effectiveness of traditional aid and conditionality, and advocated a greater role for selective aid allocation.

Some claim that the partnership model offers new hope for effective aid, but others are wary of the new emphasis on unconditional budget support, and criticize the lack of attention paid to ensuring that policies are evidence-based, and to the reform of some long-standing weaknesses in donor practices.³⁷ Moreover, since the partnership model is partly an attempt to stake out common ground, it is inevitably incomplete, and leaves the harder dilemmas unresolved. The rest of the chapter will consider the model in detail, and draw attention to some of these open questions.

8.1 Should democracies get the most aid?

Given the new emphasis on autonomy and accountability, a natural suggestion is that aid should be directed mainly toward democracies. This idea has come to greater prominence for several reasons. Democratic governments should be accountable, and more likely to use aid to good effect. Political competition and accountability could lead to better outcomes in areas such as education and health. And the 1990s saw a wave of democratizations, which has influenced the default assumptions and policies of the leading donors. Dollar and Levin (2006) show that, conditional on recipient GDP *per capita* and institutional quality, bilateral donors are relatively generous toward democratic countries, and this effect was stronger in the early 2000s than in the 1990s. The Millennium Challenge Account, launched by the US government in January 2004, explicitly links aid to political freedoms, control of corruption, and respect for civil liberties and the rule of law, among more traditional-free market conditions.

At first glance, the large number of “emerging democracies” offers new hope for measures such as budget support. There are some grounds for caution, however. There are many countries which remain undemocratic. In some countries, institutions are formally democratic but political practice is not. Even among democracies, some are not consolidated, and an elected government may be at constant risk of a coup. Indeed, precisely because many countries have democratized only recently, the proportion currently at risk may be unusually high. Hence, allocating aid mainly to democracies is more controversial than it may seem at first.

Donors might well seek to promote democracy, not least for the reasons powerfully articulated in Sen (1999).³⁸ For aid effectiveness, the most important benefits of democracy are likely to arise in three ways: representation, accountability, and selection (e.g., Besley & Kudamatsu, 2006). Governments in democracies should give more weight to the well-being of the poor than a country dominated by a narrow elite or an autocrat. Accountability should encourage governments to seek long-term development and poverty alleviation, restrain corruption, and use aid to deliver effective public services. Political selection increases the likelihood that incompetent or dishonest leaders will be removed from office. All three mechanisms imply that democratic governments will share many of the objectives of altruistic donors. In contrast, among autocratic regimes, the degree to which objectives are aligned with donors will vary widely. All of this makes intuitive sense, but drawing conclusions for aid policy may be less straightforward than it seems.

8.2 Democracy as a necessary condition

One way to organize this discussion is to consider two extreme positions, and ask whether democracy should be either a necessary or sufficient condition for aid. Historically, most aid recipients have been undemocratic. Among examples sometimes given of aid-assisted success stories, Indonesia and South Korea were ruled by military leaders for much of the period over which rapid growth took place. Botswana has greater claims to be an example of growth under democracy, but even there, the same party has held power continuously since independence in 1967. These examples hint that democracy should not always be a precondition for aid.

Implicitly, those who advocate concentrating resources on democracies tend to group other regimes together, as if they are a homogeneous set of environments in which aid is inevitably unproductive. But rather like Tolstoy’s unhappy families, every nondemocracy is nondemocratic in its own way. After all, some national leaders have been avowed socialists, and might place a high weight on the well-being of the poor. By contrast, there are notional democracies where the terms of political debate are dominated by rich elites, or effective political competition is largely absent: so the claim that democracies are always more closely aligned with donor objectives looks problematic. Besley and Kudamatsu (2008) develop a model in which autocracy can even outperform a polarized democracy, at least when performance is assessed in narrow terms.

One finding in the literature on political regimes is that outcomes are more variable under autocracies than under democracy (Almeida and Ferreira, 2002; Rodrik, 1997). One explanation is the absence of a selection mechanism: autocracy becomes a gamble on the characteristics of the leader, with consequences that may last decades. In some aid-dependent countries, rulers have indeed stayed in power for a long time. van de Walle (2005, p. 19) emphasizes that the longer a ruler stays in power, “the more power they accumulate, the more they escape accountability, and the less other political actors check their power.” Consistent with this, Besley and Kudamatsu (2008) find that longevity of leaders is associated with weak performance. In explaining differences across autocracies, they emphasize the importance of a group with a stable hold on power that can remove ineffective leaders. This could explain why some autocracies have been associated with strong performance in terms of growth, health, and education.

These considerations invite the traditional historical debate about the extent to which particular individuals matter. Casual empiricism and formal evidence both suggest that the aims and characteristics of individual leaders play a significant role (Jones & Olken, 2005, 2009). Autocrats seem to vary widely in their commitment to development, for reasons that are not well understood. Suharto wielded something close to absolute power in Indonesia, but succeeded in improving living conditions, and sometimes took unpopular decisions in the interests of long-term growth (see Temple, 2003). His regime promoted rural development and achieved reductions in poverty. The observation that aid may have contributed to these outcomes, but also played a role in sustaining a brutally undemocratic leadership, raises an acute moral dilemma. Lawyers sometimes remark that difficult cases make bad law, and a possible conclusion is that such dilemmas have to be assessed case-by-case, using discretion rather than rules.³⁹ An alternative position is that military governments, in particular, should always be denied aid. van de Walle (2005) argues that “these are easy calls.” Placing such regimes beyond the pale is the best way to promote change, and might lower the risk of military coups.

8.3 Democracy as a sufficient condition

Should democracy be a sufficient condition for aid? As many have noted, institutions can be formally democratic without ensuring political competition. Research has indicated various facets of developing-country politics that can undermine accountability, which go beyond capture by elites. Keefer (2004) and Keefer and Khemani (2005) discuss several problems, including lack of information about performance, social polarization, and the difficulties that opposition leaders can face in making promises that are credible, which tends to encourage clientelism. The potential importance of effective political competition can be seen in the contrast between the Indian states of Uttar Pradesh and Kerala. These states share similar levels of income *per capita* and formal institutions, but differ widely in outcomes such as literacy rates, school enrollment,

and infant mortality (Drèze & Sen, 1995, 1996). At one level, this variation is good news, because it testifies to the scope for effective public action. But the contrast also reveals the importance of political competition in determining final outcomes.

The extent of political competition is an equilibrium outcome, and may be influenced by the form of democracy. Political scientists have suggested that parliamentary systems have benefits for emerging democracies. In contrast, many aid-dependent countries have presidential systems, with a weak legislature and poorly institutionalized political parties. van de Walle (2005) argues that the activities of donors have often risked sidelining the legislature. More fundamentally, he argues that presidential systems tend to work against effective democracy, not least in allowing rulers to extend their grip on power. This suggests that donors should promote term limits, which have forced the retirement of leaders like Daniel Arap Moi in Kenya (2003), Konaré in Mali (2002), and Frederick Chiluba in Zambia (2001). It also suggests that, if constitutions are to determine aid allocations, then countries with parliamentary systems should be favored relative to presidential systems. But South Africa and India aside, parliamentary systems are mainly found among countries with small populations, which tend to receive disproportionate amounts of aid in any case.

8.4 Summary

This discussion implies that it would be a mistake to allocate aid based solely on the characteristics of a constitution. At the most banal level, what matters for aid effectiveness is not simply the right to vote, but whether a government is committed to growth and poverty alleviation, and can establish the policies needed to achieve them. Those goals might be attained through political will on the part of a national leader, or through a process that delivers genuine accountability. Some autocracies will meet this challenge and some democracies will not. To the extent that these various factors matter for aid effectiveness, they need to be gauged as directly as possible.

Dismissing extreme positions is perhaps too easy. The more interesting questions are how donors can distinguish within the set of democracies, and within the set of nondemocracies, in order to infer a commitment to development, or at least an incentive to act in the general interest. One view is that the character of a government is revealed nowhere more clearly than in its spending priorities. This was the view of Schumpeter, citing Goldscheid's remark that "the budget is the skeleton of the state stripped of all misleading ideologies" (quoted in Moore, 2004). The development of methods for inferring the underlying commitments of governments, from their budgets or otherwise, is an interesting research topic, and one that could have practical implications for aid allocation. Gawande, Krishna, and Olarreaga (2009) infer "welfare-mindedness" using a structural model of the political economy of trade policy. A more traditional approach examines outcomes such as life expectancy and infant mortality conditional on GDP *per capita*, as in Sen (1981) and Besley and Kudamatsu (2008). The related literature is discussed in Ravallion (2005).

At a minimum, the study of the varying records of autocracies seems essential, as in Besley and Kudamatsu (2008) and Gandhi (2008). This should highlight not only the loss of information implied by any simple-minded, binary classification of political regimes, but also the political and institutional environments that promote aid effectiveness, and the data needed to discriminate among them. At the same time, it should be acknowledged that a nuanced approach is not easy to implement in practice. Chhotray and Hulme (2009) contrast the relatively subjective and detailed approach of the UK's development ministry (DFID) with the simpler principles of the Millennium Challenge Account, in order to illustrate the difficulties and constraints that arise for both strategies.

9. CONDITIONALITY IN THEORY AND PRACTICE

This section will review the issues that arise when donors impose conditions on grants, loans, or technical assistance, linked to certain policy choices by the recipient government. Conditionality has been hugely controversial, but views differ on whether it should be criticized for the clear success in securing policy change, or the equally clear failure to do so. This hints at the difficulty in establishing the relevant set of facts, which is not unrelated to a second point. Some of the issues discussed so far, like the Dutch Disease, are primarily technical issues that can be considered as largely outside politics. This is not true of conditionality. Precisely because donors are seeking influence over the policy choices of sovereign countries, conditionality is intrinsically political, and the debate often involves questions of political philosophy and international relations, as well as economics.

For one set of critics, policy conditions have often been unfocused and intrusive. Others have argued that, on closer inspection, many of the conditions are relatively uncontroversial and potentially beneficial, but difficult to enforce in practice. Collier (2007, p. 67) gives a vivid sketch of the debate:

Conditionality turned out to be a paper tiger: governments discovered they only needed to promise to reform, not actually do it. Meanwhile, the Western left... conflated the limited reforms being urged on the governments of the bottom billion with the neoliberal savaging of the state they were fighting at home. As a result, reforms that should have been popular with all except corrupt elites become toxic in the media both within and outside Africa. The essential struggle between villains and heroes within the bottom billion became twisted into one between Africa and the IMF. [emphasis in original]

Historically, conditions have often been used, and their importance grew with the introduction of structural adjustment lending in the 1980s. This made World Bank and IMF loan disbursement conditional on macroeconomic stabilization, and often on wider policy reforms. That approach was not a conspicuous success, and the

perception that traditional policy conditionality was ineffective, or even harmful, is one of the most important forces driving the new partnership model.

This section will first review traditional policy conditionality and its problems, and then trace the shift toward greater emphasis on domestic autonomy and “country ownership” of reforms, and the move toward governance conditionality. These issues are important not only in understanding the evolution of the new partnership model, but also its strengths and weaknesses, and the scope for viable alternatives.

9.1 Policy conditionality

Attaching conditions to aid starts from a reasonable assumption, namely that recipient governments will typically have multiple objectives, not all of which are aligned with those of donors. Although some recipient governments will seek to promote growth and reduce poverty, these goals are likely to conflict with additional objectives that donors are less concerned with, or even actively oppose. This conflict of interest is essentially that of the principal-agent problem in the theory of incentives. The donor acts as the principal, while the recipient government is the agent. Under standard assumptions, the optimal policy for donors is to make transfers to the recipient government conditional on either the actions of the recipient, or final outcomes that are linked to those actions.

A quick sketch will show how this might work.⁴⁰ Imagine that recipient governments care about the welfare of the poor, but also about additional (nondevelopment) expenditure. Each government can undertake policy reforms that are in the interests of the poor, but bears a private cost of reform effort. In each country, the realized welfare of the poor depends on a combination of aid transfers and chance events, where the probability of a good state of the world is increasing in the government’s reform effort. Now consider the simple case where the donor can observe reform effort directly. In these circumstances, the donor should make its aid transfers conditional upon the extent of reform. The direct effect of the aid on the welfare of the poor is reinforced by the indirect effect of conditional assistance, which induces the recipient government to move further in the direction of reform than it would have done otherwise. The donor “buys” reform in a way that should benefit the poor. In the more complex case where reform effort is not directly observable, aid flows could be made conditional on the observed state of the world. Relative to the decisions that would be taken under full information, the donor should increase aid in good states of the world and decrease it in bad ones, to ensure that recipient governments retain some incentives to carry out reforms.

These arguments are straightforward, and appear to have much to recommend them. Yet *ex ante* policy conditionality has often been regarded as either inappropriate or a misguided failure, and the international institutions and other major donors continue to explore alternatives. Empirical studies have generally concluded that conditionality had only weak effects on policy choices in practice. This seems clearest for structural adjustment

lending. Easterly (2005) argues that, although there has been a general improvement in macroeconomic policy outcomes in developing countries, the observed extent of improvement is not strongly associated with repeated structural adjustment. The history of structural adjustment lending is mostly one of lending rather than adjustment.⁴¹

Dollar and Svensson (2000) use the World Bank's self-assessment of the outcomes of structural adjustment programs to see which characteristics distinguish the successful instances of reform. About a third of the programs in their data set were assessed as failing to meet their objectives; this proportion rises to a half in Africa. This glass might be seen as half-full or half-empty, but the historical analysis of the African case by van de Walle (2001) is more clearly negative. The relatively weak outcomes in Africa could be explained by another finding from the Dollar and Svensson study: conditionality was likely to fail in particular contexts, especially when reforms were demanded of an undemocratic regime that had been in power a long time.

It is essential to understand the reasons why policy conditionality sometimes failed. Possible reasons include defensive lending, a lack of commitment to enforcing conditions, the use of too many conditions, pressure from powerful donors for lax enforcement, and perhaps above all, the possibility that domestic political economy forces are usually too strong for donors to have much influence. A further problem is that, when aid flows are volatile and may be scaled down in the future, reforms may lack credibility with the private sector (Collier, Guillaumont, Guillaumont, & Gunning, 1997). Some of these problems could be addressed by reform of donor practices, but others indicate that even the basic principles of conditionality should be abandoned.

9.2 Commitment problems

Part of the failure of conditionality may be that, in practice, it wasn't enforced. One reason is the time-consistency problem inherent in the donation of aid. The arguments sketched above rely on donors being able to commit to their conditions. If the conditions are not met, the aid should not be disbursed. In practice, that decision may be hard to sustain. *Ex post*, a donor that can act with discretion may want to disburse the aid in any case, perhaps because this will lead to the best remaining outcome for the poor. A recipient government that knows this has no incentive to reform, and conditionality loses its rationale.

There are a number of ways in which this problem may be compounded, as Mosley, Harrigan, and Toye (1995) and Kanbur (2000, 2006) discuss. One is "budget pressure" effects, in which the relevant officials of aid agencies are likely to face significant institutional pressures to disburse aid, regardless of whether or not conditions are met. These effects seem especially likely when there is an operational division between the allocation of aid and its disbursement. The final decision to disburse aid may be subject to a range of incentives and pressures that would make a contract theorist distinctly uneasy. This is not what the "aid contract" should look like.

In some cases defensive lending is likely, where new loans are disbursed to avoid default on past loans. This may explain the tendency for the structural adjustment of the 1980s and 1990s to have been, if not an absorbing state, certainly one in which exit was relatively difficult (Easterly, 2005, 2007a). Many countries received repeated loans. A favorable interpretation is that the approach was inherently long term in its goals and might often require multiple loans. A less favorable, and more common, interpretation is that the approach was failing to achieve structural adjustment, because donors could not commit to enforcing the associated policy conditions, or the policy failed to induce reform for other reasons, or the conditions were poorly chosen. Mosley et al. (1995) document various failings of structural adjustment in action, including the “Christmas tree” effect, whereby loans would become weighed down with long lists of conditions, an approach that was ultimately self-defeating.⁴²

In principle, there are a number of ways in which commitment problems could be addressed. One approach would be to change the incentives within aid agencies, so that the interests of managers are aligned with the credible enforcement of conditions. Changes to disbursement procedures, to avoid disbursement becoming an all-or-nothing decision, are one remedy that could have a major effect. Paradoxically, another would be to delegate the disbursement of aid to an agency that cared less about the poor. As with Rogoff’s conservative central banker in the literature on monetary policy, delegation to an agent with different preferences helps to address the time-consistency problem; Svensson (2000b) derives this result formally.⁴³ Continuing the analogy with monetary policy, an alternative resolution would be based on the importance of reputation, established through a repeated game between donors and recipient governments. The academic literature has not pursued this idea, perhaps because the incentives and pressures on aid agencies, and the motivations of individuals within them, limit the scope for creating a tough reputation.

An alternative approach, emphasized in Svensson (2003), is to design mechanisms that encourage budget managers to internalize the opportunity costs of disbursing aid. He suggests that a fixed quantity of aid should be allocated not to individual countries, but to a group of countries, with the final disbursements to individual countries based on their relative performance. Like yardstick competition in the regulation literature, this allows donors to infer something about the effectiveness of individual recipient governments in the face of common shocks. But it also has a second advantage, which is that donors have stronger incentives to redirect aid from those countries which have not met conditions on policy choices, since they can infer that the aid will have greater benefits elsewhere. This approach might be politically costly, since countries are explicitly engaged in a zero-sum game, but it is easy to imagine simple variants that would avoid this problem.⁴⁴

9.3 Heroes and villains

The discussion, so far, has considered the recipient government and society to be a unified actor, the “agent” within the principal-agent problem. An alternative perspective

gives more emphasis to domestic political economy and the internal battle between reformers and nonreformers, or between those Collier (2007) calls heroes and villains. This implies a different view of what conditionality could achieve, why it may have failed, and how it should be reformed. Conditionality may strengthen the hand of a reforming government, when it needs to swing domestic opinion behind a set of policy changes. Even when the donor and the recipient government are in complete agreement on objectives and policies, the recipient government may actively want conditions to be imposed, in order to restrain its scope of action in the face of short-term domestic pressures. The donor then plays the role of Ulysses' crew members, tying him to the mast at his own request, so that Collier et al. (1997) call this form of conditionality "restraint." In this case, conditions imposed by donors can help to make reforms credible, and allow domestic reformers to deflect responsibility for measures that would otherwise be unpopular. Conditionality could also be used to influence the payoffs of groups that oppose reform, or strengthen a reformist government's bargaining power, as in Drazen (2002).

Yet it is not clear that the effects of conditionality are always so benign. Rodrik (1989) pointed out that the battle between heroes and villains may be complicated by incomplete information on the part of the private sector. Consider the case where the private sector cannot reliably distinguish between reformers and nonreformers. Conditionality may induce nonreformers to appear to behave more like reformers, at the expense of making it harder for reforming governments to reveal their true type. In the language of signaling games, conditionality makes a pooling equilibrium more likely. Reformers might find that they can only signal their type within a separating equilibrium if they implement reforms "too far, too fast," relative to the case where government type is directly observable.

The extent to which the key battle is internal, between reformers and nonreformers, is an open question. Long-term observers of African economies, notably Collier (2007), seem especially likely to take this view. It implies that aid agencies need to be politically astute, and quick to act, but also raises a dilemma. From the point of view of national self-determination, it is even more intrusive for donors to support specific political groups, than to impose loan conditions. Yet the alternative, failing to support brave individuals who try to change their societies from within, is hardly attractive either. Collier (2007, p. 96) writes that, too often, "we have stood by and watched them lose."

9.4 Conditionality versus ownership

Economists tend to assume that conditionality is a good idea, if only it can be made to work. More widely, however, policy conditionality is viewed with considerable uneasiness, and it is worth discussing the reasons why. Many of the objections lead to the same conclusion, which is that aid recipients should "own" the policy agenda. As Khan

and Sharma (2003) put it, the country carrying out reforms has to be committed to their spirit as well as their letter. What matters most is the existence of a domestic constituency genuinely willing to support and defend a set of reforms, and this cannot be achieved by conditions that have been externally imposed.

One basic issue that arises is the content of the conditions. The logic of policy conditionality risks hubris: donors have to be confident that reforms will genuinely promote the interests of the poor. The appropriate mix of policies will vary across countries and over time, in ways that are not well understood. The reality of this will be messy, and recipient governments may be understandably wary of being forced into policies that sometimes don't work, and which have a tendency to move in and out of favor depending on intellectual fads. But this argument can be overstated. A defender of policy conditionality might not want to prescribe a unified development strategy, but only seek to rule out policy choices that are known to be almost universally harmful. Sustained budget deficits, high inflation, and overvalued exchange rates can undermine growth for reasons that are reasonably well understood, command wide consensus among economists, and have some empirical support. In this view of the world, donors do not know a universal recipe for growth, but they should certainly aim to eliminate some toxic ingredients.

Something like this reasoning explains why the IMF has attached macroeconomic policy conditions to its loans. In practice, however, the conditions moved beyond macroeconomic adjustment and encompassed an increasing variety of detailed structural reforms. There are several ways in which the snowballing of the scope of conditionality may have harmed the IMF's original objectives. The attempt to impose potentially unpopular structural reforms could weaken ownership, and the legitimacy of an overall program, threatening its chances of success. Multiple conditions will also increase commitment problems: it may be hard to meet them all at once, and yet it is not clear whether a donor should disburse when conditions have been only partially fulfilled. When the disbursement decision is all-or-nothing, multiple onerous conditions become especially problematic, and risk confidence in the program and its overall success.

There is no doubt that, in some cases, the conditions became overly demanding and intrusive. Stiglitz (2002) gives the example of Korea during the East Asian financial crisis of 1997–1998, when the IMF initially sought to make its loans conditional on a change in the charter of the central bank, increasing its independence. This was surely a matter for Korea's voters and leaders, and not outsiders. The IMF has since acknowledged that its prescriptions went too far, and has moved toward a streamlined approach based on conditions seen as critical to the IMF's core responsibilities (Khan & Sharma, 2003). Although the use of policy conditions remains unpopular with many, it is surely legitimate for the IMF, as a revolving fund that is called on at times of crisis, to seek to impose conditions that help to ensure its loans are repaid.

The remaining issues surrounding conditionality are more complex for long-term development lending, such as that undertaken by the World Bank, sometimes the IMF, and bilateral donors. Again, the conditions have often involved detailed structural and microeconomic reforms. It is often argued that these attempts to secure reform and growth have failed. Even where policy conditionality induces reform, it might be short-lived, partial, and lack credibility. To the extent that it is successful in changing policies, there is a deep question about the extent to which outsiders should seek to influence recipient governments, especially when the relevant policy areas are not directly relevant to the repayment of the loans. At one extreme, critics interpret the Washington Consensus policy agenda as primarily driven by the economic self-interest of donors, so that conditionality is characterized as a new form of economic imperialism.⁴⁵

The modern view tends to sidestep this debate, by emphasizing that donors can rarely exert much influence, relative to the domestic political economy forces: the complex interactions between political and economic interests, and the beliefs of policy makers and the wider population. The means available to outsiders to effect policy reform are simply not strong enough. Trying to engineer specific policy reforms with externally imposed conditions risks becoming a quixotic exercise in which donors are tilting at windmills from a great distance. Governments may pay lip-service to the need for reform, and even pursue it temporarily, but lasting change requires a domestic constituency that is ready to promote and sustain a particular set of policy measures.

9.5 Governance conditionality

The recognition of the importance of domestic politics has encouraged a shift toward “process” or governance conditionality, as one element of the new partnership model. Rather than requesting specific economic reforms, some donors have attached broader conditions to the policy-making process, such as mechanisms for accountability, transparency, and recognition of human rights, and reform of the public sector. The Millennium Challenge Account is one example, with an emphasis on selective aid allocation. Among others, van de Walle (2005) and Collier (2007) would like to see the idea extended to other countries, so that aid is often made conditional on improvements in governance.

This form of conditionality recognizes that policy reform and growth will often rely on the fundamentals of governance, rather than short-lived loan conditions. Governance conditionality could be used to promote simple, concrete changes that are likely to lead to better outcomes, such as the term limits strongly advocated by van de Walle (2005). More subtly, it may lessen the costs for reformers of signaling their type, if nonreforming governments find it especially costly to improve mechanisms for accountability. Last but not least, the approach seems inherently more legitimate than attempts to micro-manage economic policy, especially when it is focused on strengthening accountability to the

domestic population. As Collier (2007, p. 110) observes, why should donors give aid to countries that are not willing to let their citizens see how they spend it?

The approach does inherit some of the failings of traditional policy conditionality. It does not resolve commitment problems, and in some ways makes them worse, because certain governance conditions may be difficult to verify. Since governance conditionality might often be temporary, the approach relies on some degree of path dependence in political institutions, so that improvements in governance are locked in over time. The extent to which this happens in practice is highly uncertain. But overall, these misgivings look modest compared to the potential advantages of governance conditionality.

10. SELECTIVE AID ALLOCATION

The various failings of conditionality have led to greater emphasis on three separate but related ideas: selective aid allocation, greater country ownership, and social involvement in the policy-making process. These ideas have come to dominate the rhetoric and stated policy commitments of donors since the late 1990s. An overall theme is that the donor-recipient relationship should be cast in more equal terms, as a partnership engaged in continual dialog about priorities and the means to achieve them. This and the next section will discuss how donor strategies have moved beyond policy conditionality, starting with the use of selective aid allocation.

The idea here is that, rather than making aid disbursements conditional on specific policy choices, aid should instead be targeted at those countries where it is most likely to be effective. This approach, sometimes called selectivity or “*ex post*” conditionality, and also known as performance-based allocation, is associated with the World Bank report *Assessing Aid* (1998) and an underlying empirical paper published as Burnside and Dollar (2000). It has been developed into explicit rules for aid allocation by Collier and Dollar (2001, 2002, 2004).

The advocates of this approach argue that aid is most beneficial in good policy environments. Donors should redirect aid toward those environments where its marginal effect is greatest, and disburse it with fewer conditions. Aid can then make a bigger impact, and lift more individuals out of poverty, than if the same expenditure was allocated independently of the policy environment. This is intuitively plausible, and avoids many of the problems of *ex ante* conditionality, but is not without weaknesses of its own. If we start from a purely statistical viewpoint, although Burnside and Dollar (2000) advanced the literature in certain respects, the limitations of the data are clearly significant, and mean that caution is essential. Their findings on conditional effects rest heavily on a specific interaction term, between aid and policy, in a cross-country growth regression. By construction, interaction terms are highly correlated with at least two other variables in the regression. Trying to identify their effect with

any degree of reliability, in a small data set, is difficult. In this context, it is not surprising that Easterly et al. (2004) showed, in effect, that the models favored by Burnside and Dollar lacked predictive power out of sample. Other researchers have tended to agree that the Burnside and Dollar evidence is inherently fragile.⁴⁶ One of the problems is the number of alternative candidates for interaction terms. For example, Svensson (1999) argues that the growth impact of aid varies with the extent of political and civil rights. At least eight different interaction terms have appeared in the literature (see Angeles & Neanidis, 2009). There are not enough countries for a statistical analysis to be able to discriminate reliably among tightly defined dimensions of policy. Moreover, the approach is likely to give too much emphasis to the variables most easily measured, such as macroeconomic policy outcomes.

This is not simply an arcane dispute about *t*-statistics. Burnside and Dollar (2000) and subsequent papers are intended to make reliable statements about the necessary conditions for aid effectiveness, something of a holy grail for the empirical literature, and just as elusive. Their own approach sees the benefits of aid as conditional on macroeconomic policy outcomes and the Sachs-Warner index. If we consider the economics of this, the argument looks a little suspect. It is not clear that the marginal effect of all types of aid would be strongly reliant on macroeconomic stability. Aid that is directed toward health or education, or even infrastructure, may have long-term benefits that are largely independent of current macroeconomic policy outcomes. An alternative and more reasonable interpretation is that aid should be directed toward those governments that are likely to use it well, and the Burnside-Dollar policy measure helps to indicate where the government has lost control, or where the decision-making process should be regarded as dysfunctional. Following this logic to its conclusion, aid should be allocated using broad indicators of the policy environment, with the World Bank's Country Policy and Institutional Assessment (CPIA) as the leading candidate. This aggregates a wide range of policy and governance indicators into a single index, and has been prominent in work on the principles of aid allocation.⁴⁷ Using multiple indicators should make it easier to assess the quality of policy, or even capture more interesting latent variables, such as the aims of government leaders or the extent of accountability.

At the same time, this could feel somewhat amorphous and imprecise. It hints at the serious limitations of statistical evidence in establishing the conditions for effective aid. At the risk of stating the obvious, the analysis of these conditions should draw on economic theory, country studies, and statistical analyses that study channels of influence, not simply on reduced-form relationships between aid and growth. There are too many problems with interaction terms and related methods to treat the results as anything more than indicative. What may be less obvious is that aid allocation rules can be developed in ways that are robust to uncertainty about specific candidates for interaction terms, and this will be discussed next.

10.1 Developing rules for aid allocation

If the results of Burnside and Dollar (2000) are too fragile to form the basis for policy, and some of their claims ultimately mistaken, at least the mistakes have been the productive kind. Their work has prompted researchers to give new attention to aid and growth, the conditions for aid effectiveness, and also to rules for aid allocation, an area that has been underresearched, with a few notable exceptions such as Behrman and Sah (1984). For the sake of the current argument, imagine that we can take seriously an estimated relationship between growth and aid, and between the marginal effect of aid and the CPIA index. The quantitative implications for aid allocation are striking. Collier and Dollar (2002) argue that if aid flows were selectively allocated on a more systematic basis, the number of people lifted out of poverty each year would roughly double.

At one level, this should not be a surprise: given so many donors, and their lack of coordination, we should expect the current allocation of aid to be inefficient. Yet to start to quantify this is a major contribution, and picking holes in the specifics of the Collier-Dollar analysis, which is easily done, seems a rather trivial exercise by comparison. And their approach is more general than it seems at first glance. It might be thought that rules for aid allocation require an explicit stance on the relative importance of the mechanisms that intervene between aid flows and outcomes. That would not be good news, because the literature is many papers away from establishing this, as the length of this chapter testifies. But the Collier-Dollar approach depends on a country-specific elasticity of growth with respect to aid, and not on the forces driving the variation in that elasticity. In other words, since different conditioning variables are likely to be correlated with one another, a range of empirical models, based on alternative interaction terms, could give broadly similar findings about the variation in the growth-aid elasticity across countries. This avoids some of the criticisms leveled at the more ambitious interpretations of Burnside and Dollar (2000). It also suggests that the design of aid allocation rules could build in an unexpected degree of robustness to model uncertainty.

The Collier-Dollar approach has already influenced aid allocation in practice, and is likely to lead to interesting further work; see Wood (2008) for references. Cogneau and Naudet (2007) and Wood (2008) modify the framework in order to make aid allocations forward-looking, and the first of these also suggests an additional objective, namely to equalize poverty risks across countries. But one question is whether there is any practical hope of coordinating multiple donors on a more efficient allocation. One contribution of formal analyses is to strengthen the hand of those demanding greater coordination, by illustrating the scale of the inefficiency that arises in its absence.

10.2 Dynamic aid allocation

For understandable reasons, selective aid allocation is usually discussed in terms of a static problem, based on a snapshot of policy indicators and poverty levels. Some

important and unresolved questions relate to the dynamics of aid allocation. Should aid be directed toward new democracies, countries emerging from conflict, and other fragile states where there may be significant risks of future conflict? The argument for making these circumstances a priority is often based on expected payoffs. Although the chance that peace or democracy can be sustained will sometimes be low, the payoff to success could be massive. But making such arguments more precise is difficult. Dollar and Svensson (2000) found that World Bank structural adjustment programs were more likely to be judged successful when undertaken by newly elected governments. This tends to suggest a window of opportunity that donors might want to exploit. In other cases, there are relatively few examples to draw on, especially for transitions from conflict. Collier and Hoeffer (2004) report the practitioner view that post-conflict countries are sufficiently dissimilar that generalizations are almost impossible. One of their empirical findings is that aid tends to taper off relatively rapidly, once a crisis is over, even though continued support for development and peace-building might have a large payoff. Collier (2007, p. 106) argues that aid to postconflict countries has traditionally been “too little and too soon.” Boyce (2007) also discusses a range of relevant issues.

The many problems and uncertainties of assistance to postconflict countries suggest an alternative approach, which would be to use aid to forestall conflict, so that aid becomes partly a preventive measure. Since low incomes may increase the risk of future conflict, Fisman and Miguel (2008) propose that aid should be temporarily increased for poor countries that face negative shocks from low rainfall or changes in world prices. This is consistent with the evidence in Besley and Persson (2008b) that volatility in world prices is associated with greater risks of civil war. Hence swift responses by donors could have substantial benefits, and conceiving aid allocation rules entirely in static terms could have substantial dangers.

11. THE NEW CONDITIONALITY

The rhetoric of the new partnership model sometimes places great emphasis on the importance of country ownership of any reform agenda. This is seen as desirable in itself, but also in leading to better development outcomes in the long run. The idea of ownership is sometimes used to justify unconditional budget support, and to recommend alternative forms of conditionality. This section discusses some of the most recent ideas and policy initiatives, and the associated controversies. For some, the rhetoric of ownership, dialog, and participation is empty, and perhaps inevitably so. For others, “the need for popular participation is not just sanctimonious rubbish. Indeed, the idea of development cannot be disassociated from it” (Amartya Sen, quoted in Pritchett & Woolcock, 2008).

11.1 The meaning of ownership

As we have seen, country ownership is often justified in terms of the inherent desirability of autonomy. This line of argument is attractive but also incomplete. The first responsibility of donors should be to those most in need, and the long-term interests of the poor may be very different from those of a recipient government or local elite. The divergence in objectives was the motivation for policy conditionality, and there seem few good reasons to make the autonomy of political leaders a priority, rather than the autonomy of the poor. Especially when governments are not accountable, it could make sense to impose conditions on governance. As Collier (2006a) puts it, this limits the sovereignty of the government, but not the sovereignty of the country.⁴⁸

There is more to the case for country ownership than the desirability of autonomy, however. The hard-headed argument is that, unless aid recipients “own” a set of aims and policies, and unless there is a domestic constituency which supports them, donors will face an uphill battle in trying to secure favorable changes in policy. This is reform seen in longer perspective, less interested in changes in specific policies from year to year, and instead placing more emphasis on dialog and evolution (e.g., Morrissey, 2005). It also recognizes that reform sometimes depends on multiple stakeholders. Even within a government, there may not be a single policy line, and the fact that officials in a central bank or finance ministry have signed up to a reform program does not imply that it can be sustained, given the other parties involved. These arguments indicate that discussion, learning, and flexibility are needed. In the new partnership model, the role for donors is to offer advice and finance as part of an ongoing dialog: ownership is a process, not simply an outcome.

In taking this further, one potential source of confusion is that, in everyday use, ownership is a binary category: you either own something or you don’t. In this view of the world, ownership of the reform program must be held entirely by the donor or entirely by the recipient government. In contrast, it is better to think in terms of autonomy, which has multiple dimensions and is inherently a continuum. Moreover, supporters of country ownership are committed not only to some degree of autonomy, but also to consultation and dialog. This perspective helps to answer critics like Buitter (2005), who argue that the term “country ownership” has been emptied of practical meaning. Another response is to distinguish between government ownership and country ownership. As noted previously, when there is heterogeneity in interests within countries, conditionality may strengthen the hand of reformers, including governments that face internal opposition. This implies that government ownership can sometimes coexist with conditionality, and avoids the charge that ownership is a vacuous concept (Drazen, 2002).

11.2 Poverty Reduction Strategy Papers

In practice, the centerpiece of the drive for greater ownership has been the Poverty Reduction Strategy Papers (PRSPs), led by the IMF and especially the World Bank,

with the support of other donors, and initiated in 1999. In principle, each aid recipient draws up an explicit long-term strategy for addressing poverty, and links this to an overall development strategy. In practice, it has been observed that the PRSPs completed to date often seem remarkably similar in emphasis to the Washington Consensus, and also remarkably alike.⁴⁹ This has led to the charge of “ventriloquism”: aid recipients are led to understand which policies will be needed to gain donor support, and however dutiful their own effort to make a case for these policies, the exercise remains a long way from genuine ownership. The critical account of the PRSP process in van de Walle (2005) suggests that it has focused too heavily on the social sectors, has failed to influence World Bank thinking in any meaningful way, and is hard to reconcile with the logic of the traditional pressures and priorities faced by donors. If countries lack “ownership capacity” in various dimensions, or the donors remain too prescriptive, the stated commitments to autonomy will never be anything more than window dressing. The approach could even be harmful: the contributors to Gould (2005) argue that it works against the consolidation of democratic forces and civil society, partly by sidelining parliaments and the influence of political parties.

In defense of the PRSP process, it is ambitious and wide ranging, and should not be judged too soon in its development. Some of the prominent criticisms of PRSPs embody the binary view of ownership, which is often too simplistic. Whether the process succeeds in the long run will depend on how much freedom it has to evolve, whether ownership capacity can be established over time, and whether the donors will be able to let go of the process to a greater extent. Arguably, the scope for genuine autonomy has been enhanced by the more open-minded approach to development strategies that is beginning to emerge (e.g., Rodrik, 2005, 2007b).

11.3 Social participation

In discussing the record of PRSPs to date, van de Walle (2005) argues that their most important benefit has been to encourage greater consultation. It is often argued that country ownership requires wider social and grassroots participation in the design of aid programs. Some of the arguments risk naïvety about what will happen in practice, however. It is worth unpicking some of the assumptions, in the same way that researchers have steadily unpicked those of policy conditionality. Even thorough discussion may not be able to close a substantial gap between what is popular and what is necessary. An emphasis on social participation puts a great deal of faith in the ability of governments and civil society to consult, to agree on priorities, and to arrive at effective solutions. This could seem to ask a lot of a developing society, given the usual problems of limited resources and the technical complexity of many policy issues. As Hirschman (1958) noted in a different context, any society ready to implement this process effectively should probably be classed as already developed.

One obvious question will always be: who participates? Bräutigam (2004) discusses some examples of participatory economic policy. These include the well-known budget-setting experiments of Porto Alegre in Brazil, but also other forms of participation tried in Ireland and Mauritius. The possible models vary from reliance on corporatism and social partners, to more ambitious attempts at deliberative democracy and direct citizen involvement, as in Porto Alegre. These measures have sometimes shifted policy in a propoor direction. Open questions include the extent to which this relies on a propoor party holding power, and how social participation interacts with traditional representative democracy. It is not clear that participatory budgeting has always worked well when tried in other contexts. Bräutigam concludes that sustainable policy changes will still require traditional party politics, while the need to involve the business sector should not be ignored.

11.4 New forms of conditionality

The ideal of country ownership helps to motivate some new forms of conditionality, in which specific reforms continue to be advocated by donors, but the recipient keeps more freedom of maneuver than in the past. One of the relevant possibilities, admittedly conservative, is that of “floating tranche” conditionality. As in more traditional approaches, conditions are imposed and loans disbursed only if the conditions have been met. The difference is that the conditions are not evaluated at a specific date, so reforms can be brought forward or delayed as the recipient government sees fit; only once the reforms are undertaken will the funds be disbursed. This has advantages for the recipient, since it brings greater autonomy and flexibility in the timing of policy changes. It also has advantages for the objectives of the donor, since there is less pressure on the donor to disburse on a given date even though the conditions have not been met. The World Bank introduced this approach in sub-Saharan Africa in 1995, as part of the Higher Impact Adjustment Lending (HIAL) initiative. The assessments carried out to date have been broadly favorable, and it is hard to see any significant disadvantages other than those created for donor budgeting.

A more radical idea is that conditionality should shift from input and process variables, to an emphasis on performance, based on direct assessment of final outcomes. This is sometimes called performance-based conditionality, and might also be called outcome conditionality. The idea links naturally to the Millennium Development Goals, since the assessed outcomes could be infant mortality or primary school enrolment. Most economists should be instinctively sympathetic to approaches that make rewards conditional on final outcomes, rather than attempting to prescribe input choices, at least when government efforts and actions are hard to observe directly. Recipient governments would retain important dimensions of autonomy, including some flexibility in priorities and the details of policy.

With advantages like this in mind, the European Commission has experimented with performance-based aid for the African, Caribbean, and Pacific (ACP) countries. Adam and Gunning (2002) and Adam, Chambas, Guillaumont, Guillaumont Jeanneney, and Gunning (2004) explore its operation in practice. It accounts for only a small fraction of European aid, but they are still able to indicate some of its benefits and problems. Moving to performance-based aid can create a “results culture,” but there are difficult problems in measuring the final outcomes of most interest. The approach then risks falling back on the prescriptive policy and input variables that characterized previous forms of conditionality. A second problem is that the mapping from outcomes to the amount disbursed needs to be clear and explicit, to avoid commitment and incentive problems in disbursement.

Other suggestions also emphasize rewards for good outcomes, especially for basic needs. Barder and Birdsall (2006) propose a centralized, international “price list” of rewards for outcomes such as individual vaccinations, or primary school completions. The details of achieving these outcomes are for domestic governments to determine. The idea is attractive partly in providing better incentives for recipient governments to deliver effective services. This form of “cash on delivery” aid would also lessen commitment problems, at the expense of making donor budgeting more complex. But perhaps the central objection is that the approach requires strong mechanisms for monitoring outcomes. In their discussion of suitable performance indicators, Adam et al. (2004) note that relatively accurate child mortality statistics are available through the DHS (Demographic and Health Survey) but only at 3–5-year intervals. The problems in measuring educational achievement, as opposed to input variables like enrolment rates or class sizes, are even more serious. To gain some sense of the difficulty, it is worth remembering that measuring public sector productivity is a formidable problem even in the context of the world’s richest countries, let alone the poorest. There is also a chicken-and-egg problem. Donors will be reluctant to make payments-for-progress until the data needed for evaluating progress are reliable. At the same time, there is little incentive for donors and recipient governments to attempt to collect and verify such data until more is at stake for their policies. That would require large-scale commitment to evaluation in terms of final outcomes. An interesting question is the extent to which collecting more data would have independent benefits for the management and reform of public services.

One alternative would be to look for indicators that are not final outcomes, but are closely associated with government priorities and their performance, perhaps using ideas similar to those discussed earlier in the chapter. Mosley, Hudson, and Verchoor (2004) recommend that assistance should sometimes be conditional on a given composition of the recipient government’s budget, and Collier (2006a, p. 1493) also sketches how this could work in practice. It is not difficult to see problems with such proposals. The necessary statistics, such as audited public accounts, may not be available in time

for conditions to be evaluated. Some performance benchmarks risk distorting choices. There is also the issue that, if specific targets are used to evaluate conditions for assistance, the reliability of the relevant statistics might well decline. These points are instances of Goodhart's Law: once a surrogate measure is made a target for policy, its relationship with other variables and its information content will change, reducing its suitability as a target. This idea was originally developed in the context of monetary policy, but is becoming ever more relevant in the public sector.

11.5 The future of conditionality

Arguably the debate on conditionality too often assumes that the same strategy should be adopted across the board. What is needed is a strategy based on an explicit typology of developing countries. Bourguignon and Sundberg (2007) argue that current policies increasingly amount to a three-track model. Countries with good governance and policies receive budget support, intermediate countries face something more like traditional conditionality, but with greater emphasis on governance and performance, and fragile states are aided through a combination of humanitarian assistance and aid that bypasses the state, for example, by allocation to NGOs. Note that, if the right approach is likely to vary across countries, traditional policy conditionality could still have a role. At least since *Assessing Aid*, the usual argument has been that the lever is inherently weak, but this is debatable in a world of scaled-up aid flows, and when innovations like floating tranche conditionality could help to overcome the commitment and disbursement problems of the past.

Moreover, there are some good reasons to remain wary of unconditional aid. Arguably, we already have a natural experiment, namely the large resource rents that some developing countries receive, sometimes far in excess of aid transfers. This parallel is now frequently drawn in the literature (Collier, 2006a; Dalgaard & Olsson, 2008; Djankov et al., 2008; Therkildsen, 2002). The conventional wisdom is that resource rents have been associated with diminished accountability and poor outcomes, especially in countries with weaknesses in governance. For aid to achieve better outcomes, the three-track model must place considerable faith in the ability of donors to identify good governance, in the resilience of good governance in the face of large-scale transfers, and in the scope for using new forms of conditionality to achieve good outcomes where governance is weak.

12. DONOR POLICIES

So far, the chapter has largely viewed aid effectiveness in terms of the characteristics and policies of recipient countries. But a recurring theme in the literature has been the perceived failings of donors and international institutions, whether the World Bank and the IMF, or the bilateral donors. This section will review the leading issues,

including the lack of coordination among donors, the possibility that donors have been too slow to respond to experience, and the role of statistical evidence and evaluation in improving future performance.⁵⁰

12.1 Proliferation of projects and programs

The lack of coordination among the world's aid agencies, and the associated proliferation of projects and programs, is well known. To see the scale of the problem, consider the case of Vietnam discussed by Acharya, Fuzzo de Lima, and Moore (2006). In 2002, Vietnam received aid flows that amounted to about 5% of GDP. The government had to interact with 25 official bilateral donors, 19 official multilateral donors, and about 350 international NGOs. These donors collectively accounted for about 8000 projects. This is not an extreme example: the same article reports that the median aid recipient must deal with 23 official donors, and that some bilateral donors divide their aid between more than 70 recipient governments.

The costs of this are obvious for the recipients, who have to divert scarce managerial time and expertise to handling donor missions, preparing budgets and reports, and dealing with different systems that are not harmonized across donors. There could be less tangible, but significant, benefits to greater specialization by donors: more learning, and a greater role for trust and reputation, through repeated interaction. A more coordinated approach would help in achieving an efficient allocation of aid across countries, and in enforcing conditionality.⁵¹

The lack of coordination has been acknowledged for decades, with little sign of improvement. Easterly (2007a) constructs indices of the fractionalization of aid flows for both donors and recipients, and finds no evidence of a downwards trend over time. Naturally, observers such as Easterly (2006b, 2007a, 2007b) and Svensson (2006) are led to ask whether the many related declarations and stated commitments by aid agencies are not beginning to sound rather empty. The practical responses have not been especially convincing. It is now common to emphasize the benefits of multidonor budget support, in which aid from multiple donors is pooled and transferred directly into the government's budget, with few conditions other than monitoring of financial management. The problem is that this approach only seems appropriate for countries with a high level of state capacity, where the consequences of project proliferation are less severe in any case. The budget support "solution" does not address the problem for the weaker institutional environments, which is where it matters most.

Other proposed solutions have included attempts at harmonization, and sector-wide approaches in which multiple donors pool support for a particular sector. As Acharya et al. (2006) observe, these are inherently rather conservative strategies, because they take it as given that donors will continue to operate in too many countries. Donors may want to be seen to be operating on a wide scale, planting flags, and exerting influence, but

these are dubious motivations. It does not reflect well on national aid agencies that so few are willing to limit, voluntarily, the scope of their operations.

A realist might argue that reform is difficult because the current outcomes represent a prisoner's dilemma, in which the rationality of individual donors leads to an outcome that is collectively suboptimal. Yet, given the billions of dollars that are spent on aid each year, it is remarkable that so little effort has been devoted to greater specialization. Various mechanisms could be suggested; for example, pairs of agencies, both assisting countries A and B, could agree that one agency will focus mainly on A and the other mainly on B. Benchmarking exercises like Easterly and Pfutze (2008), which explicitly penalize donors for assistance that is fragmented across many countries and sectors, might be one spur to reform.

12.2 Learning and accountability

The lack of donor coordination, and its adverse consequences for aid effectiveness, has been discussed since the 1960s. Berg (2001) and Easterly (2007a) have developed a wider critique of aid agencies, which are said to learn from experience too slowly, and to repeat costly mistakes that have been identified as such for several decades. Some of these criticisms may be unfair. For example, the World Bank's shift from *ex ante* conditionality toward selectivity represented a conscious change in direction, in response to analysis of past experience. The same could be said of the IMF's move away from loan conditions that prescribed multiple structural reforms. Sometimes, the Bank has been criticized for "fads," suggesting too much flexibility rather than too little. And responsibility for some long-standing problems, such as aid tying, must rest at least partly with politicians and not with the aid agencies. It also needs to be acknowledged that learning may be especially difficult for aid agencies, relative to other large organizations. Even in contexts where performance is much easier to assess, organizational learning remains something of a black box for economists. Not much is known about how organizations can improve their methods for gathering and aggregating relevant information, evaluating evidence, and reshaping policy.

One view is that any "top-down" perspective will be associated with slow learning, and inattention to local conditions and information (Scott, 1998). This argument may sometimes go too far. After all, most successful firms rely partly on top-down planning for their internal organization, in contexts where much relevant information is also locally held. But most firms operate within a competitive environment that is merciless in exposing poor performance, and in which failure is highly visible. It is much harder to tell when an aid agency is ineffective, given the complexity of the counterfactuals, and there are few reliable mechanisms for weeding out instances of inefficiency and error. The problem—a missing selection mechanism—is compounded by the tendency for some external commentary to be misguided or badly informed.⁵² Internal debate then becomes more

important, but external pressures could easily lead to a defensive attitude toward internal debate, for the sake of credibility and public unity. Aid agencies, as recipients of public funds, have strong incentives to avoid open divisions and self-criticism.

The case for slow learning should not be considered proven. It becomes more interesting when seen as a symptom of deeper problems: the lack of a selection mechanism, and the related charge that aid agencies are not sufficiently accountable (Easterly, 2006b). There are few ways for the intended beneficiaries to hold aid agencies to account for policy failings, or to communicate the information needed to address them. This particular diagnosis seems accurate, but often hard to take further. The lack of direct accountability to recipients seems likely to be an inherent feature of most efforts at external assistance, and the interesting question is the extent to which it is ameliorated by the commitment of aid agency professionals to their mission.⁵³ As Easterly (2007a) has emphasized, it would be interesting to know more about the incentives within aid agencies, and the extent to which their internal organization can work against their key objectives. Martins, Mummert, Murrell, and Seabright (2002) is a rare example of this form of analysis.

Although it may be hard to make donors more accountable to the intended beneficiaries, there might be scope for making them more accountable to their funders, and voters in donor countries. One way to promote this form of accountability is to benchmark aid agencies using numerical rankings, as in the work of Roodman (2008) on the “Commitment to Development Index,” and the study by Easterly and Pfutze (2008). The risks of benchmarking are well known, and it may distort future behavior rather than improving it. But the criteria in these studies seem mostly well-chosen, so that the exercises are informative. One reservation is that the individual components in the Easterly and Pfutze study are not highly correlated, so the overall ranking will be sensitive to the weighting scheme.

12.3 Randomized trials and evidence-based aid policies

There is one area which seems especially promising for aid agencies to pursue. Some recent discussions of aid have strongly advocated the use of randomized controlled trials, similar to clinical trials in medicine, to evaluate specific interventions. The case for recasting aid in these terms has been made by Banerjee (2007), Duflo and Kremer (2005), and Fisman and Miguel (2008), among others. The great strength of randomized trials is the ability to quantify the average causal effect of an intervention (or “treatment”) without the need for strong identifying assumptions. Measuring the treatment effect requires a counterfactual, but this is supplied by random assignment to a control group, rather than the use of a specific model or assumptions. Hence, at least at first glance, this form of evidence is not vulnerable to all the uncertainties that surround the use of observational data, and becomes unusually persuasive.⁵⁴

Existing applications of randomized trials already go beyond health interventions, to include specific educational interventions, the effects of credit constraints, and the management of public service workers. But aid agencies have sometimes been criticized for being too slow to adopt this approach, and to act on the associated evidence. Banerjee (2007) is one of the most prominent and articulate critics. He quotes Lant Pritchett to good effect:

Nearly all World Bank discussions of policies and project design had the character of “ignorant armies clashing by the night”—there was heated debate amongst advocates of various activities but rarely any firm evidence presented and considered about the likely impact of the proposed actions. Certainly in my experience there was never any definitive evidence that would inform decisions of funding one broad set of activities versus another (e.g., basic education versus roads versus vaccinations versus macro economic reform) or even funding one instrument versus another (e.g., vaccinations versus public education about hygiene to improve health, textbook reform versus teacher training to improve educational quality).

Randomized trials will help to inform such debates, because they provide a credible and transparent way to quantify causal effects, and increase the reliability of a cost-benefit analysis. This evidence is essential in assigning priorities, and will be relevant whether the choices are to be made by an aid agency, a government, or through grass-roots participation. From this perspective, developing the body of knowledge needed to assign priorities is one of the most valuable things that social scientists can achieve.

Some further strengths of randomized trials are worth emphasizing. The design of a trial has to confront practical details that are often glossed over in discussions of abstract concepts like decentralization or accountability. Randomized trials shift the focus of research attention toward concrete questions that may be vital to better policies. It is an attractive feature of papers using randomized trials that they often start from the world, and not from the literature. Besley and Burgess (2003), Duflo, Glennerster, and Kremer (2008b), Duflo and Kremer (2005), Easterly (2009b), and Kremer and Holla (2008) include overviews of some of the results to date. Banerjee and Duflo (2009) and Duflo et al. (2008b) develop a range of persuasive arguments for the approach: trials often bring to light unexpected but important questions, can help to develop and evaluate relevant theories, and have natural connections to behavioral economics in areas such as household saving and the demand for health. But the most convincing case for a research method is usually the one made in practice. Recent papers analyze many interventions relevant to aid effectiveness, such as possible responses to worker absenteeism in public services (Duflo, Hanna, & Ryan, 2007), deworming programs (Miguel & Kremer, 2004), anti-corruption measures (Olken, 2007), the effects of electoral platforms and social networks on the demand for public goods (Wantchekon & Vermeersch, 2005), and the effects of antiviolence campaigns on voter turnout (Collier & Vicente, 2008).

The main criticism of this evidence is that it cannot always be generalized to other contexts. In the useful jargon, randomized trials cannot establish external validity. Rodrik (2008a) provides a convincing example, from the study of charging for insecticide-treated bed nets, an antimalaria measure. The evidence from a trial in Kenya initially appears definitive (Cohen & Dupas, 2007). Closer, and nonquantitative, inspection raises some doubts, not about the evidence, but about the extent to which it will generalize to other contexts. Rodrik observes that when policies are chosen in practice, the “hard” evidence of trials will have to be supplemented with a variety of “soft” evidence, which is not what the most prominent advocates have in mind.

One counterargument is that trials carried out in different contexts can establish how the effects of the treatment vary with the context (Banerjee & Duflo, 2009; see also Duflo et al., 2008b). This is a marked improvement on more traditional methods, which find it hard to separate genuine context-driven heterogeneity from the confounding variation that arises from questionable modeling assumptions. But there must be some doubt about the extent to which multiple trials are feasible, and about the professional incentives to carry them out (Rodrik, 2008a). And without an understanding of mechanisms, it will be hard to say much about why a treatment effect varies across different contexts. A meaningful synthesis of the evidence from multiple trials will sometimes be harder to achieve than their advocates imply.

These points show that the problem of extrapolation cuts deep. Put simply, the average treatment effect rarely corresponds to a well-defined structural parameter. If the effects are heterogeneous across either individuals or contexts, or both, it seems essential to investigate the underlying mechanisms. As Deaton (2009) emphasizes, heterogeneity should not be seen as just a technical problem, requiring an econometric solution: it represents “a symptom of something deeper, which is the failure to specify causal models of the processes we are examining.” He cites the useful perspective of Pawson and Tilley (1997): it is the combination of mechanism and context that generates outcomes, and scientific progress relies on understanding that combination. In the long run, the most useful trials will often be those designed to explore mechanisms and the predictions of specific theories, rather than evaluating the mean impact of discrete projects.

Inevitably, randomized trials do not answer all the questions relevant to foreign aid, or even “just” the delivery of public services. If academics are sometimes caricatured as looking for lost keys only under the streetlight, randomized trials can be seen as building new and better streetlights, but only along certain stretches of the road. This seems obvious, and at first glance, does not leave much room for controversy. If there is nevertheless unease about some of the more assertive promotion of randomized trials, it arises partly through awareness of some important questions that will never be resolved with similarly clean identification, including a range of questions that are central to aid effectiveness. Academic research cannot always aim at definitive answers, and

sometimes must work away at narrowing the space of admissible claims, or shifting the burden of proof in complex debates. In such a world, to imply that observational data is inadmissible would carry significant risks (Rodrik, 2008a). Although some policy questions would be resolved by randomized trials, many others would have to be settled on gut instinct, which might be defined as empirical work on observational data of a very low quality. From that perspective, it would be a dangerous mistake to suggest that randomized trials are the only way forward.⁵⁵

Short-lived controversies aside, the applications of randomized trials have established their value, to the extent that it is already hard to imagine development economics and the analysis of foreign aid without them. In terms of aid effectiveness, the greater controversies for the future are likely to relate to how and when the evidence is used. Translating aggregate evidence into practice can be a more uncertain process than might be expected, as the actual practice of medicine sometimes attests. Easterly (2009b) argues that there is a crucial divide in aid practice between those who feel bound by evidence and those who do not. The ultimate impact of randomized trials on aid effectiveness will partly depend on incentives, organizational learning, and mechanisms for accountability; in other words, on the same forces that have always determined the use of evidence in shaping donor policies. Since randomized trials make relatively few assumptions in establishing average treatment effects, they will strengthen the hand of those committed to evidence-based aid, and this is a substantial gain. But even this evidence is easier to dispute than some of its advocates have fully acknowledged, because of the doubts that can arise about its interpretation and its external validity.

12.4 Evaluate, evaluate, evaluate

There is a traditional answer to many donor failings: “Evaluate.” In discussing evaluation, it is worth distinguishing between two of its roles. The first role is to increase the knowledge base, helping donors learn which ideas work. The second role of evaluation is to improve incentives to act effectively, at several levels: formal evaluations can be used to make aid agencies more accountable to their funders, and the employees of agencies more accountable to their managers. If more use of systematic evaluation improved incentives, this could be even more important than extending the knowledge base.

Evaluation is hardly a new idea. Donors have often evaluated their own projects, and Riddell (2007, p. 180) estimates that hundreds of thousands of project completion reports have now been written. Some of these estimate rates of return, but others evaluate success in broader terms. Concerns might then be raised about the undesirable interaction between a lack of independence, and imprecise criteria for success. The desirability of independent evaluations has often been discussed. Birdsall (2008) recommends the establishment of a separate agency to commission evaluations, financed by a

fixed contribution from each donor. An alternative approach would be to transfer the responsibility for evaluation largely to local agencies within aid recipients, as van de Walle (2005, pp. 93–95) recommends.

Evaluations will work best when at least some of the criteria for success are objective. In some contexts, the evaluation methods could draw on the related literature in labor economics, as Levine and Savedoff (2006) recommend. Randomized trials are not always feasible or appropriate, but aid agencies could still make more use of pilot schemes and before-and-after research designs, such as differences-in-differences. These evaluations could be managed internally, contracted out, or undertaken by an independent agency.

There is fairly general agreement that effective evaluation would be highly beneficial.⁵⁶ The more subtle issue is that any evaluation procedure will be imperfect in practice. As the history of public service reform sometimes attests, there is a risk that imperfect evaluation distorts behavior, offsetting the benefits. For example, one risk is that donors become too cautious in selecting projects and countries, something that Collier (2007) warns against. This implies that the details of evaluation are critical, and signing up to the broad idea is not enough. But working out these details should be a priority. Although the criticisms of Easterly (2006b, 2007a, 2008) range widely, one important lesson is that introducing more precise mechanisms for evaluation should be a priority for donors, and would have multiple benefits.

13. MEETING BASIC NEEDS

Spurred by the Millennium Development Goals, and following a trend in policy debates in developed countries, the aid literature has started to emphasize meeting basic needs through better public services. It has become common for aid agencies and NGOs to emphasize the social sectors, especially health and education, rather than growth objectives, infrastructure, or rural development. Academic research has devoted greater attention to the many incentive and management issues that complicate the delivery of public services.

The emphasis on basic needs can be seen as a natural counterpart of the new partnership model: if donors and recipients have to agree on objectives, meeting basic needs is a good place to start. It provides an area of common ground not only for donor and recipient governments, but also national and transnational NGOs, and helps to bolster public support for aid in donor countries. Some observers regard the Millennium Development Goals as essentially a political success: agreement on goals has concentrated minds on the steps needed to achieve them. The problem with the Goals as a political tool is that they may be unrealistic: the inevitable failure to meet them risks undermining faith in the whole aid enterprise (Clemens et al., 2007). Not for the first time, aid may be judged harshly because it has failed against a benchmark that was always out of reach.

In the long term, a more fundamental concern is that targeting basic needs becomes the line of least resistance for aid agencies and NGOs. The equilibrium outcome may be that “too many” donors concentrate on this agenda at the expense of others, especially when it is favored by the public in donor countries. The balance would then have to be redressed by especially large donors, such as the World Bank. For example, infrastructure provision can be controversial, raising issues such as corruption and environmental impact. Projects such as dams will often create losers as well as gainers, in contexts where compensation is difficult.⁵⁷ It is easy to see how this kind of project may fall from favor with some donors, requiring others to make a conscious effort to look beyond basic needs.

13.1 Public service delivery

For developed countries, the literature on public services is engaged with issues such as the role of intrinsic motivation, the desired degree of competition between providers when performance is hard to observe, and the distorting effects of evaluation. But the problems for developing countries are often more basic, as Banerjee, Iyer, and Somanathan (2007b) show in the course of their review. A study in which enumerators made random checks on primary schools and health clinics in six developing countries found that, averaged across the countries, roughly a fifth of teachers were absent, and a third of health workers. In India, around half the teachers present in school were not teaching when the enumerators arrived. Filmer, Hammer, and Pritchett (2000) discuss other problems, such as rural health clinics that rely on untrained staff and which have no medicines available, sometimes because supplies have been sold on the black market.

The recent literature on public services often assumes that the traditional model, based on a centralized and hierarchical civil service, risks embedding some degree of incompetence, low-quality service, or even dishonesty. A range of options has been canvassed, and Pritchett and Woolcock (2008) list no fewer than eight alternatives.⁵⁸ These include various forms of decentralization, contracting out, and demand-side financing through voucher schemes. Considering only the specific case of schooling provision, many different proposals have been advanced, including federalization (Argentina, Brazil), localization (Indonesia), school autonomy (Nicaragua), vouchers (Chile), community control (El Salvador), and contracting out to NGOs. Which of these will work best must surely vary with the context. For example, the success of decentralization might rely on local accountability and grassroots participation. Reinikka and Svensson (2004) showed what can happen in their absence. They used an expenditure tracking survey for Uganda in the early 1990s to show that funds provided by central government to local districts, earmarked for use by schools to cover nonwage costs, rarely reached the schools: about three-quarters of the schools received less than 5% of the funds, and 90% received less than half. Reinikka and Svensson

(2005, 2007) showed how this problem might be overcome by developing mechanisms for accountability. Björkman and Svensson (2009) similarly analyze the role of community-based monitoring, using a randomized trial.

For the countries with the least effective public services, the right approach might be to direct funds through NGOs and other independent service providers, rather than governments. There are several possible objections. Independent provision might lead governments to cut back on state provision, and slow down the capacity-building of the domestic state. Another concern is that routing funds through other providers may simply relocate the relevant agency problems.⁵⁹

For donors and developing countries, there could be a useful parallel between the literature on aid and public service delivery, and the literature on growth and binding constraints (Rodrik, 2008a). Similar cautions apply. There is no unique model for public service delivery that can and should be applied universally. One reason for caution is that developed-country experts may be effective within good institutions without knowing how to build them elsewhere: relying on their advice is too much like sending a cab driver to design a car (Pritchett & Woolcock, 2008). Rather than attempt wholesale reform, it would be better to identify the key constraints that prevent important needs from being met, and concentrate efforts on those. Given the different ways of reforming public services, information is needed on which ideas will work well in particular contexts, suggesting that this area ultimately needs its own decision tree. And as with growth policy, a relatively incremental approach will make fewer demands on information and political capital, and allow experimentation and the monitoring of unintended consequences. One of the strengths of the randomized trials literature is that it has already identified some practical, incremental reforms that could significantly improve the delivery of essential services. Examples include the work on education of Banerjee, Cole, Duflo, and Linden (2007a), Duflo, Dupas, and Kremer (2008a), and Duflo, Hanna, and Ryan (2007).

13.2 Agriculture and rural development

At least temporarily, the emphasis on education and health has obscured the role of alternative forms of aid, especially measures aimed at agriculture and rural development. Although the debate rarely figures explicitly in the recent literature, donors must confront a long-standing dilemma, namely whether to address the sectors or regions where the poor are mostly found, or to promote sectors or regions where the poor are not found, but which could improve their prospects.⁶⁰ In general equilibrium, living standards in rural areas might be improved by rural development, but also by the entry of new activities and productivity growth in urban areas. Growth economists have tended to emphasize the benefits of industrial growth, but in the past, donors have tried to promote agricultural productivity and rural living standards more directly.

The verdict of history on these efforts has been equivocal at best. Easterly (2009b) provides a brief overview of policies on agriculture, and cites evidence that the proportion of World Bank agricultural projects judged successful has been lower than for other categories of aid. Based on evaluation data from the World Bank and some bilateral donors, Cassen and Associates (1994) also presented a mixed picture, marked by considerable variability across projects, and lower success rates in Africa. They note the variability across types of project, with irrigation and credit provision associated with higher returns than livestock projects and area development, such as integrated rural development. Scott (1998) uses the record of agricultural extension in developing countries as part of his more general case against oversimplified, top-down approaches that neglect existing practices and knowledge. Pritchett and Woolcock (2008) discuss similar failures of donor efforts to improve rural water supply and irrigation. As for donor emphasis on agriculture, the fact that the two *World Development Reports* on agriculture are separated by 26 years (1982 and 2008) tells its own story. Perhaps more striking still, Atwood, McPherson, and Natsios (2008) cite USAID assistance to Ethiopia in 2007, in which agriculture accounted for just 1.5% of expenditure, compared to 50% for HIV/AIDS prevention and 38% for emergency food relief.⁶¹

Easterly (2009b) argues that the diminished emphasis on aid for agriculture may represent an efficient response to past failures. Nevertheless, the strategy of addressing rural living standards as directly as possible has merits beyond directness, such as helping to relieve migration pressures on urban slums. The work of Caselli (2005) and Restuccia, Yang, and Zhu (2008) indicates that low agricultural productivity plays a central role in explaining low levels of GDP *per capita*. With all this in mind, it seems odd that mainstream aid research currently gives so little attention to rural development, and the scope for doing better than in the past. Chen, Mu, and Ravallion (2009) study the effects of a rural development program in China, but this is a rare exception. Once again, randomized trials seem promising: Duflo, Kremer, and Robinson (2008c) use trials to clarify the conditions needed for the effective use of fertilizer. At the same time, it is easy to imagine the criticisms that Scott (1998) would raise, and the fact that policies in this area have often failed is hardly irrelevant. As the investor John Templeton once noted in a different context, the most expensive words in the English language are “It’s different this time.”

14. AID FOR THE TWENTY-FIRST CENTURY

Recent discussions of foreign aid have been dominated by the anticipated scaling-up of aid flows. At the same time, there is a sense of uncertainty about how much aid can and should be expected to achieve. Kremer and Miguel (2007) sketch the history of foreign aid in terms of attempts to identify ever more fundamental causes of poverty: from the

early emphasis on famine relief and capital shortage, through to macroeconomic policies and then institutions. This could be seen as progress, except that each new cause of poverty is harder to address than the previous one. In discussing policies toward Africa, Easterly (2009b) argues that the back-and-forth of aid debates has involved a cycle in which ideas are tried, abandoned, and then tried again. He also criticizes the simultaneous escalation in what aid has been trying to achieve, so that the overall pattern has not been a dialectic of thesis and antithesis, but something more like an unstable cycle, spinning ever further away from the feasible and well understood.

The case against scaling-up is that it starts from wildly unrealistic targets, and then proceeds as if aid is simply an exercise in matching donor funds to unmet human needs. Easterly (2006a) criticizes the way in which the more ambitious proposals, such as Sachs (2005), echo the simplistic thinking of the 1950s and 1960s. The skeptical view is that each successive generation is told that it is uniquely well placed to redress global poverty, and can achieve this by throwing ever-greater amounts of money at the problem. The result is technocratic plans that are too wide ranging, when less ambitious and more focused schemes, with scope for learning, would stand a greater chance of success. And, however, ambitious the planning, there are few aid agencies that would not benefit from better mechanisms for accountability, evaluation, and organizational learning. Easterly's view seems to be that donors should address the weaknesses of their current practices, rather than simply replicate them on a larger scale.

As this begins to make clear, the debate on scaling-up has taken its clearest shape in the contributions of Easterly and Sachs. Their well-publicized exchanges make it difficult to define each of their positions without making at least some reference to the other.⁶² As many have observed, including Easterly, the tireless efforts of Sachs to raise the profile of world poverty and foreign aid deserve great credit. And some of the criticisms of his work risk unfairness. Sachs (2005) emphasizes the importance of differential diagnosis, attentive to the circumstances of individual countries. In a similar way to Collier (2006b, 2007), he has analyzed aid within the context of specific traps. This remedies an occasional tendency in the aid literature to discuss its effectiveness in isolation, without specifying the concrete obstacles to development that aid should be designed to overcome.

But Sachs is not always explicit about differential treatment, and a further objection to the Sachs worldview is well captured by Deaton (2008). He suggests that Sachs often seems to think of the developing world as a broken machine, with a list of missing parts: antimalaria bed nets at so much each, vaccines at so much a dose, roads at so much a mile. But the academic literature suggests that this is often the wrong way to think about aid. It risks ignoring the complex interactions between aid and governance. Even taken on its own terms, there are considerable difficulties in getting the right missing parts to the right places in the machine. This has been observed many times before: Scott (1998) and Easterly (2006a, 2006b, 2007b) are only the latest to criticize

the view that deep-seated conditions can always be reduced to a known set of technical subproblems, and solved by centralized, top-down planning by governments or donors.

These debates are linked to a central challenge, to make aid work more effectively in Africa. The evidence from project evaluations, case studies, and cross-country regressions often suggests that aid has been less effective in Africa than elsewhere, rarely delivering higher growth rates. Yet as countries in other regions move toward middle-income levels of development, Africa is also the area where the world's poor are increasingly concentrated. And as noted earlier, as some countries graduate from low-income levels, those that remain will inevitably be the countries facing the most severe obstacles to development, and where innovative approaches may be needed now more than ever.

The new emphasis on Africa means that issues of governance and political economy will be central to the analysis of foreign aid effectiveness. Observers such as van de Walle (2005) emphasize the role of local political dynamics, clientelism, and presidentialism, with important policy implications, such as conditionality based on term limits. In other cases, policies toward fragile and postconflict states take center stage, introducing a degree of practical and ethical complexity undreamt of in the postwar emphasis on capital shortage, and that can easily make external commentary seem glib. If the central development problem is not a capital shortage but a governance shortage, we are a long way from knowing how to address it, and the answers will need to combine expertise from political science, history, philosophy and sociology, as well as economics.

14.1 New directions for foreign aid

The rest of this section will discuss some new directions for aid. The chapter has already covered some long-standing policy debates; the aim now is to discuss some recent controversies, and some of the most innovative ideas for new forms of aid. Many of the suggestions for better external assistance can be summarized as “narrow the target and broaden the instruments” (Collier, 2007, p. 192). It has become a commonplace to warn against the dangers of attempting too much at once, and greater attention is now focused on new instruments.

Easterly (2006b) recommends that ambitious top-down planning is abandoned altogether, in favor of decentralized competition among “searchers.” In principle, donors and recipient governments could seek to foster decentralized searching for solutions to development problems. This is the idea behind social entrepreneurship or “social business,” with the Grameen Bank in Bangladesh as the best-known example. It may be possible to design institutional environments in which entrepreneurs and organizations are successful only to the extent that they provide goods and services valued by the poor. The role for the donor is sometimes clear, as in subsidizing micro-credit schemes, but elsewhere the task seems harder. Easterly (2006b, pp. 330–332) has suggested more use of voucher schemes. Donors should allocate vouchers to the poor,

and donor money then follows the holder of the voucher through the door of a school or health clinic, whether it is run by the state, an NGO, or the local community. This ensures that resources are placed directly in the hands of the poor, and encourages competition among health and education providers. This kind of scheme still requires some top-down planning, but also embeds a selection process that will reveal information to individual providers, potential entrants, and the planners.⁶³ The work of Angrist, Bettinger, and Kremer (2006) on school vouchers, allocated by lottery in Colombia, found long-term benefits for graduation rates and test scores; but these results arise in a context with many private schools already in existence, not the case for all poorer countries. One objection to voucher schemes is that, to work well, voucher holders need to be able to discriminate between the qualities of different providers, in circumstances where this is rarely straightforward. Easterly recommends that a pilot scheme is tried.

Another radical strategy would be to adopt a foundation model. Donors would invite applications, and then make decisions about which projects to fund. The applications could be from the governments of developing countries, as van de Walle (2005) proposes, and thereby promote country ownership of aid programs. But such a scheme is less desirable if general equilibrium effects, like the Dutch Disease, are central to the effects of aid. An alternative version of the foundation model would prioritize Easterly's "searchers," distributing funds within countries not only based on proposals from central and local governments, but also to NGOs and social entrepreneurs. The obvious problems include the difficulty for donors in ranking projects, and in establishing whether applicants have a strong track record.

Others have emphasized the potential of instruments beyond direct financial transfers (Birdsall, Rodrik, & Subramanian, 2005; Collier, 2007; Deaton, 2008; van de Walle, 2005). Donors could devote more resources to global public goods, including research in tropical medicine and agriculture, innovations suitable for use in low-income countries, and investment in the body of knowledge represented by randomized trials. Here, Bhagwati has made the useful distinction between aid spent *in* Africa, and aid spent *for* Africa. There are obvious market failures in the provision of global public goods, and remedying them would make it easier to meet basic needs in Africa.⁶⁴ The search for new drugs and vaccines could use a combination of "push" (direct financial support) and "pull" mechanisms (financial incentives for the private sector to carry out the necessary research). One of the most impressive contributions is the work of Kremer and Glennerster (2004) on advance purchase commitments for vaccines, creating a prize that can then be pursued by the private sector; Kremer (2008) provides an introductory overview. The case for research on tropical agriculture may be less clear-cut, since improving agricultural productivity will require not just new technologies, but also solving problems that include infrastructure provision, education, and training, and incentives for farmers; see Easterly (2009b) for discussion

and references. A newer candidate for an international public good is the body of knowledge represented by randomized trials; see Duflo and Kremer (2005) for the potential role of international agencies in supporting this approach.

In multiplying the possible forms of aid, it is not clear how donors should judge the appropriate balance between them, and between more traditional forms of aid, as Mosley and Hudson (2007) emphasize. Pilot schemes might be desirable, such as research on tropical agriculture focused on a small set of well-defined problems, and taking into account incentives as well as narrower technological questions. In some areas, such as funding prizes for vaccines, the recent proposals could be a candidate for new sources of development finance, such as the possibility of a global lottery discussed in Addison and Chowdhury (2005).

Other suggestions include the strengthening of institutions with a regional focus, including mechanisms for peer review and oversight. A concrete example is the African Peer Review Mechanism introduced by the African Union. The Commission for Africa (2005) recommended that rich countries support such initiatives, which might achieve a great deal for small sums of money. Easterly (2006b, p. 129) argues that such proposals are misguided, because what matters is not accountability to other governments, but to the domestic population. This is not wholly convincing, because the domestic population still needs information, and the relatively impartial judgments provided by peer review will then help to strengthen domestic accountability.

Another international measure would be for rich countries to give imports from Africa preferential treatment, relative to imports from Asia, through differential tariffs. The leading example is the African Growth and Opportunity Act (AGOA) of the US, enacted unilaterally in 2000. Even this imperfect measure appears to have raised imports from Africa, especially in clothing. Exploiting the selective application of the Act, across countries and products, Frazer and Van Bieseboeck (2009) find that the effect was to increase clothing imports from Africa by more than 40%, with smaller increases in other manufactured products covered by the act, and in agricultural products. They estimate that the overall causal effect of AGOA was to raise Africa's nonoil exports by 6.6%, a large effect for a relatively modest program. Collier (2007) suggests extending the horizon for the duration of AGOA to encourage investment, and adopting a similar policy across the OECD.

One consequence of globalization, including the global media, is that domestic politics and societies are increasingly influenced by external factors. Donors could seek external policies that change the incentives and constraints faced by domestic actors. Well-targeted measures, such as international charters, could limit the incidence and effects of weak governance, and strengthen domestic civil society (Collier, 2007). One of the best-known measures, the Extractive Industries Transparency Initiative, is an example of external coordination that should help to limit the misappropriation of resource rents, and reduce corruption in the allocation of contracts.

This general approach is unlikely to be costly, and has less chance than conventional aid of being actively harmful. The expected net benefits could be high even when the probabilities of good outcomes are only marginally affected. But to be effective, these measures require international coordination, and the extension of the development agenda to arms of government beyond specialist development ministries and aid agencies.⁶⁵

The idea of modifying domestic incentives by external action could take other forms, as in Oechslin (2006). It might be possible to intervene in the provision of technology licenses, and certain goods that are essential to meeting basic needs, such as medicines and school textbooks, or capital equipment used in agricultural processing or labor-intensive manufacturing. In principle, actions that lowered the world prices of these goods would encourage developing-country governments, citizens, and firms to substitute toward them, which is desirable if their use brings external benefits. This approach would be less reliant than traditional policies on the objectives and capacities of recipient governments.

But some of these proposals could be modest in their reach, and it is clear that some problems are likely to get worse rather than better, on a scale that could overwhelm the current international institutions. The record of rich countries in dealing with new and unfamiliar problems is not a strong one. Easterly (2006b) gives an account of the West's failure to respond more quickly to the HIV-AIDS crisis as it emerged, especially in sub-Saharan Africa, that is one of the most passionate and effective chapters in his book. The related and more detailed analysis of Epstein (2007) is also salutary, highlighting the basic ways in which outside interventions have gone wrong, and how simple solutions have too often been overlooked.

These policy failures do not augur well for the major problems gathering on the horizon: the dangers posed by climate change, and the separate but related possibilities of widespread water shortages, climate stress, and food insecurity. The literature is only just starting to engage with the implications of climate change for development and aid; see Collier, Conway, and Venables (2008) and Hepburn and Stern (2008). A coordinated policy on climate change could introduce a new source of transfers to poor countries, namely payments from rich countries for emissions rights. But the global policy challenges raised by these various issues are clearly enormous, and it has become trite to say that the current international agencies lack the capacity to anticipate and respond to new challenges on this scale, and coordinate an effective response. The urgent need is for practical institutional solutions that have some chance of being adopted even in the absence of well-informed public debate. The precedent of HIV-AIDS has shown that concerted public pressure emerges only when a crisis has deepened enough to become widely visible, and for many that will be too late.

15. CONCLUSIONS

Many observers have expressed frustration that the imagination and ambition needed to make a serious difference to world poverty are too often absent. But in discussing foreign aid, a persistent danger is that the best becomes the enemy of the good. It is not enough to say that aid is flawed, or falls short of its goals. That is true of much human endeavor, and certainly does not imply that the entirety of aid is wasted or harmful. The danger of asking too much becomes clearer when aid flows are calculated *per capita*: the sums involved are not vast, even in the poorest countries, as Tarp (2006) emphasizes. Some of the evidence suggests that aid can be effective, and the hypothesis that aid may be actively harmful does not have the same backing. The arguments that aid can weaken accountability and institutional development are potentially fundamental, but much of the supporting evidence remains impressionistic and anecdotal.

Although much remains unknown, arguably enough has been learnt that aid and conditionality can be more effective in the next few decades than in the past. By assigning a central role to governance and institutions, academics and policy makers may have finally identified the most important binding constraint on growth, development, and aid effectiveness. Donor policies are responding to this, and evolving toward greater flexibility in approach, changes that are assisted by improvements in governance in a significant number of poorer countries.

The task of donors is much easier when their objectives are aligned with those of the recipient government. In such cases, a combination of direct budget support, project aid, and policy dialog is likely to be uncontroversial. When applied to this subset of countries, the new partnership model has clear strengths, and is not simply rhetoric. The emphasis of the partnership model on autonomy, balanced by accountability, has ethical and practical justifications. It recognizes that attempting to impose detailed reforms from outside is invasive, places too much faith in the expertise and leverage of outsiders, underestimates the need for the support of a domestic constituency, and can undermine domestic political development. Basing the aid relationship on a long-term partnership means that budget support is a credible way to address the unpredictability and lack of coordination that have undermined past aid efforts. Hence, the new partnership model has the potential to enhance the effectiveness of aid.

Yet the partnership model is a less than complete response to critics of current aid practices. Donors and recipient governments still have to find ways to allocate funds efficiently, and to ensure that flawed projects and programs are identified and weeded out. One effect of the partnership model may be to refocus attention on these practical questions. Randomized trials strengthen the necessary evidence base, but the academic literature on aid has largely failed to address the mechanisms by which donor policies are formulated, evaluated, and modified over time. Nor does it say much about how to improve accountability, beyond benchmarking exercises. It gives relatively little

attention to the potential role of aid in stabilizing countries that face major shocks. In terms of the overall approach, it remains unclear whether the best way forward is improved planning or more radical alternatives, such as outcome-based aid, social entrepreneurship, and voucher schemes.

There are other reasons to be distinctly wary of a Whig history of foreign aid, based on a complacent narrative of progress toward enlightenment. Most obviously, the partnership model only resolves problems for those countries where the problems for donors were least serious to start with. The policy dilemmas are much greater when recipient governments are not committed to growth and development, or where governance is weak. And here the partnership model does not have a great deal to offer. In these cases, the key question is whether aid, when combined with governance conditionality, can achieve better results than in the past, and avoid the negative effects associated with resource revenues. Although the evidence for a resource curse remains ambiguous, the parallel continues to raise some awkward questions, perhaps the most awkward, for advocates of increased aid. Such advocates need considerable faith that conditions on governance, and other innovations such as floating tranche conditionality, will work better than the policy conditions of the past.

It should also be reiterated that direct financial transfers are not the only way forward. Donors should actively pursue innovative forms of development assistance, especially targeted at the problems faced by Africa. The market failures that arise in global public goods, notably global public health, represent a clear justification for new forms of large-scale aid. This approach should avoid many of the traditional objections to aid. Some forms of global action, such as charters, norms, and differential trade policies, would help to encourage development, and might strengthen domestic governance and civil societies, with only modest downside risks.

This chapter's discussion of the various issues, and all that remains unknown, also invites some closing reflections on whether economists can meet the relevant challenges. There are clearly areas in which economists can make fundamental contributions by the application of existing tools. These include general equilibrium analyses of macroeconomic phenomena like the Dutch Disease, the study of incentives as one force in shaping political outcomes, and the design of the statistical evaluations that are needed to assign priorities. But, to borrow an idea from Goethe, foreign aid divided by economics leaves a remainder. And if every academic field involves some degree of professional deformation, then reading the aid literature can increase a sense of the deformations specific to economics. In trying to assess the effects of aid on growth and conflict, many economists instinctively reach for regressions, when close historical study would sometimes be more informative. When talking about institutions, the limits of the tools and methods currently taught in graduate schools become even more apparent. What may sometimes be needed here is something closer to inductive research, based on patient observation and analysis; at times, something closer to

anthropology than mathematics or physics. As things stand, one can spend a long time reading the best economics journals without finding much of interest on how successful institutions work in practice. That is a loss for the discipline. Rather more fundamentally, it is a loss for the millions whose destinies are linked, for better or worse, to the effectiveness of foreign aid.

APPENDIX

This appendix reviews issues that arise in using cross-country data to study the effects of aid. The main focus is on growth, but many of the issues are also relevant to the study of other outcomes, such as conflict, governance, and social indicators.

Data

As discussed in Section 2, there are several measurement issues that arise when relating growth to aid flows. For present purposes, the most important relate to the aggregation of grants and loans; whether or not to include technical assistance; and whether to disaggregate aid further by its intended purpose, or perhaps in terms of its origin and the objectives of the donor. Burnside and Dollar (2000), and several of the subsequent studies, use data constructed by Chang, Fernandez-Arias, and Serven (1998). Their measure, called effective development assistance (EDA), differs from the standard net ODA measure in two ways. First, it excludes technical assistance; and second, it replaces the face value of loans with the net present value of their grant element. Since most aid is in the form of grants, the EDA measure is often fairly close to ODA. Another aspect of Burnside and Dollar's approach is that they deflate aid inflows for recipient countries by an import price index, and then divide by GDP measured in PPP terms. In contrast, Rajan and Subramanian (2008) appear to use the simpler approach of dividing nominal ODA by nominal GDP.

Endogeneity and control variables

There are many reasons why aid might be correlated with the error term of a growth regression. This is sometimes interpreted simply in terms of reverse causality: for example, negative shocks to growth may induce aid inflows. It is important to clarify, however, that the endogeneity problem goes much deeper than reverse causality. The fundamental problem is that aid is not randomly assigned, and its allocation to individual countries will depend on country characteristics. The lack of random assignment not only makes simple correlations uninformative, but also makes a regression approach deeply problematic. Since some of the country characteristics that determine aid allocation are likely to be unobservable to the econometrician, even a systematic approach to the choice of control variables will not be sufficient for clean identification. The likelihood of finding persuasive solutions is further undermined by the small number of countries available to researchers.

An IV approach to the endogeneity problem has become especially popular since Boone (1996), with precursors that include Gupta (1975) and Mosley (1980). This approach is not a panacea, especially in samples of the available size. The bias of 2SLS can be larger than the OLS bias for small departures from the identifying assumptions, and in particular, given even a weak correlation between the instrument and the error term in the growth equation. Yet some of the exclusion restrictions in the aid literature lack credibility, especially when the controls in the structural equation are selected on an arbitrary basis.

Boone's chosen instruments were population size, exploiting the small-country bias in aid allocation, and variables capturing political or geostrategic alliances and historical connections, such as those arising from colonization. These and related instruments have become popular choices in subsequent work. Deaton (2009) expresses skepticism, but arguably population size is a more convincing instrument than many of those adopted in the wider growth literature. This is because it has rarely been found to be a robust determinant of growth rates, and has a low posterior inclusion probability in comprehensive Bayesian studies of growth such as Sala-i-Martin, Doppelhofer, and Miller (2004). The remaining problem is that population size is not useful in panels, because its variation within countries is modest relative to the variation across countries.

Since Boone's work, the use of historical and strategic connections has played a key role in several important studies, including Rajan and Subramanian (2008). This approach is vulnerable to the charge that these connections have not emerged randomly. More importantly, the use of strategic connections assumes that the marginal effect of aid on growth does not vary with the intentions of the donor (Headey, 2005, 2008). This seems unlikely, since bilateral aid has often been given for strategic reasons. The essence of such aid is that development is not its sole purpose, implying that considerations such as need, accountability, and governance will carry less weight. Strategic aid could therefore be less effective, at least on average, and it is this smaller effect that will be captured when strategic alliances are used as instruments.

Panel data and dynamic responses

In the long term, and especially as longer spans of data become available, the use of panel data is likely to be the most attractive approach, as in Dalgaard et al. (2004) and Hansen and Tarp (2001). This method seeks to identify the causal effect of aid from the within-country variation in aid and growth, and is thereby immune to the problems that arise when aid allocation is based on fixed country characteristics.

The new problem is the likely relevance of time-varying confounders. If donors respond to events like leadership changes or reforms with greater aid flows, the estimated association between aid and growth may be positive even when there is no causal effect. Researchers have not yet established genuinely persuasive instruments

for aid in a time-series context. The use of lagged aid seems especially questionable. Many of the GMM panel data studies use identifying assumptions that rule out delayed effects of aid, but the economics of this are questionable, as much of this chapter attests. Pragmatic arguments based solely on statistical tests will leave many unconvinced.

A further problem is that GMM methods have occasionally been applied mechanically, sometimes with a lack of genuine understanding of the estimators.⁶⁶ Part of the problem is that, given the nature of cross-country samples, the GMM approach can be highly sensitive to the chosen identifying assumptions. This puts a premium on a thoughtful choice of assumptions and genuinely rigorous testing. As longer spans of data become available, it is possible that the best approach will be to apply instrumental variable methods to fixed-effects models, rather than eliminate the fixed effects by first-differencing.

This point is related to more general criticisms of the empirical literature on aid and growth. In the majority of the recent empirical studies of foreign aid, the headline results are based on a short-run association, often driven by a contemporaneous relationship estimated on the basis of 4-year averages. Instead, the panel structure could be exploited more often, so that causality is inferred partly from timing, and the role of lagged aid in determining current output. Although this approach is vulnerable to the criticism that aid may anticipate future growth, it could still be a useful complement to IV methods.

Error dependence

Since countries are influenced to varying degrees by common factors, such as the world business cycle and developments at the regional level, it seems likely that the errors in cross-section and panel data models will not be independent across countries. This is yet another reason for treating the existing studies with caution. Error dependence means that conventional standard errors will tend to overstate the information in the data, and when unobservable factors influence country outcomes to varying degrees, there will be an omitted variable problem that time dummies cannot fully address. Methods for testing and improved inference are starting to become available, and the expectation must be that corrected standard errors will often be higher than those reported in the literature.

Heterogeneity

One of the most important advances in the recent literature has been the introduction of various kinds of heterogeneity and nonlinearity in the effect of aid. The relevant variation could occur across countries, across time, and across different categories of aid. One reason that heterogeneity receives so much attention is that it often has direct and concrete implications for the policies of donors. There is a great deal at stake here, which makes some of the current treatments of heterogeneity look disconcertingly simplistic.

One idea is that aid involves diminishing returns: as aid to a given country increases, it may become harder to use it effectively. To capture this possibility, growth is often modeled as a quadratic function of aid relative to GDP.⁶⁷ To the extent that this effect is present, estimating a simpler linear relationship could be misleading about the benefits of aid. Yet we should be careful before placing too much weight on this empirical finding. An estimated quadratic will often be sensitive to outlying observations, and even where the quadratic term is statistically robust, diminishing returns are not the only possible explanation.⁶⁸ The overinterpretation of results becomes especially dangerous when the turning point of the estimated quadratic function is used to calculate a limit for aid intensity, beyond which further aid is said to become ineffective. Notwithstanding many other problems, such exercises rarely calculate a confidence interval for the turning point. Since the turning point is typically based on a ratio of parameter estimates, it will be hard to identify precisely in the available data.

One widely discussed claim, initially associated with Burnside and Dollar (2000), is that aid is most effective in certain environments. The economics of this claim, and its implications, are covered in more detail in the main text. The statistical evidence is not overwhelming. The best-known findings are based on pooled cross-section time-series models, as in Burnside and Dollar (2000) and Collier and Dollar (2001, 2002). Sometimes the aid term is insignificant, but the effect of an interaction term between aid and policy can be estimated more precisely. These findings do not appear especially robust, however. Dollar and Levin (2006) provide references to many studies that question the Burnside and Dollar analysis. The cross-section study by Kourtellos, Tan, and Zhang (2007) finds considerable instability in regression tree models, which suggests that threshold effects and interactions—whether based on good policy or other variables—are sensitive to modeling assumptions.

Much of the literature has focused on heterogeneity in terms of recipient characteristics. But as when discussing the IV approaches used in the literature, the marginal effects of aid could also vary with the nature, intentions, and policies of the donor. This could give rise to heterogeneity both over countries and over time, as Bobba and Powell (2007), Headley (2005), and Minoiu and Reddy (2007) have recently argued. For example, with the end of the Cold War, and the weakening of strategic motivations for aid, the marginal effect of aid in the 1990s might be stronger than in previous decades. Across countries, perhaps bilateral aid from “enlightened” donors has a larger marginal effect than aid given for strategic reasons; for similar reasons, aid from multilateral donors may have greater benefits than bilateral aid. But these are subtle effects to be readily discernible in the cross-country data. When bilateral aid from relatively small donors (in absolute terms) is found to be a significant growth determinant, this raises more questions than answers. In particular, it seems likely to point not toward the importance of enlightened aid, but toward heterogeneity in aid assignment rules. Donors appear to differ substantially in the extent to which their aid responds to

particular recipient characteristics, as Headley (2005) notes. In that case, in growth regressions, disaggregation of aid by donor may end up capturing the effects of nonrandom assignment rather than the intended causal effect. Even panel data studies are not immune to this critique, given that time-varying confounders could influence both aid allocation and growth.

Another approach is to disaggregate aid receipts into different categories, perhaps with different effects on growth. If aid is only partially fungible, then the effect of aid will vary with its type, and Bhagwati (1972) and Papanek (1972) both warned that aggregating different types of aid can be seriously misleading. Clemens et al. (2004) have developed an operational approach, which classifies aid flows into categories expected to have a fast-acting impact on growth, and others which may be more long term or neutral in their effects. Disaggregation is justified if different types of aid work over different horizons, or vary in their effects on productivity. This makes obvious sense from an economic point of view, and helps to provide a more complete picture of dynamic responses. The main drawback is the limitations of the underlying data needed for the classification. Headley (2005) recommends a simpler approach, which is to subtract humanitarian assistance from recorded ODA flows. The disaggregation of aid flows is likely to be an important direction for future work, across the full range of cross-country research on aid effectiveness.

Publication bias

In assessing the literature as a whole, publication bias may be a serious problem. To take an extreme example, imagine that aid has no effect, but journal editors or referees favor empirical results that feature large *t*-statistics. The journals may be filled with the 5% of studies that reflect Type I errors, while the file drawers of researchers are filled with the remaining 95% in which the aid coefficient was insignificant (Rosenthal, 1979). More subtle forms of publication bias can also arise. Perhaps journal reviewers apply higher standards of rigor when assessing papers that suggest a negative effect of aid, in which case the studies viewed in aggregate are too optimistic. Others argue that it is difficult to publish work which indicates positive effects of aid, and that “contrarian” studies may be especially likely to get wide publicity, risking misinterpretation of the balance of evidence.

In other fields, methods for formal meta-analysis have been developed, which can be used to combine multiple findings in a scientific way. But these methods are probably best suited to environments where there are multiple data sets, with randomized trials in medicine as the ideal benchmark. The approach seems less well suited to the context of aid and growth, where the same data are repeatedly revisited, and choices over the specification and estimation methods are central to the variation in results. Moreover, even published studies vary widely in their quality. Some of these factors can be taken into account in meta-analysis, but for now, the systematic investigation

and extension of individual studies, as in Roodman (2007a), seems a better way forward. For a study which does apply meta-analysis methods to foreign aid, see Doucouliagos and Paldam (2008).

End Notes

- *. I am grateful to Patrick Carter, Oliver Morrissey, Paul Mosley, Lant Pritchett, James Rockey, Dani Rodrik, and Adrian Wood for helpful discussions, and especially to Chris Adam, Patrick Carter, Carl-Johan Dalgaard, Oliver Morrissey and Dani Rodrik for detailed written comments on an earlier draft. I am also grateful to participants at the Pre-Conference for the Handbook of Development Economics, including my discussant Arvind Subramanian, and participants at the first IDS/ESRC Development Economics Conference at Sussex. I am solely responsible for the errors and misinterpretations that remain.
- 1. This tendency is criticized by Banerjee (2007) in particular.
- 2. For lively accounts of the history of foreign aid, see Hjertholm and White (2000), Kanbur (2006), and Moss (2007). The early history is covered in more detail in Little and Clifford (1965), Mikesell (1968), and Riddell (2007). For introductions to the issues raised by aid, some of the most useful resources are the articles collected in Easterly (2008) and Tarp (2000), together with the World Bank report *Assessing Aid* (1998).
- 3. This list is not exhaustive. See <http://www.developmentgoals.org/> for the complete list.
- 4. Eaton (1989, Section 7.2.1) discusses the grant element in more detail, including the relationship between the grant element and the net burden on the donor, which generally differ.
- 5. David Roodman has recently constructed a data set on net aid transfers, which nets out interest repayments and the cancellation of past non-ODA loans. Chang et al. (1998) constructed a measure of “effective” development assistance which converted loans into grant equivalents, and subtracted technical assistance.
- 6. The patterns shown in Figure 5 are somewhat sensitive to the criteria used for arriving at a consistent sample of countries observed throughout the period. The figure is based on a sample of 78 countries with populations greater than 250,000 in 1960.
- 7. See Clemens and Moss (2007) for a detailed critique of the 0.7% target.
- 8. The conventional wisdom may apply better in related areas, such as clothing: although consumers will benefit from foreign donations of used clothing, producers will lose. Frazer (2008) finds that donations of clothing to African countries account for a substantial fraction of the decline in domestic clothing production over 1981–2000; this may seem surprising, but used clothing is consistently one of the top 10 US exports to African countries.
- 9. One theoretical complication is introduced by the budgetary process, and the divisions between those planning expenditure, and those implementing it (McGillivray & Morrissey, 2001). This links to a more general issue in empirical studies of fungibility, which is that many of the estimated models are relatively static (McGillivray & Morrissey, 2004). For empirical work, one major limitation is the lack of complete and detailed data on the earmarking of aid for specific purposes.
- 10. Bourguignon et al. (2009) is a rare attempt to quantify the current effect of aid in redistributing income, finding that the effects on international inequality are relatively modest.
- 11. Although usually associated with a representative consumer, the results can also be taken as indicating the outcomes in a model with a limited amount of heterogeneity; see Caselli and Ventura (2000), while Barro and Sala-i-Martin (2004) provide a textbook presentation. Scholl (2009) integrates conditionality within a Ramsey model. An alternative approach emphasizes life-cycle considerations, as in overlapping-generations models; see Eaton (1989) and Dalgaard (2008) for references.

12. The relevant intuition is likely to be similar to that in the quantitative analysis of capital mobility by Gourinchas and Jeanne (2006). They showed that, in the context of growth models with optimizing consumers, opening an economy to private capital inflows leads to only modest gains in welfare. This is because the relevant distortion arises from the gap between foreign and domestic returns to capital, but domestic capital accumulation under autarky will close this gap in any case; and because capital inflows lead to higher investment, they accelerate the decline in the marginal product of capital. Bringing forward the accumulation of capital also brings forward a reduction in its marginal benefit.
13. For example, some economists contest the view that Africa's defining problem is low rates of investment. Collier and Gunning (1999) argue that the relevant constraints for African private sector firms are not primarily the availability of external finance, but a combination of inadequate infrastructure, weak institutions, and risky investment environments.
14. One danger is that other approaches risk being crowded out, especially where they do not fit neatly into the new framework. For example, in a poverty trap model, it may not be possible to reduce the underlying mechanisms to a single constraint that can be targeted by a donor. The view that development is primarily a collective action problem fits into the framework less readily than, say, credit constraints.
15. Empirical tests of the main ideas are not straightforward. Growth economics has often relied on linear regressions in which, implicitly, the growth effects of failings in one dimension can be offset by strengths elsewhere. If growth outcomes are not an additive function of country characteristics, but rely on something more like a fixed-coefficients, Leontief technology, then new empirical approaches are needed. These could be models that are nonlinear in parameters, or that allow multiple regimes. Some ideas along these lines are pursued in Sirimaneetham and Temple (2009). Using the growth accelerations approach, Dovern and Nunnenkamp (2007) find that aid has a small but significant positive effect on the probability of a growth acceleration.
16. Boone (1996) is one of the leading exceptions, but as Hansen and Tarp note, he does find a strong effect of aid on the investment rate in one of his full-sample regressions.
17. For example, where corruption is a problem, higher rates of investment may be one way for state officials to extract bribes and illegal payments, so an association between aid and public investment is not especially informative. See Keefer and Knack (2007).
18. As Deaton (2009) puts it, "to give up the econometric analysis [of aid and growth] is simply to abandon precise statements for loose and unconstrained histories of episodes selected to support the position of the speaker." History mining is not much better than data mining, and sometimes involves fewer constraints.
19. For surveys of aid and growth, see Clemens et al. (2004), Deaton (2009), Easterly (2003), Hansen and Tarp (2000), McGillivray, Feeny, Hermes, and Lensink (2005), Morrissey (2001), Radelet (2008), Roodman (2007a, 2007b), and Tarp (2006). Many of the general issues that arise in the empirical study of growth are reviewed in Temple (1999) and Durlauf, Johnson, and Temple (2005, 2009).
20. This point is made by Mosley and Hudson (2007). Note that there is no reason to expect aid transfers to raise consumption and investment immediately. Some have suggested that aid inflows may initially be reflected in capital account outflows and increased foreign exchange reserves, with effects on spending emerging only later on. See Aiyan and Ruthbah (2008) for analysis and references.
21. It is clear from some theoretical papers that "aid is ineffective" is now dangerously close to being elevated to a stylized fact in the literature. But there is a big difference between a zero that is precisely estimated and one that is not. For example, the confidence intervals in Rajan and Subramanian (2008) are sometimes relatively wide.
22. See also Radelet (2008), who sketches a number of ways in which the literature could evolve.
23. For more discussion of the transfer problem, see Bhagwati et al. (1985), Bhagwati, Panagariya, and Srinivasan (1998, Chapter 16), Eaton (1989), and Suwa-Eisenmann and Verdier (2007). Yano and

Nugent (1999) analyse a model where aid augments the productive potential of either an import-competing sector or a nontraded sector, in the presence of tariffs. This can lead to a nonclassical, small-country transfer paradox even when world prices are unchanged, using logic similar to the analysis of immiserizing growth.

24. The exposition draws on Neary and van Wijnbergen (1986). Their motivation is the study of natural resource discoveries, but the analysis carries over to aid.
25. For more general surveys of the macroeconomic challenges of scaling-up aid, see Gupta, Powell and Yang (2006) and Isard, Lipschitz, Mourmouras, and Yontcheva (2006), and Berg et al. (2007) for case-study evidence.
26. When these effects are present, the likely severity of Dutch disease has to be assessed using general equilibrium models that account simultaneously for productivity effects, movements in the real exchange rate, and distributional consequences, as in Adam and Bevan (2006). The literature on “aid for trade,” including trade facilitation, is also relevant here. See Suwa-Eisenmann and Verdier (2007) for a review.
27. Moreover, it may remain quite difficult to identify the effects of aid separately from the strong performance of the East Asian countries in labor-intensive exports. Their success in these sectors places downward pressure on the world price of labor-intensive exports, causing contraction in those sectors elsewhere, including in those countries that are generally performing poorly and receiving large amounts of aid.
28. Among others, Malik and Temple (2009) confirm that volatility in the terms of trade is a major determinant of volatility in aggregate output.
29. Collier (1999) includes an explicit comparison of the two concepts, pointing out that the well-known poverty trap for individuals, arising from the means-tested withdrawal of individual welfare benefits, does not apply at the level of countries: the implied marginal tax rates are lower by at least an order of magnitude.
30. As Tilly (1975) put it, “war made the state and the state made war.” Besley and Persson (2008a) is a recent analysis of the relation between external war and state capacity.
31. Acemoglu et al. (2004) argue that group-specific taxes and transfers might be especially plausible for Africa, given clear geographic and ethnic distinctions between producer groups, and the observed prevalence of differential taxes and price manipulation through marketing boards (Bates, 1981).
32. Donors can only avoid this problem if they can form a credible commitment to a rule for disbursing aid; in that case, aid can help to enforce cooperation rather than undermining it.
33. One weakness of the cross-country empirical literature is that it cannot always distinguish sharply between different forms of corruption. The ICRG index of corruption used by Mauro (1998) and Svensson (2000a) combines political and bureaucratic forms of corruption: it measures whether highly placed officials are likely to demand special payments, and whether illegal payments are generally expected throughout lower levels of government. Aggregating different forms of corruption seems risky when considering implications for the effectiveness of foreign aid. Treisman (2007) reviews measurement issues in more detail.
34. The problems are analogous to those that appear in Western-style welfare states: some transfers will be made to the “undeserving” or even the outright fraudulent, but this hardly implies that such a scheme should be abandoned altogether.
35. Casamatta and Vellutini (2008) provide a recent analysis of the links between aid and clientelism, in which aid can take policies nearer or further away from optimality.
36. His argument is that a rebellion may take years to prepare, in which case current aid flows are a secondary consideration, while coups occur quickly, and may be more influenced by the rents to sovereignty. In either case, a credible commitment to withhold aid from nondemocratic governments would help to lessen the risks.

- 37. Koeberle, Stavreski, and Walliser (2006) draw together a range of perspectives on budget support.
- 38. He argues that development should be seen mainly in terms of an expansion of freedoms, and these freedoms are symbiotic with economic growth. Others have argued that, to the extent that developing country states become more effective, the checks and balances of democratic institutions may become a vital constraint on the uses to which a newly effective state is turned. These constraints are needed given the well-known point in political philosophy that, with the potential for the state to do good, also comes the power to do great harm. The development of effective states in Europe led to a huge militarization of society and two World Wars; more recently, the genocide in Rwanda in 1994 may have been partly enabled by a bureaucracy that was relatively well developed by African standards (see Robinson, 2002 for references). Easterly, Gatti, and Kurlat (2006) provide some cross-country evidence on whether democratic institutions have been associated with fewer mass killings.
- 39. In this particular instance, it is often argued that the role of aid in sustaining Suharto's grip on power must have been minor, but this observation risks glossing over the compromises involved.
- 40. This sketch is based on Svensson (2000b); see also Kanbur (2000, 2006). Additional formal analyses of conditionality can be found in Azam and Laffont (2003), Cordella and Dell'Arccia (2002), Drazen (1999, 2000, 2002), Epstein and Gang (2009), Federico (2004), and Scholl (2009), among others. For recent overviews of conditionality in practice, see Koeberle (2003), Koeberle, Bedoya, Silarksy, and Verheyen (2005) and Riddell (2007). As Kanbur (2006) notes, an early statement of the desirability of conditionality can be found in Little and Clifford (1965, pp. 93-94).
- 41. An econometric literature has developed to test the effects of IMF programs on economic growth and other outcomes, but the endogeneity of program participation represents a formidable obstacle. See Conway (2006) for analysis and references.
- 42. There is also a literature on the political economy of condition enforcement. For example, Kilby (2009) finds that World Bank loan disbursements were less dependent on macroeconomic performance in recipient countries aligned with the US, based on voting alignment at the United Nations.
- 43. A related idea, again discussed in Svensson (2000b), would be for the donor to enter into contracts with third-parties that will tie the hands of the donor *ex post*. This is an unusual argument for tied aid, but much would depend on implementation of the policy conditions being verifiable in the technical sense (i.e., capable of being confirmed or rejected in a legal setting).
- 44. In principle, donors could commit to withholding aid if their conditions are not met, instead redirecting the aid toward a fund that benefits developing countries as a whole. This option would increase the perceived opportunity cost of allocating aid toward countries that have not met specific conditions, and thereby increase the chance that these conditions will be enforced in practice.
- 45. Such a claim often starts from the belief that the Washington Consensus policies were primarily driven by the interests of donor country firms, especially multinationals, rather than a genuine attempt to promote development. Adherents of this view rarely take the trouble to engage with the academic literature on alternative development strategies, such as import substitution in Latin America and the more *ad hoc* state-led policies that have been common in Africa; see Bruton (1998) for a discussion of import substitution in particular. The economic imperialism view may be one of those ideas that is wrong in most respects, but has just enough truth in it to survive.
- 46. See, for example, Dalgaard and Hansen (2001), Hansen and Tarp (2000), and Roodman (2007a), while Dollar and Levin (2006) cite some additional references. Roodman compares the performance of a variety of interaction terms, and finds that the Hansen and Tarp (2000) interaction between aid and the tropics is the most robust. As some have noted, given the balance of countries in most cross-country samples, this is close to saying that aid has been least effective in Africa.
- 47. See Riddell (2007, 233-235) for a critical discussion of the CPIA.
- 48. A closely related distinction can be found in Rawls (1999, Section 2.2), who argues that governments should not be considered the authors of their own powers, but "peoples." The consequence, as in

- political philosophies based on human rights, is to restrict a government's internal sovereignty. Similar reasoning, based on the priority of the interests of peoples, might allow donors to justify conditionality as morally legitimate in some circumstances.
- 49. See Gould (2005), Riddell (2007, pp. 238–241), and van de Walle (2005) for references and further discussion. Some versions of ownership seem to echo the English football manager Brian Clough, on responding to disagreements with his players (“I'd ask him how he thinks it should be done, have a chat about it for twenty minutes and then decide I was right”).
 - 50. See Birdsall (2008) for a more comprehensive discussion of donor failings than space allows here.
 - 51. More speculatively, as Knack and Rahman (2007) and Djankov, Montalvo, and Reynal-Querol (2009) argue, donor fragmentation seems associated with weak governance in the cross-country data, although the direction of causality must be considered highly uncertain.
 - 52. It is often the World Bank and IMF that bear the brunt of criticism, not all of it equally thoughtful. Some attacks on these institutions fall into the long and dubious tradition that attempts to cast all international development issues in terms of a morality play, ideally one with prominent and emotionally satisfying villains.
 - 53. See Besley and Ghatak (2005) for an analysis of the role of intrinsic motivation in public organizations.
 - 54. In particular, the average treatment effect can be recovered in a way that is model-free (e.g., see Heckman, 1992). Stronger assumptions are needed to identify other quantities, such as median differences. This point is more important than it may first appear: after all, a mean effect of zero is consistent with large positive effects for a few, and losses for many. See Deaton (2009) for more discussion.
 - 55. A different perspective, leading to much the same conclusion, can be found in Scott (1998). He argues that experiments in agriculture have typically shown a lack of interest in existing practices, to an extent that should be regarded as unscientific. He writes “From a narrow scientific view, *nothing* is known until it is proven in a tightly controlled experiment. Knowledge that arrives in any form other than through the techniques and instruments of formal scientific procedure does not deserve to be taken seriously.... traditional practices, codified as they are in practice and in folk sayings, are seen presumptively as not meriting attention, let alone verification.” (pp. 305–306, emphasis in original).
 - 56. The contributions in Banerjee (2007) offer an informative range of viewpoints on the form and role of evaluation.
 - 57. For example, Duflo and Pande (2007) analyze the distributional impact of Indian dams built for irrigation, and show that downstream populations tend to benefit, while those in the vicinity of the dam experience greater variability in agricultural output.
 - 58. For additional discussions of public services in developing countries, see Devarajan and Reinikka (2004) and the 2004 World Development Report.
 - 59. Some NGOs may be more likely than governments to have objectives in line with those of donors, not least when they are staffed by those who care about the outcomes directly (i.e., those with a form of intrinsic motivation). Yet, even if those with intrinsic motivation sort into independent service providers, this sorting is an equilibrium outcome, one that may alter when large sums are allocated to these providers.
 - 60. See Fields (2007) and Gutierrez, Orecchia, Paci, and Serneels (2007) for versions of this long-standing distinction.
 - 61. See Bezemer and Headey (2008) for a more detailed quantitative analysis of the declining importance of agricultural assistance, relative to other sources of aid, and as a topic of World Bank research.
 - 62. It is not solely facetious to say that their contributions have become the Yin and Yang of foreign aid, each incomplete without the other. More seriously, the aid literature is a good example of a research area where diversity in approach and outlook has a functional role. For an excellent account and critique of the different views of Collier, Easterly, and Sachs, see Clemens (2007).

63. Easterly suggests that vouchers could be used for several purposes. This would allow recipients to make their own decisions about priorities, while prompted by donors to increase the consumption of goods such as education, health care, or even safer cooking stoves, where externalities may be significant, or where lack of information or self-control problems can lead to underconsumption.
64. For a more detailed discussion of international public goods, and references, see Kanbur (2006).
65. This might suggest that development ministries or aid agencies need cabinet representation. The consequences of donor policies that are not “joined up” can be seen in the effects of the agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS) on the cost of medicines, and the tendency for developed-country health services to draw on doctors and nurses from the developing world.
66. It is not uncommon to find that researchers have used most of the available moment conditions, but this practice may lead to severe finite sample bias and undermines the power of Sargan-type tests of the overidentifying restrictions. A warning sign is *p*-values from the Sargan or Hansen test that are close to unity. Some papers in the aid literature report such *p*-values, which probably reflect a collapse in the power of the test. Without a meaningful test of the overidentifying restrictions, the reliability and interpretation of the parameter estimates is uncertain at best. See Roodman (2009) for more discussion.
67. Collier and Dollar (2001, 2002) estimate growth regressions in which aid enters in quadratic form, and interacted with a policy indicator; see Hansen and Tarp (2000) for further references, and a discussion of the problems in distinguishing in practice between quadratic terms and alternatives, such as interactions between aid and measures of policy.
68. Given that aid is not randomly assigned, it would be easy to argue that a negatively signed quadratic term is really detecting adverse, but omitted, country characteristics that influence either aid allocations or the marginal effect of aid, or both. Alternatively, countries where aid is sometimes especially high may also be countries where aid flows are especially volatile.

References

- Acemoglu, D., & Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy*, 115, 925–985.
- Acemoglu, D., Robinson, J. A., & Verdier, T. (2004). Kleptocracy and divide-and-rule: A model of personal rule. *Journal of the European Economic Association*, 2(2–3), 162–192.
- Acharya, A., Fuzzo de Lima, A. T., & Moore, M. (2006). Proliferation and fragmentation: Transactions costs and the value of aid. *Journal of Development Studies*, 42(1), 1–21.
- Adam, C. S., & Bevan, D. L. (2006). Aid and the supply side: Public investment, export performance, and Dutch Disease in low-income countries. *World Bank Economic Review*, 20(2), 261–290.
- Adam, C. S., Chambas, G., Guillaumont, P., Guillaumont Jeanneney, S., & Gunning, J. W. (2004). Performance-based conditionality: A European perspective. *World Development*, 32(6), 1059–1070.
- Adam, C. S., & Gunning, J. W. (2002). Redesigning the aid contract: Donors’ use of performance indicators in Uganda. *World Development*, 30(12), 2045–2056.
- Adam, C. S., & O’Connell, S. A. (1999). Aid, taxation and development in sub-Saharan Africa. *Economics and Politics*, 11(3), 225–253.
- Addison, T., & Chowdhury, A. R. (2005). A global lottery and a global Premium Bond. In A. B. Atkinson (Ed.), *New sources of development finance*. Oxford: UNU-WIDER and OUP.
- Agénor, P. R., & Yilmaz, D. (2008). *Aid allocation, growth and welfare with productive public goods*. CGEBCR Discussion Paper No. 95. University of Manchester.
- Aidt, T. (2003). Economic analysis of corruption: A survey. *Economic Journal*, 113, F632–F652.
- Aiyar, S., & Ruthbah, U. (2008). *Where did all the aid go?* IMF Working Paper No. 08/34.

- Alesina, A., & Dollar, D. (2000). Who gives foreign aid to whom and why? *Journal of Economic Growth*, 5(1), 33–63.
- Alesina, A., & Weder, B. (2002). Do corrupt governments receive less foreign aid? *American Economic Review*, 92(4), 1126–1137.
- Almeida, H., & Ferreira, D. (2002). Democracy and the variability of economic performance. *Economics and Politics*, 14(3), 225–257.
- Angeles, L., & Neanidis, K. C. (2009). Aid effectiveness: The role of the local elite. *Journal of Development Economics*, 90(1), 120–134.
- Angrist, J., Bettinger, E., & Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96(3), 847–862.
- Arellano, C., Bulíř, A., Lane, T., & Lipschitz, L. (2009). The dynamic implications of foreign aid and its variability. *Journal of Development Economics*, 88, 87–102.
- Arndt, C. (2000). Technical cooperation. In F. Tarp (Ed.), *Foreign aid and development*. London: Routledge.
- Arndt, C., Jones, S., & Tarp, F. (2007). Aid and development: The Mozambican case. In S. Lahiri (Ed.), *Frontiers of economics and globalization: Theory and practice of foreign aid* (pp. 235–288). Amsterdam: Elsevier.
- Aslaksen, S., & Torvik, R. (2006a). A theory of civil conflict and democracy in rentier states. *Scandinavian Journal of Economics*, 108(4), 571–585.
- Atkinson, A. B. (2005a). The way forward. In A. B. Atkinson (Ed.), *New sources of development finance*. Oxford: UNU-WIDER and OUP.
- Atkinson, A. B. (Ed.), (2005b). *New sources of development finance*. Oxford: UNU-WIDER and OUP.
- Atwood, J. B., McPherson, M. P., & Natsios, A. (2008, November/December). Arrested development: Making foreign aid a more effective tool. *Foreign Affairs*, 87(6), 123–132.
- Azam, J. P., Devarajan, S., & O'Connell, S. A. (1999). *Aid dependence reconsidered*. CSAE Working Paper No. 99–5.
- Azam, J. P., & Laffont, J. J. (2003). Contracting for aid. *Journal of Development Economics*, 70, 25–58.
- Balla, E., & Reinhardt, G. Y. (2008). Giving and receiving foreign aid: Does conflict count? *World Development*, 36, 2566–2585.
- Banerjee, A. (2008). Big answers for big questions: The presumption of growth policy. Manuscript. MIT.
- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007a). Remedyng education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3), 1235–1264.
- Banerjee, A., & Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics*, forthcoming.
- Banerjee, A., Iyer, L., & Somanathan, R. (2007b). Public action for public goods. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of development economics* (vol. 4). Amsterdam: North-Holland.
- Banerjee, A. V. (2007). *Making aid work*. Cambridge, MA: Boston Review/MIT Press.
- Barder, O., & Birdsall, N. (2006, December). *Payments for progress: A hands-off approach to foreign aid*. CGD Working Paper No. 102.
- Barro, R., & Sala-i-Martin, X. (2004). *Economic growth*. Cambridge: MIT Press.
- Bates, R. (1981). *Markets and states in tropical Africa*. Berkeley: University of California Press. (New Edition 2005).
- Bauer, P. T. (1969). Dissent on development. *Scottish Journal of Political Economy*, 75–94.
- Bauer, P. T. (1971). *Dissent on development*. London: Weidenfeld and Nicolson.
- Behrman, J. R., & Sah, R. K. (1984). What role does equity play in the international distribution of development aid? In M. Syrquin, L. Taylor & L. E. Westphal (Eds.), *Economic structure and performance* (pp. 295–315). New York: Academic Press.
- Berg, E. (2001). Why are aid organizations such poor learners? In J. Carlsson & L. Wohlgemuth (Eds.), *Learning in development cooperation*. Stockholm: Almqvist and Wiksell International.
- Berg, A., Aiyar, S., Hussain, M., Roache, S., Mirzoev, T., & Mahone, A. (2007). *The macroeconomics of scaling up aid: Lessons from recent experience*. Occasional Paper No. 253. Washington: IMF.
- Besley, T., & Burgess, R. (2003). Halving global poverty. *Journal of Economic Perspectives*, 17(3), 3–22.

- Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, 95(3), 616–636.
- Besley, T., & Kudamatsu, M. (2006). Health and democracy. *American Economic Review*, 96(2), 313–318.
- Besley, T., & Kudamatsu, M. (2008). Making autocracy work. In E. Helpman (Ed.), *Institutions and economic performance*. Cambridge: Harvard University Press.
- Besley, T., & Persson, T. (2008). Wars and state capacity. *Journal of the European Economic Association*, 6(2–3), 522–530.
- Besley, T., & Persson, T. (2008b). The incidence of civil war: Theory and evidence. Manuscript, London School of Economics.
- Bezemer, D., & Headey, D. (2008). Agriculture, development, and urban bias. *World Development*, 36(8), 1342–1364.
- Bhagwati, J. N. (1970). Alternative estimates of the real cost of aid. In P. Streeten (Ed.), *Unfashionable economics: Essays in honour of Lord Balogh*. London: Weidenfeld and Nicolson.
- Bhagwati, J. N. (1972). Amount and sharing of aid. In C. R. Frank Jr., J. N. Bhagwati, R. d'A. Shaw & H. B. Malmgren (Eds.), *Assisting developing countries: Problems of debts, burden-sharing, jobs, and trade*. New York: Overseas Development Council and Praeger Publishers.
- Bhagwati, J. N., Brecher, R., & Hatta, T. (1985). The generalized theory of transfers and welfare: Exogenous (policy-imposed) and endogenous (transfer-induced) distortions. *Quarterly Journal of Economics*, 100(3), 697–714.
- Bhagwati, J. N., Panagariya, A., & Srinivasan, T. N. (1998). *Lectures on international trade* (2nd ed.). Cambridge, MA: MIT Press.
- Birdsall, N. (2008). Seven deadly sins: Reflections on donor failings. In W. Easterly (Ed.), *Reinventing foreign aid*. Cambridge, MA: MIT Press.
- Birdsall, N., Rodrik, D., & Subramanian, A. (2005). How to help poor countries. *Foreign Affairs*, 84(4), 136–152.
- Björkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment of a community-based monitoring project in Uganda. *Quarterly Journal of Economics*, forthcoming.
- Bobba, M., & Powell, A. (2007). *Aid and growth: Politics matters*. IADB Working Paper No. 601.
- Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review*, 40, 289–329.
- Bourguignon, F., & Leipziger, D. (2006, April). *Aid, growth, and poverty reduction: Toward a new partnership model*. Washington DC: World Bank.
- Bourguignon, F., Levin, V., & Rosenblatt, D. (2009). International redistribution of income. *World Development*, 37(1), 1–10.
- Bourguignon, F., & Sundberg, M. (2007). Aid effectiveness—Opening the black box. *American Economic Review*, 97(2), 316–321.
- Boyce, J. K. (2007). *Public finance, aid and post-conflict recovery*. Department of Economics Working Paper No. 2007–09. University of Massachusetts, Amherst.
- Bräutigam, D. (2004). The people's budget? Politics, participation and pro-poor policy. *Development Policy Review*, 22(6), 653–668.
- Bräutigam, D., Feldstadt, O. H., & Moore, M. (Eds.). (2008). *Taxation and state-building in developing countries*. Cambridge: Cambridge University Press.
- Bräutigam, D., & Knack, S. (2004). Foreign aid, institutions, and governance in sub-Saharan Africa. *Economic Development and Cultural Change*, 52(2), 255–285.
- Bruton, H. J. (1998). A reconsideration of import substitution. *Journal of Economic Literature*, 36(2), 903–936.
- Buiter, W. (2005). Country ownership: A term whose time has gone. In S. Koeberle, H. Bedoya, P. Silarksy & G. Verheyen (Eds.), *Conditionality revisited: Concepts, experiences and lessons*. Washington, DC: World Bank.
- Bulíř, A., & Hamann, A. J. (2003). Aid volatility: An empirical assessment. *IMF Staff Papers*, 50(1), 64–89.
- Bulíř, A., & Hamann, A. J. (2008). Volatility of development aid: From the frying pan into the fire? *World Development*, 36(10), 2048–2066.
- Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *American Economic Review*, 90(4), 847–868.
- Casamatta, G., & Vellutini, C. (2008). Clientelism and aid. *Journal of Development Economics*, 87, 322–332.

- Caselli, F. (2005). Accounting for cross-country income differences. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (pp. 679–741). Amsterdam: North-Holland.
- Caselli, F., & Feyrer, J. (2007). The marginal product of capital. *Quarterly Journal of Economics*, 122(2), 535–568.
- Caselli, F., & Ventura, J. (2000). A representative consumer theory of distribution. *American Economic Review*, 90(4), 909–926.
- Cassen, R., & Associates. (1994). *Does aid work?* (2nd ed.). Oxford: Clarendon Press.
- Celasun, O., & Walliser, J. (2008). Predictability of aid: Do fickle donors undermine aid effectiveness? *Economic Policy*, 55, 545–586.
- Chang, C. C., Fernandez-Arias, E., & Serven, K. (1998). Measuring aid flows: A new approach. Manuscript, World Bank.
- Chatterjee, S., Sakoulis, G., & Turnovsky, S. J. (2003). Unilateral capital transfers, public investment, and economic growth. *European Economic Review*, 47(6), 1077–1103.
- Chatterjee, S., & Turnovsky, S. J. (2004). Substitutability of capital, investment costs, and foreign aid. In S. Dowrick, R. Pitchford & S. J. Turnovsky (Eds.), *Economic growth and macroeconomic dynamics: Recent developments in economic theory*. Cambridge: Cambridge University Press.
- Chatterjee, S., & Turnovsky, S. J. (2005). Financing public investment through foreign aid: Consequences for economic growth and welfare. *Review of International Economics*, 13(1), 20–44.
- Chen, S., Mu, R., & Ravallion, M. (2009). Are there lasting impacts of aid to poor areas? *Journal of Public Economics*, forthcoming.
- Chhotray, V., & Hulme, D. (2009). Contrasing visions for aid and governance in the 21st century: The White House Millennium Challenge Account and DFID's Drivers of Change. *World Development*, forthcoming.
- Clemens, M. A. (2007). Smart Samaritans: Is there a third way in the development debate? Review essay. *Foreign Affairs*, 86(5), 132–140.
- Clemens, M. A., Kenny, C. J., & Moss, T. J. (2007). The trouble with the MDGs: Confronting expectations of aid and development success. *World Development*, 35(5), 735–751.
- Clemens, M. A., & Moss, T. J. (2007). The ghost of 0.7 per cent: Origins and relevance of the international aid target. *International Journal of Development Issues*, 6(1), 3–25.
- Clemens, M. A., Radelet, S., & Bhavnani, R. (2004, November). *Counting chickens when they hatch: The short-term effect of aid on growth*. CGD. Working Paper No. 44, revised.
- Cogneau, D., & Naudet, J. D. (2007). Who deserves aid? Equality of opportunity, international aid, and poverty reduction. *World Development*, 35(1), 104–120.
- Cohen, D. (2006). *Globalization and its enemies*. Cambridge, MA: MIT Press.
- Cohen, D., & Soto, M. (2002). *Why are poor countries poor? A message of hope which involves the resolution of a Becker/Lucas Paradox*. CEPR Discussion Paper No. 3528.
- Cohen, J., & Dupas, P. (2007). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment in Kenya. Manuscript, UCLA.
- Collier, P. (1999). Aid ‘dependency’: A critique. *Journal of African Economies*, 8(4), 528–545.
- Collier, P. (2006a). Is aid oil? An analysis of whether Africa can absorb more aid. *World Development*, 34(9), 1482–1497.
- Collier, P. (2006b). African growth: Why a ‘Big Push?’ *Journal of African Economies*, 15(S2), 188–211.
- Collier, P. (2007). *The bottom billion*. Oxford: Oxford University Press.
- Collier, P., Conway, G., & Venables, A. (2008). Climate change and Africa. *Oxford Review of Economic Policy*, 24(2), 337–353.
- Collier, P., & Dollar, D. (2001). Can the world cut poverty in half? How policy reform and effective aid can meet international development goals. *World Development*, 29(11), 1787–1802.
- Collier, P., & Dollar, D. (2002). Aid allocation and poverty reduction. *European Economic Review*, 46, 1475–1500.
- Collier, P., & Dollar, D. (2004). Development effectiveness: What have we learnt? *Economic Journal*, 114(496), F244–F271.
- Collier, P., & Goderis, B. (2008). Commodity prices, growth, and the natural resource curse: Reconciling a conundrum. Manuscript, University of Oxford.

- Collier, P., Guillaumont, P., Guillaumont, S., & Gunning, J. W. (1997). Redesigning conditionality. *World Development*, 25(9), 1399–1407.
- Collier, P., & Gunning, J. W. (1999). Explaining African economic performance. *Journal of Economic Literature*, 37(1), 64–111.
- Collier, P., & Hoeffer, A. (2004). Aid, policy and growth in post-conflict societies. *European Economic Review*, 48, 1124–1145.
- Collier, P., & Hoeffer, A. (2007). Unintended consequences: Does aid promote arms races? *Oxford Bulletin of Economics and Statistics*, 69(1), 1–27.
- Collier, P., & Rohner, D. (2008). Democracy, development, and conflict. *Journal of the European Economic Association*, 6(2–3), 531–540.
- Collier, P., & Sambanis, N. (Eds.). (2005). *Understanding civil war (Volume 1: Africa): Evidence and analysis*. Washington DC: World Bank.
- Collier, P., & Vicente, P. C. (2008). Votes and violence: Evidence from a field experiment in Nigeria. Manuscript, University of Oxford.
- Commission for Africa (2005). *Our common interest*. London: Penguin Books.
- Conway, P. (2006). Endogenous IMF conditionality: Theoretical and empirical implications. In G. Ranis, J. Vreeland & S. Kosack (Eds.), *Globalization and the Nation State*. London: Routledge.
- Cordella, T., & Dell'Ariccia, G. (2002). Limits of conditionality in poverty reduction programs. *IMF Staff Papers*, 49, 68–86.
- Cordella, T., & Dell'Ariccia, G. (2007). Budget support versus project aid: A theoretical appraisal. *Economic Journal*, 117, 1260–1279.
- Corden, M. W., & Neary, P. J. (1982). Booming sector and de-industrialisation in a small open economy. *Economic Journal*, 92(368), 825–848.
- Coviello, D., & Islam, R. (2006). *Does aid help improve economic institutions?*. World Bank Policy Research Working Paper No. 3990.
- Dalgaard, C. J. (2008). Donor policy rules and aid effectiveness. *Journal of Economic Dynamics and Control*, 32, 1895–1920.
- Dalgaard, C. J., & Hansen, H. (2001). On aid, growth and good policies. *Journal of Development Studies*, 37 (6), 17–41.
- Dalgaard, C. J., Hansen, H., & Tarp, F. (2004). On the empirics of foreign aid and growth. *Economic Journal*, 114, F191–F216.
- Dalgaard, C. J., & Olsson, O. (2008). Windfall gains, political economy and economic development. *Journal of African Economies*, 17(AERC supplement), i72–i109.
- Deaton, A. (2008). Maximum prophet. *The Lancet*, 372, 1535–1536.
- Deaton, A. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. Manuscript, Princeton University.
- de Ferranti, D. (2006). The World Bank and the middle-income countries. In N. Birdsall (Ed.), *Rescuing the World Bank*. Washington: CGD.
- DeLong, J. B. (1997). Cross-country variations in national economic growth rates: The role of technology. In J. Fuhrer & J. S. Little (Eds.), *Technology and growth*. Boston: Federal Reserve Bank of Boston.
- Depetrini Chauvin, N., & Kraay, A. (2007). Who gets debt relief? *Journal of the European Economic Association*, 5(2–3), 333–342.
- de Ree, J., & Nillesen, E. (2009). Aiding violence or peace? The impact of foreign aid on the risk of civil conflict in sub-Saharan Africa. *Journal of Development Economics*, 88, 301–313.
- Devarajan, S., Dollar, D. R., & Holmgren, T. (Eds.), (et al., 2001). *Aid and reform in Africa: Lessons from ten case studies*. Washington DC: World Bank.
- Devarajan, S., & Reinikka, R. (2004). Making services work for poor people. *Journal of African Economies*, 13(AERC supplement), i142–i166.
- Djankov, S., Montalvo, J. G., & Reynal-Querol, M. (2008). The curse of aid. *Journal of Economic Growth*, 13, 169–194.
- Djankov, S., Montalvo, J. G., & Reynal-Querol, M. (2009). Aid with multiple personalities. *Journal of Comparative Economics*, forthcoming.

- Dollar, D., & Levin, V. (2006). The increasing selectivity of foreign aid, 1984–2003. *World Development*, 34(12), 2034–2046.
- Dollar, D., & Svensson, J. (2000). What explains the success or failure of structural adjustment programmes? *Economic Journal*, 110, 894–917.
- Doucouliagos, H., & Paldam, M. (2008). Aid effectiveness on growth: A meta study. *European Journal of Political Economy*, 24(1), 1–24.
- Dovern, J., & Nunnenkamp, P. (2007). Aid and growth accelerations: An alternative approach to assessing the effectiveness of aid. *Kyklos*, 60(3), 359–383.
- Drazen, A. (1999). What is gained by selectively withholding foreign aid? Manuscript, University of Maryland.
- Drazen, A. (2000). Political economy in macroeconomics. Princeton: Princeton University Press.
- Drazen, A. (2002). Conditionality and ownership in IMF lending: A political economy approach. *IMF Staff Papers*, 49, 36–67.
- Dreher, A., Nunnenkamp, P., & Thiele, R. (2008). Does aid for education educate children? Evidence from panel data. *World Bank Economic Review*, 22, 291–314.
- Drèze, J., & Sen, A. (1995). India: Economic development and social opportunity. Delhi: Oxford University Press.
- Drèze, J., & Sen, A. (Eds.). (1996). *Indian development: Selected regional perspectives*. Delhi: Oxford University Press.
- Duflo, E., Dupas, P., & Kremer, M. (2008a). Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya. Manuscript, MIT.
- Duflo, E., Glennerster, R., & Kremer, M. (2008b). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: North-Holland.
- Duflo, E., Hanna, R., & Ryan, S. (2007). Monitoring works: Getting teachers to come to school. Manuscript, MIT.
- Duflo, E., & Kremer, M. (2005). Use of randomization in the evaluation of development effectiveness. In G. Pitman, O. Feinstein & G. Ingram (Eds.), *Evaluating development effectiveness*. New Brunswick, NJ: Transaction Publishers.
- Duflo, E., Kremer, M., & Robinson, J. (2008c). How high are rates of return to fertilizer? Evidence from field experiments in Kenya. *American Economic Review*, 98(2), 482–488.
- Duflo, E., & Pande, R. (2007). Dams. *Quarterly Journal of Economics*, 122(2), 601–646.
- Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. (2005). Growth econometrics. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1A) (pp. 555–677). Amsterdam: North-Holland.
- Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. (2009). The methods of growth econometrics. In T. C. Mills & K. Patterson (Eds.), *Palgrave handbook of econometrics, Vol. 2: Applied econometrics*. London: Palgrave Macmillan.
- Easterly, W. (2001). *The elusive quest for growth*. Cambridge: MIT Press.
- Easterly, W. (2003). Can foreign aid buy growth? *Journal of Economic Perspectives*, 17(3), 23–48.
- Easterly, W. (2005). What did structural adjustment adjust? The association of policies and growth with repeated IMF and World Bank adjustment loans. *Journal of Development Economics*, 76(1), 1–22.
- Easterly, W. (2006a). The Big Push Deja Vu: A review of Jeffrey Sachs's The End of Poverty: Economic possibilities for our time. *Journal of Economic Literature*, 44(1), 96–105.
- Easterly, W. (2006b). *The White Man's Burden: Why the West's efforts to aid the rest have done so much ill and so little good*. Oxford: Oxford University Press.
- Easterly, W. (2006c). Reliving the 1950s: The big push, poverty traps, and takeoffs in economic development. *Journal of Economic Growth*, 11(4), 289–318.
- Easterly, W. (2007a). Are aid agencies improving? *Economic Policy*, 52, 633–678.
- Easterly, W. (2007b). Was development assistance a mistake? *American Economic Review*, 97(2), 328–332.
- Easterly, W. (Ed.), (2008). *Reinventing foreign aid*. Cambridge, MA: MIT Press.
- Easterly, W. (2009a). How the Millennium Development Goals are unfair to Africa. *World Development*, 37(1), 26–35.
- Easterly, W. (2009b). Can the West save Africa? *Journal of Economic Literature*, 47(2), 373–447.

- Easterly, W., Gatti, R., & Kurlat, S. (2006). Development, democracy, and mass killings. *Journal of Economic Growth*, 11(2), 129–156.
- Easterly, W., Levine, R., & Roodman, D. (2004). Aid, policies, and growth: Comment. *American Economic Review*, 94(3), 774–780.
- Easterly, W., & Pfutze, T. (2008). Where does the money go? Best and worst practices in foreign aid. *Journal of Economic Perspectives*, 22(2), 29–52.
- Eaton, J. (1989). Foreign public capital flows. In H. Chenery & T. N. Srinivasan (Eds.), *Handbook of development economics* (Vol. II). Amsterdam: North-Holland.
- Eifert, B., & Gelb, A. (2008). Reforming aid: Toward more predictable, performance-based financing for development. *World Development*, 36(10), 2067–2081.
- Epstein, G. S., & Gang, I. N. (2009). Good governance and good aid allocation. *Journal of Development Economics*, forthcoming.
- Epstein, H. (2007). *The invisible cure: Africa, the West, and the fight against AIDS*. London: Viking.
- Federico, G. (2004). Aid and policies under weak donor conditionality. *Journal of African Economies*, 13(3), 395–416.
- Feyzioğlu, T., Swaroop, V., & Zhu, M. (1998). A panel data analysis of the fungibility of foreign aid. *World Bank Economic Review*, 12(1), 29–58.
- Fields, G. S. (2007). *Labor market policy in developing countries: A selective review of the literature and needs for the future*. World Bank Policy Research Working Paper No. 4362.
- Filmer, D., Hammer, J. S., & Pritchett, L. (2000). Weak links in the chain: A diagnosis of health policy in poor countries. *World Bank Research Observer*, 15, 199–224.
- Fisman, R., & Miguel, E. (2008). *Economic gangsters*. Princeton: Princeton University Press.
- Frazer, G. (2008). Used-clothing donations and apparel production in Africa. *Economic Journal*, 118, 1764–1784.
- Frazer, G., & Van Bieseboeck, J. (2009). Trade growth under the African Growth and Opportunity Act. *Review of Economics and Statistics*, forthcoming.
- Gandhi, J. (2008). *Political institutions under dictatorship*. Cambridge: Cambridge University Press.
- Gawande, K., Krishna, P., & Olarreaga, M. (2009). What governments maximise and why: The view from trade. *International Organization*, forthcoming.
- Gelb, A. H., & Associates. (1988). *Oil windfalls: Blessing or curse?*. Oxford: Oxford University Press.
- Gomanee, K., Morrissey, O., Mosley, P., & Verschoor, A. (2005). Aid, government expenditure, and aggregate welfare. *World Development*, 33(3), 355–370.
- Gould, J. (Ed.), (2005). *The new conditionality: The politics of poverty reduction strategies*. London: Zed Books.
- Gourinchas, P. O., & Jeanne, O. (2006). The elusive gains from international financial integration. *Review of Economic Studies*, 73(3), 715–741.
- Graham, B. S., & Temple, J. R. W. (2006). Rich nations, poor nations: How much can multiple equilibria explain? *Journal of Economic Growth*, 11(1), 5–41.
- Gunter, B. G., Rahman, J., & Wodon, Q. (2008). Robbing Peter to pay Paul? Understanding who pays for debt relief. *World Development*, 36(1), 1–16.
- Gupta, K. L. (1975). Foreign capital inflows, dependency burden, and saving rates in developing countries: A simultaneous equation model. *Kyklos*, 28(2), 358–374.
- Gupta, S., Clements, B. J., Pivovarsky, A., & Tiengson, E. R. (2004). Foreign aid and revenue response does the composition of aid matter? In S. Gupta, G. Inchauste & B. Clements (Eds.), *Helping countries develop: The role of fiscal policy*. Washington: IMF.
- Gupta, S., Pattillo, C., & Wagh, S. (2006). Are donor countries giving more or less aid? *Review of Development Economics*, 10(3), 535–552.
- Gupta, S., Powell, R., & Yang, Y. (2006). *Macroeconomic challenges of scaling up aid to Africa*. Washington: IMF.
- Gutiérrez, C., Orechchia, C., Paci, P., & Serneels, P. (2007). *Does employment generation really matter for poverty reduction?* World Bank Policy Research Working Paper No. 4432.
- Hansen, H., & Tarp, F. (2000). Aid effectiveness disputed. *Journal of International Development*, 12, 375–398.
- Hansen, H., & Tarp, F. (2001). Aid and growth regressions. *Journal of Development Economics*, 64, 547–570.

- Hanushek, E. A., & Wößmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668.
- Hausmann, R., Klinger, B., & Wagner, R. (2008, September). *Doing growth diagnostics in practice: A ‘Mind-book’*. CID Working Paper No. 177.
- Hausmann, R., Pritchett, L., & Rodrik, D. (2005). Growth accelerations. *Journal of Economic Growth*, 10, 303–329.
- Hausmann, R., Rodrik, D., & Velasco, A. (2008). Growth diagnostics. In J. Stiglitz & N. Serra (Eds.), *The Washington Consensus reconsidered: Towards a new global governance*. New York: Oxford University Press.
- Headey, D. (2005). *Foreign aid and foreign policy: How donors undermine the effectiveness of overseas development assistance*. Working Paper No. 05/2005. CEPA, University of Queensland.
- Headey, D. (2008). Geopolitics and the effect of foreign aid on economic growth: 1970–2001. *Journal of International Development*, 20(2), 161–180.
- Heckman, J. J. (1992). Randomization and social policy evaluation. In C. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs* (pp. 201–230). Harvard: Harvard University Press.
- Hepburn, C., & Stern, N. (2008). A new global deal on climate change. *Oxford Review of Economic Policy*, 24(2), 259–279.
- Herbst, J. (2000). *States and power in Africa*. Princeton: Princeton University Press.
- Hirschman, A. O. (1958). *The strategy of economic development*. Newhaven: Yale University Press.
- Hjertholm, P., & White, H. (2000). Foreign aid in historical perspective: Background and trends. In F. Tarp (Ed.), *Foreign aid and development*. London: Routledge.
- Hudson, J., & Mosley, P. (2008). Aid volatility, policy and development. *World Development*, 36(10), 2082–2102.
- Imbs, J., & Ranciere, R. (2007). The overhang hangover. Manuscript, HEC Lausanne.
- Isard, P., Lipschitz, L., Mourmouras, A., & Yontcheva, B. (2006). *The macroeconomic management of foreign aid: Opportunities and pitfalls*. Washington: IMF.
- Isham, J., & Kaufmann, D. (1999). The forgotten rationale for policy reform: The productivity of investment projects. *Quarterly Journal of Economics*, 114(1), 149–184.
- Isham, J., Kaufmann, D., & Pritchett, L. H. (1997). Civil liberties, democracy, and the performance of government projects. *World Bank Economic Review*, 11(2), 219–242.
- Jayachandran, S., & Kremer, M. (2006). Odious debt. *American Economic Review*, 96(1), 82–92.
- Jones, B., & Olken, B. (2005). Do leaders matter? National leadership and growth since World War II. *Quarterly Journal of Economics*, 120(3), 835–864.
- Jones, B., & Olken, B. (2009). Hit or miss? The effect of assassinations on institutions and war. *American Economic Journal: Macroeconomics*, forthcoming.
- Kaldor, N. (1963, January). Will underdeveloped countries learn to tax? *Foreign Affairs*.
- Kanbur, R. (2000). Aid, conditionality and debt in Africa. In F. Tarp (Ed.), *Foreign aid and development*. London: Routledge.
- Kanbur, R. (2006). The economics of international aid. In S. C. Kolm & J. M. Ythier (Eds.), *Handbook of the economics of giving, altruism, and reciprocity* (Vol. 2). Amsterdam: North-Holland.
- Keefer, P. (2004). What does political economy tell us about economic development—and vice versa? *Annual Review of Political Science*, 7, 247–272.
- Keefer, P., & Khemani, S. (2005). Democracy, public expenditures, and the poor: Understanding political incentives for providing public services. *World Bank Research Observer*, 20, 1–27.
- Keefer, P., & Knack, S. (2007). Boondoggles, rent-seeking and political checks and balances: Public investment under unaccountable governments. *Review of Economics and Statistics*, 89(3), 566–572.
- Khan, M. S., & Sharma, S. (2003). IMF conditionality and country ownership of adjustment programs. *World Bank Research Observer*, 18(2), 227–248.
- Kilby, C. (2009). The political economy of conditionality: An empirical analysis of loan disbursements. *Journal of Development Economics*, forthcoming.
- Kimbrough, K. (1986). Foreign aid and optimal fiscal policy. *Canadian Journal of Economics*, 19(1), 35–61.
- Knack, S., & Rahman, A. (2007). Donor fragmentation and bureaucratic quality in aid recipients. *Journal of Development Economics*, 83, 176–197.

- Koeberle, S., Bedoya, H., Silarksy, P., & Verheyen, G. (Eds.). (2005). *Conditionality revisited: Concepts, experiences and lessons*. Washington, DC: World Bank.
- Koeberle, S., Stavreski, Z., & Walliser, J. (2006). *Budget support as more effective aid?*. Washington, DC: World Bank.
- Koeberle, S. G. (2003). Should policy-based lending still involve conditionality? *World Bank Research Observer*, 19(2), 249–273.
- Kourtellos, A., Tan, C. M., & Zhang, X. (2007). Is the relationship between aid and growth nonlinear? *Journal of Macroeconomics*, 29, 515–540.
- Kraay, A., & Raddatz, C. (2007). Poverty traps, aid, and growth. *Journal of Development Economics*, 82, 315–347.
- Kremer, M. (2008). Making vaccines pay. In W. Easterly (Ed.), *Reinventing foreign aid*. Cambridge: MIT Press.
- Kremer, M., & Glennerster, R. (2004). *Strong medicine*. Princeton: Princeton University Press.
- Kremer, M., & Holla, A. (2008). Pricing and access: Lessons from randomized evaluations in education and health. Manuscript, Harvard.
- Kremer, M., & Miguel, E. (2007). The illusion of sustainability. *Quarterly Journal of Economics*, 122, 1007–1065.
- Lahiri, S., & Raimondos-Møller, P. (2004). Donor strategy under the fungibility of foreign aid. *Economics and Politics*, 16(2), 213–231.
- Lensink, R., & Morrissey, O. (2000). Aid instability as a measure of uncertainty and the positive impact of aid on growth. *Journal of Development Studies*, 36(3), 30–48.
- Levine, R., & Savedoff, W. D. (2006). The evaluation agenda. In N. Birdsall (Ed.), *Rescuing the World Bank*. Washington: CGD.
- Levinsohn, J., & McMillan, M. (2006). Does food aid harm the poor? Household evidence from Ethiopia. In A. Harrison (Ed.), *Globalization and poverty*. The University of Chicago Press for the NBER.
- Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, 22(2), 139–191.
- Little, I. M. D., & Clifford, J. M. (1965). *International aid*. London: George Allen and Unwin.
- Little, I. M. D., Cooper, R. N., Corden, W. M., & Rajapatirana, S. (1993). *Boom, crisis, and adjustment: The macroeconomic performance of developing countries*. New York: OUP.
- López, J. H., Molina, L., & Bussolo, M. (2008). Remittances, the real exchange rate, and the Dutch Disease phenomenon. In P. Fajnzylber & J. H. López (Eds.), *Remittances and development: Lessons from Latin America*. Washington: World Bank.
- Malik, A., & Temple, J. R. W. (2009). The geography of output volatility. *Journal of Development Economics*, 90(2), 163–178.
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2), 407–437.
- Martins, B., Mumment, U., Murrell, P., & Seabright, P. (2002). *The institutional economics of foreign aid*. Cambridge: Cambridge University Press.
- Masud, N., & Yontcheva, B. (2005). *Does foreign aid reduce poverty? Empirical evidence from nongovernmental and bilateral aid*. IMF Working Paper No. 05/100.
- Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics*, 69, 263–279.
- McGillivray, M., Feeny, S., Hermes, N., & Lensink, R. (2005). *It works; it doesn't; it can, but that depends*. UNU-WIDER Research Paper No. 2005/54.
- McGillivray, M., & Morrissey, O. (2001). Aid illusion and public sector fiscal behaviour. *Journal of Development Studies*, 37(6), 118–136.
- McGillivray, M., & Morrissey, O. (2004). Fiscal effects of aid. In T. Addison & A. Roe (Eds.), *Fiscal policy for development* (pp. 72–96). Palgrave/WIDER.
- Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159–217.
- Mikesell, R. F. (1968). *The economics of foreign aid*. New Brunswick: Aldine.

- Minoiu, C., & Reddy, S. G. (2007). *Development aid and economic growth: A positive long-run relation*. Working Paper, Barnard College.
- Mishra, P., & Newhouse, D. (2007). *Health aid and infant mortality*. IMF Working Paper No. 07/100.
- Moore, M. (2004). Revenues, state formation, and the quality of governance in developing countries. *International Political Science Review*, 25(3), 297–319.
- Morrissey, O. (2001). Does aid increase growth? *Progress in Development Studies*, 1(1), 37–50.
- Morrissey, O. (2005). Alternatives to conditionality in policy-based lending. In S. Koerberle, H. Bedoya, P. Silarksy & G. Verheyen (Eds.), *Conditionality revisited: Concepts, experiences and lessons*. Washington, DC: World Bank.
- Morss, E. R. (1984). Institutional destruction resulting from donor and project proliferation in sub-Saharan African countries. *World Development*, 12(4), 465–470.
- Mosley, P. (1980). Aid, savings and growth revisited. *Oxford Bulletin of Economics and Statistics*, 42(2), 79–95.
- Mosley, P. (1986). Aid effectiveness: The macro-micro paradox. *Institute of Development Studies Bulletin*, 17, 22–28.
- Mosley, P. (1987). *Overseas aid: Its defence and reform*. Brighton: Wheatsheaf Books.
- Mosley, P., Harrigan, J., & Toye, J. (1995). *Aid and power: The World Bank and policy-based lending* (2nd ed., Vol. 1). London: Routledge.
- Mosley, P., & Hudson, J. (2007). Aid effectiveness: Literature review. Manuscript, University of Sheffield.
- Mosley, P., Hudson, J., & Verchoor, A. (2004). Aid, poverty reduction and the ‘new conditionality’. *Economic Journal*, 114, F217–F243.
- Moss, T. (2007). *African development: Making sense of the issues and actors*. Boulder, CO: Lynne Rienner Publishers.
- Moss, T., Pettersson, G., & van de Walle, N. (2008). An aid-institutions paradox? A review essay on aid dependency and state building in sub-Saharan Africa. In W. Easterly (Ed.), *Reinventing foreign aid*. Cambridge, MA: MIT Press.
- Neary, J. P., & van Wijnbergen, S. J. G. (1986). Natural resources and the macroeconomy: A theoretical framework. In J. P. Neary & S. van Wijnbergen (Eds.), *Natural resources and the macroeconomy*. Oxford: Blackwell.
- Nelson, R. R. (1956). A theory of the low-level equilibrium trap in underdeveloped economies. *American Economic Review*, 46(5), 894–908.
- Obstfeld, M. (1999). Foreign resource inflows, saving, and growth. In K. Schmidt-Hebbel & L. Serven (Eds.), *The economics of saving and growth*. Cambridge: CUP.
- Oechslin, M. (2006, November). *Foreign aid, political instability, and economic growth*. Working Paper No. 310. University of Zurich, Institute for Empirical Research in Economics.
- Olken, B. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2), 200–249.
- Overland, J., Simons, K., & Spagat, M. (2005). Political instability and growth in dictatorships. *Public Choice*, 125(3), 445–470.
- Pack, H., & Pack, J. R. (1990). Is foreign aid fungible? The case of Indonesia. *Economic Journal*, 100, 188–194.
- Pack, H., & Pack, J. R. (1993). Foreign aid and the question of fungibility. *Review of Economics and Statistics*, 75(2), 258–265.
- Pallage, S., & Robe, M. A. (2001). Foreign aid and the business cycle. *Review of International Economics*, 9(4), 641–672.
- Pallage, S., Robe, M. A., & Bérubé, C. (2006). The potential of foreign aid as insurance. *IMF Staff Papers*, 53(3), 453–672.
- Pande, R. (2008). Understanding political corruption in low income countries. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: North-Holland.
- Papanek, G. F. (1972). The effect of aid and other resource transfers on savings and growth in less developed countries. *Economic Journal*, 82(327), 934–950.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

- Prasad, E. S., Rajan, R. G., & Subramanian, A. (2007). Foreign capital and economic growth. *Brookings Papers on Economic Activity*, 1, 153–230.
- Pritchett, L. (2000). Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains. *World Bank Economic Review*, 14(2), 221–250.
- Pritchett, L., & Woolcock, M. (2008). Solutions when the solution is the problem: Arraying the disarray in development. In W. Easterly (Ed.), *Reinventing foreign aid*. Cambridge: MIT Press.
- Radelet, S. (2006). The role of the Bank in low-income countries. In N. Birdsall (Ed.), *Rescuing the World Bank*. Washington: CGD.
- Radelet, S. (2008). Foreign aid. In A. Dutt & J. Ros (Eds.), *International handbook of development economics*. London: Edward Elgar.
- Rajan, R., & Subramanian, A. (2006). *Aid, Dutch Disease, and manufacturing growth*. Mimeo. IMF.
- Rajan, R., & Subramanian, A. (2007). Does aid affect governance? *American Economic Review*, 97(2), 322–327.
- Rajan, R., & Subramanian, A. (2008). Aid and growth: What does the cross-country evidence really show? *Review of Economics and Statistics*, 90(4), 643–665.
- Ravallion, M. (2005). On measuring aggregate “social efficiency” *Economic Development and Cultural Change*, 53(2), 273–292.
- Rawls, J. (1999). *The law of peoples*. Harvard: Harvard University Press.
- Reinikka, R., & Svensson, J. (2004). Local capture: Evidence from a central government transfer program in Uganda. *Quarterly Journal of Economics*, 119(2), 679–705.
- Reinikka, R., & Svensson, J. (2005). Fighting corruption to improve schooling: Evidence from a newspaper campaign in Uganda. *Journal of the European Economic Association*, 3(2–3), 1–9.
- Reinikka, R., & Svensson, J. (2007). *The returns from reducing corruption: Evidence from education in Uganda*. CEPR Discussion Paper No. 6363.
- Remmer, K. L. (2004). Does foreign aid promote the expansion of government? *American Journal of Political Science*, 48(1), 77–92.
- Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2), 234–250.
- Riddell, R. (2007). *Does foreign aid really work?*. Oxford: Oxford University Press.
- Robinson, J. A. (2002). States and power in Africa by Jeffrey I. Herbst: A review essay. *Journal of Economic Literature*, 40, 510–519.
- Rodrik, D. (1989). Promises, promises: Credible policy reform via signalling. *Economic Journal*, 99, 756–772.
- Rodrik, D. (1996). Why is there multilateral lending? In M. Bruno & B. Pleskovic (Eds.), *Annual World Bank conference on development economics 1995*. World Bank.
- Rodrik, D. (1997). Democracy and economic performance. Manuscript, Harvard University.
- Rodrik, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. *Journal of Economic Growth*, 4(4), 385–412.
- Rodrik, D. (2005). Growth strategies. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth*. Amsterdam: North-Holland.
- Rodrik, D. (2007a). *One economics, many recipes*. Princeton: Princeton University Press.
- Rodrik, D. (2007b). Goodbye Washington Consensus, hello Washington confusion. *Journal of Economic Literature*, 44(4), 973–987.
- Rodrik, D. (2008a). The new development economics: We shall experiment, but how shall we learn? In J. Cohen & W. Easterly (Eds.), *What works in development? Thinking big vs. thinking small*. Washington, DC: Brookings Institution Press.
- Rodrik, D. (2008b). The real exchange rate and economic growth. *Brookings Papers on Economic Activity*, forthcoming.
- Roodman, D. (2007a). The anarchy of numbers: Aid, development, and cross-country empirics. *World Bank Economic Review*, 21(2), 255–277.
- Roodman, D. (2007b, December). *Macro aid effectiveness research: A guide for the perplexed*. Working Paper No. 135. Center for Global Development.

- Roodman, D. (2008). *The commitment to development index: 2008 edition*. Manuscript, Center for Global Development.
- Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, forthcoming.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Sachs, J. D. (2005). *The end of poverty*. London: Penguin.
- Sachs, J. D., McArthur, J. W., Schmidt-Traub, G., Kruk, M., Bahadur, C., Faye, M., et al. (2004). Ending Africa's poverty trap. *Brookings Papers on Economic Activity*, 35, 117–240.
- Sala-i-Martin, X., Doppelhofer, G., & Miller, R. I. (2004). Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach. *American Economic Review*, 94(4), 813–835.
- Scholl, A. (2009). Aid effectiveness and limited enforceable conditionality. *Review of Economic Dynamics*, forthcoming.
- Scott, J. C. (1998). *Seeing like a state*. New Haven: Yale University Press.
- Sen, A. (1981). Public action and the quality of life in developing countries. *Oxford Bulletin of Economics and Statistics*, 43(4), 287–319.
- Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.
- Sirimaneetham, V., & Temple, J. R. W. (2009). Macroeconomic stability and the distribution of growth rates. *World Bank Economic Review*, forthcoming.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1), 65–94.
- Stiglitz, J. (2002). *Globalization and its discontents*. London: Penguin Books.
- Suwa-Eisenmann, A., & Verdier, T. (2007). Aid and trade. *Oxford Review of Economic Policy*, 23(3), 481–507.
- Svensson, J. (1999). Aid, growth and democracy. *Economics and Politics*, 11(3), 275–297.
- Svensson, J. (2000a). Foreign aid and rent-seeking. *Journal of International Economics*, 51, 437–461.
- Svensson, J. (2000b). When is foreign aid policy credible? Aid dependence and conditionality. *Journal of Development Economics*, 61(1), 61–84.
- Svensson, J. (2003). Why conditional aid does not work and what can be done about it? *Journal of Development Economics*, 70(2), 381–402.
- Svensson, J. (2006). The institutional economics of foreign aid. *Swedish Economic Policy Review*, 13, 115–137.
- Tarp, F. (2000). *Foreign aid and development*. London: Routledge.
- Tarp, F. (2001). Aid and reform in Africa: Review of World Bank volume by S. Devarajan, D. R. Dollar and T. Holmgren. *Journal of African Economies*, 10(4), 341–353.
- Tarp, F. (2006). *Aid and development*. Discussion Paper No. 06–12. Department of Economics, University of Copenhagen.
- Tavares, J. (2003). Does foreign aid corrupt? *Economics Letters*, 79, 99–106.
- Temple, J. R. W. (1999). The new growth evidence. *Journal of Economic Literature*, 37(1), 112–156.
- Temple, J. R. W. (2003). Growing into trouble: Indonesia after 1966. In D. Rodrik (Ed.), *In search of prosperity: Analytic narratives on economic growth*. Princeton: Princeton University Press.
- Therkildsen, O. (2002). Keeping the state accountable: Is aid no better than oil? *IDS Bulletin*, 33(3), 41–49.
- Tilly, C. (1975). *The formation of national states in Western Europe*. Princeton: Princeton University Press.
- Tornell, A., & Lane, P. (1999). The voracity effect. *American Economic Review*, 89(1), 22–46.
- Treisman, D. (2007). What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annual Review of Political Science*, 10, 211–244.
- van de Walle, N. (2001). *African economies and the politics of permanent crisis, 1979–1999*. Cambridge: Cambridge University Press.
- van de Walle, N. (2005). *Overcoming stagnation in aid-dependent countries*. Washington: Center for Global Development.
- van de Walle, D., & Mu, R. (2007). Fungibility and the flypaper effect of project aid: Micro-evidence for Vietnam. *Journal of Development Economics*, 84, 667–685.

- van der Ploeg, R., & Venables, A. (2008). *Harnessing windfall revenues in developing economies: Sovereign wealth funds and optimal tradeoffs between citizen dividends, public infrastructure and debt reduction*. Oxcarre Working Paper No. 2008–09, Oxford.
- Vreeland, J. R. (2007). *The International Monetary Fund*. London: Routledge.
- Wantchekon, L., & Vermeersch, C. (2005). Information, social networks and the demand for public goods: Experimental evidence from Benin. Manuscript, New York University.
- White, H. (Ed.), (1998). *Aid and macroeconomic performance: Theory, empirical evidence and four country cases*. Basingstoke: Macmillan Press.
- Wood, A. (2008). Looking ahead optimally in allocating aid. *World Development*, 36(7), 1135–1151.
- World Bank. (1998). *Assessing aid: What works, what doesn't, and why*. Oxford: Oxford University Press.
- World Bank. (2004). *Strategic framework for assistance to Africa: IDA and the emerging partnership model*. Washington: World Bank.
- Yano, M., & Nugent, J. B. (1999). Aid, nontraded goods, and the transfer paradox in small countries. *American Economic Review*, 89(3), 431–449.

This page intentionally left blank

Property Rights and Economic Development*

Timothy Besley and Maitreesh Ghatak

Department of Economics, R532, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

Contents

1. Introduction	4526
2. Resource Allocation and Property Rights	4528
2.1 The role of property rights in limiting expropriation	4529
2.2 Insecure property rights as a barrier to trade	4534
2.3 Optimal assignment of property rights	4545
2.4 Evidence	4552
3. Endogenous Property Rights	4559
3.1 Expropriation	4559
3.2 Improving state effectiveness	4583
4. Concluding Comments	4588
End Notes	4589
References	4592

Abstract

This chapter develops a unified analytical framework, drawing on and extending the existing literature on the subject, for studying the role of property rights in economic development. It addresses two fundamental and related questions concerning the relationship between property rights and economic activity. (i) What are the mechanisms through which property rights affect economic activity? (ii) What are the determinants of property rights? In answering these, it surveys some of the main empirical and theoretical ideas from the extensive literature on the topic.

JEL classifications: H210, O120, O170, P480, Q150

Keywords

property rights
investment incentives
economic development

1. INTRODUCTION

The term *property right* refers to an owner's right to use a good or asset for consumption and/or income generation (referred to as "use rights"). It can also include the right to transfer it to another party, in the form of a sale, gift, or bequest (referred to as "transfer rights"). A property right also typically conveys the right to contract with other parties by renting, pledging, or mortgaging a good or asset, or by allowing other parties to use it, for example, in an employment relationship.¹

While the classical economists, from Smith to Marx, accorded a central position to the role of property rights (or, "relations of production") in the process of economic development, it is only recently that mainstream economics has come around to this point of view. The core welfare results of economics concerning the role of competitive markets assume that property rights are well defined and costlessly enforced. The literature on economic growth has traditionally focused on savings and capital accumulation in an institution-free world with perfect property rights.

The new institutional approach to development economics (North, 1990) has, however, put concerns about effective property rights at the centre of thinking about development, recognizing that this requires an explicit departure from a frictionless world. According to North:

Institutions are the rules of the game in a society, or more formally, are the humanly devised constraints that shape human interaction. In consequence, they structure incentives in human exchange, whether political, social, or economic.

Seen from this perspective, property rights are an important element of the institutional structure of an economy. However, property rights are not exogenously given—they evolve over time, driven by economic and political forces. Therefore, a study of property rights also requires consideration of the arrangements, both formal and informal, that ensure that property rights are well defined and enforced. Recent advances in political economy have given greater prominence to the role of the state in codifying and protecting such rights.²

By property rights economists typically refer to *private* property rights a key feature of which is being able legally to exclude others from using a good or asset. This affects resource allocation by shaping the incentives of individuals to carry out productive activities involving the use of the good or asset, undertake investments that maintain or enhance its value, and also, to trade or lease it for other uses.³

However, other forms of property rights, such as communal property rights, are important in many societies.⁴ In the case of common property, such as a lake or a forest, individuals have use rights but do not have the right to exclude others from using it. There are also assets where the transfer rights of owners are circumscribed. For example, slavery is prohibited in modern economies. In general, property rights (both use rights

and transfer rights) are always circumscribed—for example, the owner of a plot of land does not have the right to carry out illegal activities on it. The nature of these restrictions depends on the political, legal, and enforcement system in place at a particular time and place.

When unpacking these ideas, it quickly becomes clear that there are many important facets of property rights which go to the heart of how economies work and give incentives for individuals and firms to make productive investments. The term effective property rights refer to a number of economically relevant concepts. First and foremost, it refers to the fact that ownership structures (whether collective or individual) are well defined. This has a first-order impact on the distribution of wealth and consumption. By the same token, property rights affect the pattern of production by influencing who has use rights to an asset and allowing separation of ownership from control. Thus the depth and nature of rental markets depend on the development of property rights. Such rights also affect the inter-generational evolution of the wealth distribution, by having an impact on whether assets can be transferred from parents to children. Rights may also affect the development of markets, particularly credit markets, to the extent assets can be pledged against default.

An important conceptual issue concerns the relationship between contracts and property rights. Both specify a set of decision rights: rights to take some actions or to proscribe others. In a world with perfect contracting, a rental contract is effectively equivalent to a change in ownership because these rights can be specified for every foreseeable contingency. This idea lies at the heart of the celebrated Coase theorem (Coase, 1960): in a world with complete information and no contracting costs, resource allocation will be independent of the allocation of property rights.

In a world with costly contracting, owning and renting are not equivalent since not all uses of a good or asset can be specified for all eventualities up front. A corollary of this is the idea that property rights convey *residual* control rights to the owner (Hart, 1995). These rights represent a source of freedom to those who hold them, allowing them to decide what he or she would like to do with the object (subject to any legal or technical constraints). This will also affect the holder's incentives to invest in enhancing the value of the asset, as well as those of others who might also have contractual rights to use the asset.

This chapter develops a unified analytical framework, contributing to and drawing on the existing literature on the subject, to address two fundamental and related questions concerning the relationship between property rights and economic development.

- (i) What are the mechanisms through which property rights affect economic activity?
- (ii) What are the determinants of property rights? In each case, the aim of the chapter is to survey the main ideas in the field rather than to provide an exhaustive review of the literature.

In terms of the first question, we emphasize four main aspects of how property rights affect economic activity.⁵ The first is expropriation risk—insecure property rights imply that individuals may fail to realize the fruits of their investment and efforts. Second, insecure property rights lead to costs that individuals have to incur to defend their property which, from the economic point of view, is unproductive. The third is failure to facilitate gains from trade—a productive economy requires that assets are used by those who can do so most productively and improvements in property rights facilitate this. In other words, they enable an asset’s mobility as a factor of production (e.g., via a rental market). The fourth is the use of property in supporting other transactions. Modern market economies rely on collateral to support a variety of financial market transactions and improving property rights may increase productivity by enhancing such possibilities. We will explore these arguments and discuss some of the relevant empirical evidence.

As far as the second question goes, the contribution of the chapter is to explore how systems of property rights are created and evolve over time. To understand this requires an appreciation of the gainers and losers from such rights and the institutions that shape the process by which rights are created and destroyed. Here, we look at lessons from history as well as contemporary experiences.⁶

This chapter is organized as follows. In Section 2, we take a microeconomic approach to studying how property rights affect resource allocation in theory. We use this approach as a basis for reviewing some of the empirical evidence on how property rights affect household behavior. We also review some general equilibrium implications of property rights improvements. Section 3 then discusses endogenous property rights. We look in detail at forces that shape expropriation risk. We also discuss investing resources to improve state effectiveness in improving property rights. Section 4 offers some concluding comments.

2. RESOURCE ALLOCATION AND PROPERTY RIGHTS

In this section, we examine in detail the key economic arguments about the economic role of property rights and how they affect productivity. In this section, we unify and extend the arguments for secure property rights studied in Besley (1995). We classify the various channels through which property rights affect efficiency of resource allocation under two broad categories: first, limiting expropriation, and second, facilitating market transactions. The former includes two subcategories: enhancing investment incentives by limiting expropriation risk, and reducing the need to divert private resources to protect property. The latter too includes two subcategories: facilitating trade in assets and improving collateralizability of assets, thereby facilitating credit transactions. We will discuss both individual behavioral responses as well as general

equilibrium implications. We will also discuss the insights from the literature on the property rights approach to the theory of the firm, which suggests a theory of optimal allocation of property rights.

2.1 The role of property rights in limiting expropriation

2.1.1 The basic model

We begin with a very simple set up which will allow us to illustrate a series of arguments very transparently. We begin by looking at a single producer economy. For the moment, we assume that there are no markets or any form of exchange. To fix ideas, think of this as a farmer who is endowed with a quantity of land.

We work with a very simple stochastic output model where the farmer commits effort (time) $e \in [0, 1]$ of which he has an endowment $\bar{e} \leq 1$. This yields output A with probability \sqrt{e} and zero with probability $1 - \sqrt{e}$. Expected output y is therefore:

$$y = A\sqrt{e}. \quad (1)$$

In this single input setting, the farmer's decision is to choose his optimal level of e . Since there are no labor markets, this choice will be driven by his own disutility cost of supplying labor.

We assume that the farmer's utility function is linear in consumption (c) and leisure (l):

$$u(c, l) = c + l. \quad (2)$$

This formulation rules out income effects and risk aversion.

We assume that property rights are imperfect in the sense that there is an exogenously given probability $\tau \in [0, 1]$ of expropriation. This could apply to the output that is produced, or the land which is needed to produce output. These are equivalent, so long as labor is a sunk input prior to whether or not there is going to be expropriation.⁷

Given this formulation, expected consumption is $c = (1 - \tau)A\sqrt{e}$. At this stage, we make no distinction between expropriation and taxation nor do we consider the choice of τ . The implicit assumption, which we will make more precise later, is that there is an actor in the economy with coercive power which can be used to tax, confiscate, or steal. In Section 3, we discuss the factors that determine the choice of τ .

The producer selects e to maximize:

$$(1 - \tau)A\sqrt{e} + \bar{e} - e \quad (3)$$

subject to the constraint $e \leq \bar{e}$. The first-order condition for an interior solution is:

$$\frac{(1 - \tau)A}{2\sqrt{e}} = 1. \quad (4)$$

The optimal choice of labor of the producer is therefore given by:

$$e^* = \left[\frac{(1 - \tau)A}{2} \right]^2. \quad (5)$$

Since we require that $e \leq 1$, we assume throughout that $A \leq 2$. Correspondingly, (expected) gross output is $\gamma(\tau) = ((1 - \tau)A^2)/2$, and the producer's net surplus (taking into account the cost of e) is given by $\pi(\tau) = [(1 - \tau)A/2]^2 + \bar{e}$. Using this, we have the following observation:

Result 1. *Labor supply, output and profits are strictly decreasing in τ .*

This is really only like a standard model in which taxes create a disincentive to commit effort. In this risk neutral setting, it also does not matter whether τ is a fixed or known proportion of output, as with a tax or a probability of full expropriation of all output. This result underpins the standard "security" argument in favor of property rights which allow lower τ . The same logic would extend to other inputs such as fertilizer or land improvements.

There are three key assumptions that drive this result. First, the input is sunk before the farmer knows whether there is going to be expropriation or not. Second, more efficient instruments for transfer are not available. Therefore, as with any form of outcome-contingent transfer policy, there is a standard disincentive effect. A lump-sum tax or a "profit-tax" would benefit both the farmer and the coercive authority.⁸ Third, the resource-endowment constraint (here, labor) is not binding.

To explore the importance of the latter, suppose that the resource constraint is binding, that is, $e^* = \bar{e}$. In this case, gross output is $A\sqrt{\bar{e}}$, and the producer's net surplus is $(1 - \tau)A\sqrt{\bar{e}}$. At this corner solution, marginal changes in τ have only distributional implications: labor supply and gross output are unaffected.

If competitive labor markets exist, then resource constraints are unlikely to be binding.⁹ To see this, suppose that e can be sold in the market with $w = 1$. We would get the same outcome in terms of productive efficiency in the benchmark model irrespective of the specific form of preferences of the producer, or his endowments, such as \bar{e} . In particular, the outcome will be the same whether or not the labor endowment constraint binds. If, for example, $e^* \geq \bar{e}$ the producer would hire in labor from the labor market.¹⁰ The effect of τ would, of course, stay the same: like a tax, it distorts labor usage.

2.1.2 Guard labor

In the basic model, there is only one margin of choice: how much labor to put into production. Suppose now that labor can also be used to reduce the risk of expropriation. This potentially creates an additional margin of distortion caused by imperfect property rights. Poor property rights not only reduce incentives to supply productive

labor, it also diverts resources (here labor) from productive to unproductive uses. Improvements in the protection of property rights can then free up labor and enable households to make unconstrained decisions.

There are two cases to consider. First, where the asset that is subject to insecure property rights is involved in the production or income-generation process, as in our basic model. A good example of this is agricultural land. Second, where the asset subject to insecure property rights is not directly involved in the production or income-generation process. Residential property is a good example of this.

To explore this, we modify the model by having two types of labor. Let $e_1 \in [0, 1]$ denote “productive” labor and $e_2 \in [0, 1]$ denote “guard” labor that reduces the probability of expropriation. We use a simple technology to describe the probability of expropriation: $\tau(1 - \gamma\sqrt{e_2})$, where $\tau \in [0, 1]$ and $\gamma \in [0, 1]$. This captures very simply the idea that expropriation is lower if e_2 is higher with γ representing the effectiveness of efforts put into guard labor. Otherwise the model is the same as the basic model, with $A\sqrt{e_1}$ denoting expected output. Now the producer’s decision problem is:

$$\max_{e_1, e_2} \left(1 - \tau(1 - \gamma\sqrt{e_2})\right) A\sqrt{e_1} + \bar{e} - e_1 - e_2. \quad (6)$$

Solving the first-order conditions for both effort choices yields:

$$e_1 = \left(\frac{2(1 - \tau)A}{4 - (\tau\gamma A)^2} \right)^2 \text{ and } e_2 = \left(\frac{\gamma\tau(1 - \tau)A^2}{4 - (\tau\gamma A)^2} \right)^2. \quad (7)$$

Several interesting implications follow immediately from these two expressions:¹¹

Result 2. *If the insecure asset is involved in the production process, then in the case where the resource constraint is not binding: (i) improved property rights (lower τ) increases productive labor; (ii) there exists $\bar{\tau} \leq 1$ such that guard labor is increasing in τ so long as $\tau \leq \bar{\tau}$ and decreasing otherwise; and (iii) economic efficiency is increasing in improved property rights (lower τ).*

This result says that the link between productive labor and secure property rights remains. However, the effect of property rights security on guard labor is ambiguous in sign.¹²

The intuition for this finding is as follows. As productive and guard labor are *complementary*, more effort to protect property rights will raise the expected marginal returns from efforts to produce more output.¹³ Formally, e_1 is increasing in γ , and so compared to the basic model, introducing guard labor increases productive labor. Given this, there are two effects of increasing τ on e_1 as can be seen from the first-order condition. The direct effect is negative for the same reasons as in the basic model. But there is an indirect effect operating via e_2 in the presence of guard labor. However, this

effect is always dominated by the direct effect. For (ii) observe that an increase in τ raises the expected marginal return from guard labor while lowering e_1 . The complementarity between e_1 and e_2 means that this tends to reduce the expected marginal return from guard labor. For small values of τ the first effect dominates and for larger values of τ , the second effect dominates. However, as one would expect, economic efficiency increases when property rights are more secure following the logic of the previous section: namely, because it is a first-order “tax” on output.

Consider now what happens when the resource (i.e., labor endowment) constraint is binding (i.e., $(1 - \tau)^2 A^2 (4 + \tau^2 \gamma^2 A^2) / (4 - \tau^2 \gamma^2 A^2)^2 > \bar{e}$). Then the first-order conditions are:

$$\begin{aligned} (1 - \tau + \tau\gamma\sqrt{e_2})A \frac{1}{2\sqrt{e_1}} &= 1 + \lambda, \\ \tau\gamma \frac{1}{2\sqrt{e_2}} A\sqrt{e_1} &= 1 + \lambda, \end{aligned} \tag{8}$$

where λ is the Lagrangian multiplier associated with the binding resource constraint (the shadow price of labor). Using these two conditions together with the binding labor-endowment constraint, we obtain the following quadratic equation determining $\sqrt{e_2}$:

$$2\tau\gamma e_2 + (1 - \tau)\sqrt{e_2} - \tau\gamma\bar{e} = 0. \tag{9}$$

Solving (and picking the larger root as the smaller root is negative) we obtain:

$$\begin{aligned} e_1 &= \bar{e} - \left[\frac{1}{4\gamma} \left(1 - \frac{1}{\tau} \right) + \sqrt{\left\{ \frac{1}{4\gamma} \left(1 - \frac{1}{\tau} \right) \right\}^2 + \frac{\bar{e}}{2}} \right]^2, \\ e_2 &= \left(\frac{1}{4\gamma} \left(1 - \frac{1}{\tau} \right) + \sqrt{\left\{ \frac{1}{4\gamma} \left(1 - \frac{1}{\tau} \right) \right\}^2 + \frac{\bar{e}}{2}} \right)^2. \end{aligned} \tag{10}$$

It is now straightforward to check that e_2 is always increasing in τ and e_1 is always decreasing in τ . Also, now anything that raises e_2 (e.g., an increase in γ) will directly reduce e_1 via the binding labor-endowment constraint. In this case, productive and guard labor are *substitutes*, and the intuition that guard labor diverts resources away from productive uses applies quite clearly.

We next consider the case where the insecure asset is not involved in the production process. This could apply, for example, if residential property is subject to insecure property rights. This might affect labor supply decisions even though the asset is not directly used for income generation. Suppose the asset is worth \bar{h} to the producer if property rights are not violated and is worth \underline{h} otherwise. As before, let e_1 and e_2 be productive and guard labor. In this case, $A\sqrt{e_1}$ is expected income and $(1 - \tau(1 - \gamma\sqrt{e_2}))$ is the probability that property rights are not violated. Therefore, the producer's decision is now characterized by:

$$\max_{e_1, e_2} \left(1 - \tau(1 - \gamma\sqrt{e_2})\right)\bar{h} + \tau(1 - \gamma\sqrt{e_2})\underline{h} + A\sqrt{e_1} + \bar{e} - e_1 - e_2. \quad (11)$$

For this case, we have:

Result 3. *If the insecure asset is not involved in the production process, then in the case where the resource constraint is not binding, the productive and guard labor supply decisions are independent and accordingly, e_1 is unaffected by τ .*

If the labor endowment constraint is binding, as before, e_1 and e_2 are substitutes and any reduction in guard labor will increase productive labor. In this case, if τ goes up, then e_2 goes up and therefore e_1 has to go down. Therefore, e_1 is decreasing in τ . If e_1 and e_2 are substitutes in the disutility of labor (e.g., the cost of labor being $e_1 + e_2 + \phi e_1 e_2$ where $\phi > 0$) then this effect is further reinforced.

Note, however, that a binding labor-endowment constraint is an issue only when the labor market is imperfect or absent. Otherwise, the opportunity to hire labor at a given wage rate should, in principle, make the cost function linear and separable as is the case when the labor endowment constraint is not binding. However, it may be that there are difficult agency problems in hiring guard labor, that is, preventing the hired guards from appropriating the asset which would need to be considered.

We have abstracted so far from income effects by making the assumption of linear preferences over consumption and leisure. If this is not the case, then there is a further channel through which property rights can affect resource allocation. To see this, consider a slight modification of the above model. Suppose that the insecure asset is not involved in the production process. However, in the utility function of the producer, consumption (e.g., food) and the asset (e.g., consumption value of housing) are complements. The producer then maximizes expected utility as follows:

$$\max_{e_1, e_2} \left(1 - \tau(1 - \gamma\sqrt{e_2})\right) \left(A\sqrt{e_1}\right)^\alpha \left(\bar{h}\right)^\beta + \tau(1 - \gamma\sqrt{e_2}) \left(A\sqrt{e_1}\right)^\alpha \left(\underline{h}\right)^\beta + \bar{e} - e_1 - e_2, \quad (12)$$

where $\alpha \in (0, 1)$, $\beta \in (0, 1)$ and $\alpha + \beta < 1$. In this case, the first-order conditions are:

$$\begin{aligned} \left\{ (1 - \tau(1 - \gamma\sqrt{e_2}))\bar{h}^\beta + \tau(1 - \gamma\sqrt{e_2})\underline{h}^\beta \right\} \frac{\alpha}{2} A^\alpha (e_1)^{(\alpha/2)-1} &= 1, \\ \frac{\tau\gamma}{2\sqrt{e_2}} A^\alpha e_1^{\alpha/2} (\bar{h}^\beta - \underline{h}^\beta) &= 1. \end{aligned} \quad (13)$$

Substituting e_2 from the second equation to the first, and then totally differentiating with respect to τ it is straightforward to verify that $\partial e_1 / \partial \tau < 0$ for small values of τ . Thus worsening property rights protection reduces productive effort. The intuition is as follows: the expected marginal return from supplying productive labor falls when τ goes up as consumption is complementary with the asset that is subject to insecure property rights. Clearly, if there is a competitive insurance market then the risk of losing the asset can be insured away, and once again τ will not affect e_1 .

To summarize, there is a variety of ways that guard labor supplied in response to insecure property rights can be modeled. Moreover, the theoretical predictions are somewhat sensitive to the case being considered. Thus broad brush conclusions are probably not warranted even though there are a number of reasonable cases where the intuitive idea, that less secure property rights encourage the use of guard labor, emerges from the analysis.

There is a literature that deals with the general equilibrium effects of guard labor (or, more broadly, self-defense) in a model similar to the one above, but with many producers. The key idea is that individual investments in protection entail a negative externality on the other producers as predators are deflected from the protected to the unprotected properties. This implies that the decentralized equilibrium is generally inefficient as it has too much protection.¹⁴

2.2 Insecure property rights as a barrier to trade

The effects that we have studied so far could be studied in the absence of markets. One key role of property rights is to facilitate exchange and allow producers/consumers to exploit gains from trade. In the following two sections we examine the role of property rights in facilitating exchange in land markets (rental, sales) and in credit markets, respectively.

2.2.1 Property rights and trade in assets

Economic efficiency is enhanced by having assets managed by those who can use them most productively. But this depends on being able to write efficient contracts to trade. In our basic model everyone has the same amount of land, and also, everyone has the same skill level. As a result, so long as there is a competitive labor market, there are no efficiency gains from having a land market. Now we relax this assumption and

allow some agents to have more land than they want to optimally cultivate themselves, and some agents to have less. This creates potential gains from trade via a rental or sales market in land. But a necessary (but not sufficient) condition for this to take place is to have well-defined property rights in land. Otherwise, land will not be offered for rental or sale driven by the fear that lenders could lose the land with some probability, or equivalently, receive only a fraction of the market returns to land due to imperfect property rights in land. This will create an additional margin of distortion due to imperfect property rights. As a consequence, potentially gainful trades will be lost.

To model this in a simple way we assume there is a continuum of agents divided into landed (a fraction of δ) and landless (a fraction $(1 - \delta)$). Suppose that time is infinite and rental contracts involve an up-front payment from the landless farmer to the landlord. However, there is a probability τ of losing ownership of the land at the end of the rental contract which we assume to be one period.

At the beginning of each period a farmer receives a productivity shock $\theta \in \{\underline{\theta}, \bar{\theta}\}$ with $0 \leq \underline{\theta} < \bar{\theta} \leq 1$. Let the probability of low productivity $\theta = \underline{\theta}$ be p . This is assumed to be distributed independently and identically across individuals, as well as over time (for the same individual).

Given θ , output is $\theta A\sqrt{e}$. Therefore, for a given θ , a producer who owns land chooses:

$$\max_e \theta A\sqrt{e} + \bar{e} - e. \quad (14)$$

This yields, given perfect property rights (and ignoring corner solutions): $e^* = [\theta A/2]^2$ and $\pi^*(\theta) = [\theta A/2]^2 + \bar{e}$. From now on, we set $\bar{e} = 0$.

For a landless individual or someone who leases out land, there is an alternative activity which could be thought of as working for a wage, that yields utility $\bar{u} \geq 0$. We assume that:

$$\pi^*(\underline{\theta}) > \bar{u}, \quad (15)$$

that is, that any landowner prefers to operate his land to taking the outside opportunity. In this situation, there are clearly gains from trade.

Suppose both landed and landless farmers face the same distribution of productivity shocks. Then there is a fraction $p\delta$ of which is low productivity and landed and a fraction $(1 - p)(1 - \delta)$ which is high productivity and landless. Assume that

$$(1 - p)(1 - \delta) > p\delta \text{ or } 1 > \delta + p. \quad (16)$$

This says that there are more high productivity and landless than there are low productivity and landed. Given this, in a competitive market, land is scarce and rents will accrue to land owners.

In a perfect rental market land trades at a price

$$r^* = \pi^*(\bar{\theta}) - \bar{u}. \quad (17)$$

All land is fully utilized and has high productivity.

Now let us consider the decision problem when there is a probability τ that the tenant will not return the land. Now we contrast two strategies for a low-productivity landlord: renting out the land and bearing the risk of losing his land or cultivating it himself. As productivity shocks are assumed to be i.i.d. over time, and in any future period when the landowner is lucky and draws $\bar{\theta}$ he would prefer to cultivate the land himself as this way he does not bear the risk of losing it. Following this argument, we can now set up two value functions, one which we call V when in the current period land is rented out, and one which we call W when in the current period the landowner cultivates the land himself. Then,

$$\begin{aligned} V &= \pi^*(\bar{\theta}) + \beta(1 - \tau)[(1 - p)W + pV], \\ W &= \pi^*(\bar{\theta}) + \beta[(1 - p)W + pV]. \end{aligned} \quad (18)$$

Solving for W as a function of V yields

$$W = \frac{\pi^*(\bar{\theta}) + \beta p V}{1 - \beta(1 - p)}. \quad (19)$$

We can now plug W into V , and after some manipulation we obtain

$$V = \frac{1 - \beta\tau(1 - p)}{1 - (1 - \tau p)\beta}\pi^*(\bar{\theta}). \quad (20)$$

Observe that V is decreasing in τ , as we would expect.

Consider the autarky option whereby a landowner always cultivates his own land. Let V' and W' denote his lifetime expected payoff from autarky when, respectively, he has a low- and a high-productivity shock in the current period:

$$\begin{aligned} V' &\equiv \pi^*(\underline{\theta}) + \beta\{pV' + (1 - p)W'\}, \\ W' &\equiv \pi^*(\bar{\theta}) + \beta\{pV' + (1 - p)W'\}. \end{aligned} \quad (21)$$

Solving these, we get:

$$V' = \frac{\pi^*(\underline{\theta})(1 - \beta(1 - p)) + \beta(1 - p)\pi^*(\bar{\theta})}{1 - \beta}. \quad (22)$$

Comparing V and V' we can see that if τ is small then $V > V'$ because in the limit when $\tau = 0$, V has to exceed V' as the land is always with a high-productivity producer and the owner gets the full surplus. Consider the opposite case when τ is high. Now there is a trade-off: with autarky there are periods when the land is used unproductively, and with tenancy, there is a risk that the owner may lose the land. Take the extreme case where $\tau = 1$. Now it is easy to check that if

$$\frac{\pi^*(\underline{\theta})}{\pi^*(\bar{\theta})} > \left(1 - \frac{\beta}{1 - \beta + \beta p}\right)$$

then $V' > V$. A sufficient condition for this is $\beta > 1/(2 - p)$, in which case the right-hand side is negative and so even if $\pi^*(\underline{\theta}) = 0$ the condition would be satisfied. Naturally, if $V' > V$ for $\tau = 1$ by continuity and the fact that V is monotonically decreasing in τ , we have the following result:

Result 4. *If $\beta > 1/(2 - p)$, then there is a $\hat{\tau} \in (0, 1)$ such that for $\tau \geq \hat{\tau}$ there is no trade in assets and land is cultivated by low-productivity farmers.*

The insecure property rights now lead to no trade and a *per capita* output loss equal to $\delta p[\pi^*(\bar{\theta}) - \pi^*(\underline{\theta})]$. In this case, a fall in τ constitutes a Pareto improvement because those who rent out their land are better off, while those who rent in land are indifferent.

In the case $\pi^*(\underline{\theta}) = 0$ the autarky option, in a period the producer receives a low-productivity shock, is equivalent to keeping the land idle. This is consistent with the fact that in the developing world assets are often kept undeveloped or idle due to insecure property rights.¹⁵ Increasing the security of property rights can therefore reduce the extent to which assets are underutilized.

2.2.2 Property rights and collateralizability of assets

Above, we showed that property rights facilitate trade in assets and thereby achieve efficient allocation of resources. In the presence of agency costs, effective property rights can facilitate the use of assets to mitigate agency costs, thereby facilitating trade. A prime example of this is in the credit market; when agency or enforcement costs are important, lenders may not be willing to lend an efficient amount or, in some cases, lend at all. Property rights improve the ability of borrowers to pledge their assets as collateral, and thereby relax credit constraints.¹⁶

A recent influential advocate of the importance of this link between property rights and economic efficiency is de Soto (2000, 2001) who calls this the problem of “dead capital.” For example, he argues that:

What the poor lack is easy access to the property mechanisms that could legally fix the economic potential of their assets so that they could be used to produce, secure, or guarantee greater value in the expanded market. de Soto (2001).

He proposes the following metaphor:

Just as a lake needs a hydroelectric plant to produce usable energy, assets need a formal property system to produce significant surplus value. de Soto (2000, p. 48).

While de Soto is the modern incarnation of this view, it has an important lineage. For example, in his perceptive study of West African trade, Bauer (1954) also recognizes the importance of poorly developed property rights and the impediment to trade that they create when he observes that:

Both in Nigeria and in the Gold Coast family and tribal rights in rural land are unsatisfactory for loans. This obstructs the flow and application of capital to certain uses of high return, which retards the growth of income and hence accumulation. (p. 9).

To explore these issues, we use the same basic model as above. Thus, \sqrt{e} remains the probability that output is A . We now assume explicitly that $e \in [0, 1]$ is private information to the producer (borrower) and set $\bar{e} = 0$ for simplicity. In addition to committing effort, we now allow the producer to use capital to enhance productivity. For simplicity, capital x is a discrete variable that takes on the values 0 and 1. When $x = 1$, output is $A(1 + \Delta)$ with probability \sqrt{e} and 0 with probability $1 - \sqrt{e}$. Thus, expected output is $A(1 + \Delta)\sqrt{e}$. The cost of a unit of capital is ρ , which for now is exogenously given. We abstract from any direct insecurity of property rights to focus on how they work through the ability to pledge assets. Given this, and absent any frictions, the producer's decision problem is:

$$\max_{e \in (0,1), x \in \{0,1\}} A(1 + \Delta x)\sqrt{e} - e - \rho x. \quad (23)$$

The optimal choice of effort, e , is given by:

$$e = \left(\frac{A(1 + \Delta x)}{2} \right)^2. \quad (24)$$

In this model the capital good x and effort are complements. The expected surplus at the optimal effort level is

$$\frac{1}{4}A^2(1 + \Delta x)^2 - \rho x. \quad (25)$$

For concreteness sake, we assume

$$\frac{1}{4}[A(1 + \Delta)]^2 - \rho > \frac{1}{4}A^2 \quad \text{and} \quad \frac{A(1 + \Delta)}{2} < 1. \quad (26)$$

The first condition ensures that under the first best (where effort is observable), it is profitable to use the capital good. The second assumption ensures an interior solution for e .¹⁷ We will therefore refer to $e^* = [A(1 + \Delta)/2]^2$ as the first-best level of effort.

If the producer owned the capital, or if there were no moral hazard, that is, a lender could specify a level of effort as a condition of lending to the producer, then effort as above would be efficient and $x = 1$ would be optimal. The analysis is more interesting when we make two key assumptions: (i) effort is unobservable and hence cannot be specified in lending contracts (moral hazard) and (ii) the producer has insufficient wealth to post as a bond in the event that he defaults (limited liability). To capture the latter, we suppose that the producer has an illiquid asset whose value is w . We assume, however, that the assets can be pledged as collateral against borrowing x from the lender. Limited liability implies that he can pay only up to $A(1 + \Delta) + w$, when output is high and w when output is low.

If illiquid wealth were large enough, we would be back to first-best case. It is as if effort could be specified in the contract. By varying the level of collateral demanded, the lender could make the stakes high enough for the borrower so that he puts in the first-best effort level.

It is now clear why property right imperfections will enter the story. Even if the producer has some illiquid wealth that could be pledged as collateral, it is necessary that the legal environment be able to support its use as a bond against not repaying the loan. This is particularly striking in the case where the level of illiquid wealth that the producer owns is large enough to alleviate the moral hazard problem entirely but is prevented from doing so by insecurity of title to that wealth. The illiquid wealth in this case is “dead capital” in do Soto’s sense. As we shall see, an economy could then be constrained (in terms of output and efficiency) by the absence of secure title rather than by absence of wealth.

For the purposes of our exposition here, we model this constraint on contracting in a very simple way. Suppose that if a borrower has wealth w , then its collateral value is $(1 - \tau)w$, that is, only a fraction of that wealth can be used as effective collateral. This

could be given a stochastic interpretation: with probability $1 - \tau$ the lender will be able to foreclose on the asset that was pledged as collateral if output is low and the borrower is unable to repay his loan from the output/revenue of his project.¹⁸ In concrete terms, the parameter τ reflects that in many countries registering assets as property is time consuming and costly.

To understand how property rights matter, we now solve for the optimal debt contract as a function of τ . We will then explore how changing τ affects optimal debt contracts. A debt contract is an interest payment on a successful project, denoted by r , and a level of collateral, denoted by c , to be paid if the project is unsuccessful. The expected payoff of the producer with a contract (r, c) is:

$$\sqrt{e}\{A(1 + \Delta) - r\} - (1 - \sqrt{e})c - e \quad (27)$$

while that of a lender is:

$$\sqrt{e}r + (1 - \sqrt{e})c - \rho. \quad (28)$$

The producer always has the option of not borrowing x . This creates an outside option equal to $\frac{1}{4}A^2$. Assumption (26) guarantees that (in principle) there are gains from trade as long as effort can be specified in the contract. A loan transaction takes place so long as the producer's expected payoff is above her outside option and the lender makes nonnegative expected profits. Otherwise, the producer is credit-constrained.

Given r and c the producer chooses her effort to maximize her expected payoff, which yields the first-order condition:

$$\frac{1}{2\sqrt{e}}\{A(1 + \Delta) - (r - c)\} = 1. \quad (29)$$

Solving this yields an optimal effort level:

$$e = \left[\frac{A(1 + \Delta) - (r - c)}{2} \right]^2. \quad (30)$$

This is the incentive compatibility constraint of the borrower. Observe that e and r are negatively related, while e and c are positively related. This is intuitive as r is a tax on success, while c is a penalty for failure.

In addition, the contract also has to satisfy the limited liability constraint:

$$(1 - \tau)w \geq c. \quad (31)$$

This says that the payment demanded from the producer when the project is unsuccessful cannot exceed her pledgeable wealth.

Inspecting Eq. (30), it may appear as if it is possible to achieve the first-best effort level by setting $r = c$. However, since c cannot exceed $(1 - \tau) w$ this might not be enough for the lender to recover the opportunity cost of capital (ρ). If that is the case, then the lender will need to set $r > \rho > c$. This will imply that effort will fall below the efficient level. This illustrates how agency costs have bite in this world.

We now sketch how the lender will fix the optimal contract when the incentive compatibility and limited liability constraints are binding. Substituting Eqs. (30) and (31) into the lender's payoff function yields the following single variable decision problem to determine the optimal interest payment:

$$\max_r \frac{A(1 + \Delta) - (r - w(1 - \tau))}{2} (r - w(1 - \tau)) + w(1 - \tau) - \rho. \quad (32)$$

Solving this yields:

$$r = \frac{A(1 + \Delta)}{2} + w(1 - \tau). \quad (33)$$

In this case, the lender takes one half the return from a successful project in addition to the value of the pledged collateral. The effort level that the producer puts in is therefore:

$$e = \left[\frac{A(1 + \Delta)}{4} \right]^2 \quad (34)$$

which is below the first-best level. Notice that this result does not depend on the security of collateral— τ . The borrower's and the lender's expected payoffs are, respectively: $u \equiv \{A(1 + \Delta)/4\}^2 - w(1 - \tau)$ and

$$\pi \equiv \frac{1}{2} \left\{ \frac{A(1 + \Delta)}{2} \right\}^2 + w(1 - \tau) - \rho.$$

For trade to take place on these terms, we require that $\frac{1}{4}u \geq \frac{1}{4}\pi^2$. This will happen when

$$w(1 - \tau) \leq \frac{A^2}{4} \left[\frac{(1 + \Delta)^2}{4} - 1 \right] \equiv \underline{\omega}.$$

When the outside option is a binding constraint, then r will be determined by:

$$\left\{ \frac{A(1 + \Delta) - (r - w(1 - \tau))}{2} \right\}^2 - w(1 - \tau) = \frac{1}{4} A^2. \quad (35)$$

This yields

$$r = A(1 + \Delta) - 2\sqrt{\frac{A^2}{4} + w(1 - \tau)} + w(1 - \tau), \quad (36)$$

with effort equal to $(A^2/4) + w(1 - \tau)$.¹⁹ Now effort is a (decreasing) function of the security of collateral.

We can now define precisely when pledgeable wealth is a constraint on economic efficiency. This will be the case if wealth is insufficient for the first-best effort level to be attainable, that is,

$$\sqrt{\frac{A^2}{4} + w(1 - \tau)} \leq \frac{A(1 + \Delta)}{2}$$

or

$$w(1 - \tau) \leq \frac{A^2}{4} [(1 + \Delta)^2 - 1] \equiv \bar{\omega}. \quad (37)$$

If $w(1 - \tau) > \bar{\omega}$ then we have a first-best outcome. Evidently, this requires that the availability of illiquid assets (w) has to be large enough. However, this is not sufficient— τ must also be far enough away from one. An economy is constrained by property rights when $w \geq \bar{\omega} > w(1 - \tau)$. For $\bar{\omega} > w$ imperfect property rights increase the existing level of inefficiency, while for $w \geq \bar{\omega} > w(1 - \tau)$ imperfect property rights create new inefficiencies.

As in previous sections, we turn our focus now to what happens when τ changes marginally. Our simple setup allows us to get a complete understanding of the comparative statics of the optimal contract. Our main result drops cleanly out of the analysis.²⁰

Result 5. *For $w(1 - \tau) \in [\underline{\omega}, \bar{\omega}]$, the interest payment, r , is lower and producer effort is greater after a marginal increase in the security of collateral which increases the level of pledgeable wealth, $w(1 - \tau)$. For $w(1 - \tau) < \underline{\omega}$, or $w(1 - \tau) > \bar{\omega}$, marginal improvements in the security of collateral do not affect resource allocation (i.e., loan size and effort) in the credit market. However, in the former case, it has a redistributive effect with lenders gaining relative to borrowers.*

The result captures the mechanism suggested by de Soto (2000) linking property rights that increase the use of collateral and efficiency. However, it also makes precise

the range of illiquid wealth for which this argument is relevant. If wealth is very low, that is, $w(1 - \tau) < \underline{\omega}$, then the outside option constraint is not binding. In this case, the terms of the contract are affected by improvements in property rights, but there is no increase in effort conditional on credit being granted. However, improvement in property rights eases the constraint of transferring resources from the borrower to the lender, and this benefits the lender at the expense of the rent that the borrower gets. Improving property rights have a purely redistributive effect in this case. Similarly, if wealth is very high, the resource allocation is already efficient at the first-best level, and therefore, marginal improvements in property rights will not have any effect.

The upshot of this discussion is that even where there is a “de Soto effect” on effort observed (or, loan size), we would expect that effect to be heterogeneous with $\partial e / \partial \tau$ being proportional to illiquid wealth w . Those with larger levels of illiquid wealth will respond more strongly to a given improvement in property rights. However, beyond the pledgeable wealth of $\bar{\omega}$, the effect again becomes zero.

This illustrates the importance of modeling in seeking to study the impact of property rights improvements on economic outcomes through the collateral channel. Looking for an average effect across a group of producers with heterogeneous wealth could well underestimate the impact which we would expect to find only in the middle wealth group.

There are also implications for looking at the effect of improving property rights in aggregate data. The size of the gains from reducing τ will depend on the distribution of wealth. In particular, in very rich, very poor, or very unequal societies (comprising only very rich or very poor) the overall effect will not be large.

Our model can also highlight another set of effects that have been largely ignored in the empirical and theoretical literature to date. So far our analysis has not considered how changing property rights affects the structure of the credit market and who trades with whom. To illustrate this, suppose that there are many potential lenders who vary in their opportunity costs of capital, ρ , determined by their access to loanable funds.

A simple way to thinking about this is to consider a two-sector model using the labels formal and informal to describe the lenders. In the formal sector, there is a common transactions technology $1 - \tau_F$ and access to funds $\rho = \rho_F$. We imagine that producers are also connected to potential lenders through social networks in which case they face property rights enforcement $1 - \tau_N$ and lenders with cost of loanable funds ρ_N . The most natural and interesting case to study is where $\tau_F > \tau_N$ and $\rho_F < \rho_N$. This says that formal lenders have better access to loanable funds while the informal sector is better at enforcing contracts. If networks had both lower ρ and lower τ then they would clearly dominate the formal sector.

We will not provide a complete treatment of how people are assigned to the two sectors—that would require a more involved analysis than can be undertaken here. Instead, we will look at some of the issues that arise as property rights change. The

analytical change that is needed to study this is to recognize that the relevant outside option for a producer may no longer be $\frac{1}{4}A^2$ but trading with another lender.

Suppose (somewhat unrealistically) that both networks and markets are competitive so that lender rents are bid to zero in each. Then it is straightforward to show that the level of producer utility is:

$$U(\tau_i, \rho_i) = \left[\frac{A(1 + \Delta) + \sqrt{[A(1 + \Delta)]^2 + 8[w(1 - \tau_i) - \rho_i]}}{4} \right]^2 - w(1 - \tau_i), \quad (38)$$

where $i \in \{F, N\}$. We assume $u_i > \frac{1}{4}A^2$ because, otherwise given Eq. (26) no trade will take place. In this competitive world, we would expect the producer to match with the lender for whom this zero profit utility is greatest. Thus, the formal sector will dominate if $U(\tau_F, \rho_F) > U(\tau_N, \rho_N)$. It is clear now that improving formal sector property rights can potentially lead to a move from networks to formal lending as τ_F falls. Since effort is now set by the outside option and is equal to $[(A(1 + \Delta) + \sqrt{[A(1 + \Delta)]^2 + 8[w(1 - \tau_i) - \rho_i]})/4]^2$, moving to a more efficient producer now leads also to greater efficiency. This is a general equilibrium response to an improvement in property rights allowing trade to prosper in its most efficient form. It is related to the effect identified in Section 2.2.1. However, it is now the effect of improved property rights to allow superior trade in another market, the credit market, that drives the result.

There are other possible general equilibrium effects to consider if we move away from the perfect competition story. In the other extreme suppose that there is a single network lender and a single formal sector lender. Each gets to propose a contract to a producer and she picks her preferred outcome. In this case, the reservation outcome is now set by the outside opportunity available either in autarky or else by trading in the other sector. Suppose that the latter is the case. In this case, a producer who chooses to trade in a network will be affected by an improvement in formal sector property rights even if she chooses not to obtain credit in the formal sector. This is because of a pure outside option effect. Improving formal property rights now, through this route, increases effort in the network. However, if trading in the other sector does not provide a good enough outside option (e.g., the borrowers are poor, or the cost difference is large), then an improvement in property rights will benefit the lender and hurt the borrower without having any efficiency effects, as discussed earlier.

Finally, there is the possibility that improving property rights increases competition.²¹ To see this, we need to suppose that there are different possible levels of ρ_F with some formal sector firms being more efficient. Suppose, for example, there is no informal sector, but two formal sector lenders with different levels of ρ_F but the same level

of τ_F . Suppose that the cost difference between the two lenders and the level of τ_F are such that the higher cost formal sector lender cannot provide any competition to the lower cost lender, and autarchy is the only outside option of a borrower. A further effect of improving τ_F can now be to induce entry in the formal sector increasing the outside option of the producer. This leads to a redistribution of surplus from the efficient formal sector producer to the producer. But it also increases efficiency by increasing the outside option of the producer. This will increase producer effort.

The latter effects that we have identified come from thinking about how the improvement of property rights affects the set of potential trades that can be sustained between lenders and producers. One feature of formal sector enforcement is that it is a freely available contracting technology, whereas the τ_N is available only for trades between people who know each other. When considering property rights that improve trading possibilities the benefits from the creation of formal property rights may in significant measure be due to the fact that these are widely available, that is, to all producers rather than just those who are socially connected. This highlights a potential downside in the use of networks in enforcing trade.

2.3 Optimal assignment of property rights

So far we have discussed how insecure property rights impede efficiency by undermining investment incentives, and creating barriers to trade. Consequently, our analysis has focused on the channels through which making property rights more secure for the producer will improve efficiency. This implicitly assumes that the initial assignment of property rights to the producer is optimal.

In this section, we question this and discuss the role of property rights in assigning ownership to an asset to maximize its productive potential. We have already looked at one aspect of this issue in Section 2.2.1 where we allowed for the possibility that the current owner may not be the most efficiency potential user of an asset. The aspect that we address here allow for the possibility that more than one party can invest to improve the productivity of an asset. Our discussion of these issues is based on the literature on the property rights approach to the theory of the firm developed in Grossman and Hart (1986) and Hart and Moore (1990).

2.3.1 Optimal ownership of an asset

We extend our benchmark model above by considering two individuals, A and B, who undertake investments e_A and e_B that, in combination with the asset, generate returns $a\sqrt{e_A} + b\sqrt{e_B}$. The costs of these investments to A and B are, respectively, e_A and e_B . The terminology “investment” here as opposed to “effort” in the last section emphasizes the durability of the activity. We have in mind that the effort undertaken by each party creates something which is potentially of value to the future output from the asset even if the party who makes the decision is separated from the asset.

The first-best levels of these investments are:

$$e_A^* = \frac{a^2}{4} \quad \text{and} \quad e_B^* = \frac{b^2}{4}. \quad (39)$$

The associated total surplus is:

$$S^* = \frac{1}{4}(a^2 + b^2). \quad (40)$$

Without any contracting problems, ownership does not have allocative implications, that is, a contract can be written in which investment levels $\{e_A^*, e_B^*\}$ are prescribed.

The key insight of the property rights approach is that ownership matters due to contractual incompleteness. In this example, the owner has some bargaining power as he can threaten to exclude the other party from using the asset (i.e., he can “fire” the other party and exclude him from the returns from his investment). Ownership is now different from residual claimancy of a profit stream: it is the residual control right over the asset.

If the owner of an asset rents it out to someone, the tenant has residual claimancy. However, the owner retains the right not to renew the lease. This will potentially affect the incentive of the tenant to improve the asset. It is these residual control rights that give the owner a bargaining advantage over the nonowner.

As we shall see, this improves investment incentives for the owner while worsening them for the tenant. The optimal assignment of ownership takes into account how important is the investment decisions of each party and how severe is the holdup problem from having each party not owning the asset. The term holdup here refers to the fact that the owner can limit the value of an investor’s input to the project by firing him *ex post*.

To illustrate these arguments more precisely, assume that e_A and e_B are observable but nonverifiable. The last of these assumptions implies that a court could not enforce stipulated effort levels as it would be impossible to verify whether they were implemented. Thus investment levels are noncontractible *ex ante*. The two parties are assumed to bargain over the *ex post* surplus once it has been created.

Suppose first that party A is owner. Then at the bargaining stage, he has the right to fire B. Let \bar{u}_i^j denote the disagreement payoff or outside option of i when j is the owner. We assume that even if A fires B at the bargaining stage, he can still make some use of the results of B’s investments. Specifically, a fraction λ of the investment remains to be exploited by A in B’s absence. It is useful to think of λ as measuring the extent of asset specificity. In the model of the previous section where e_B would be generic “effort” then $\lambda = 1$. However, where there is something special about B’s human

capital which requires his continued involvement in the project to make the most of it, then $\lambda < 1$.

Putting this together, the outside options of the two parties are $\bar{u}_A^A = a\sqrt{e_A} + \lambda b\sqrt{e_B}$ and $\bar{u}_B^A = \bar{u}_B$ where \bar{u}_B is the exogenously given level of the disagreement payoff of B. Using the symmetric Nash bargaining formula,²² the *ex post* payoff of A is:

$$\frac{1}{2}(a\sqrt{e_A} + b\sqrt{e_B}) + \frac{1}{2}(\bar{u}_A^A - \bar{u}_B^A), \quad (41)$$

which simplifies to

$$a\sqrt{e_A} + \frac{1}{2}(1 + \lambda)b\sqrt{e_B} - \frac{\bar{u}_B}{2}.$$

Similarly, the *ex post* payoff of B is

$$\frac{1}{2}(a\sqrt{e_A} + b\sqrt{e_B}) + \frac{1}{2}(\bar{u}_B^A - \bar{u}_A^A), \quad (42)$$

which in turn simplifies to

$$\frac{1}{2}(1 - \lambda)b\sqrt{e_B} + \frac{\bar{u}_B}{2}.$$

The two parties will choose e_A and e_B at the *ex ante* stage anticipating the *ex post* payoffs derived above. As a result the optimal choice of these variables are

$$\hat{e}_A^A = \frac{a^2}{4} \quad \text{and} \quad \hat{e}_B^A = \frac{b^2(1 - \lambda)^2}{16}. \quad (43)$$

This yields a second-best net expected surplus of

$$\hat{S}^A = \frac{a^2}{4} + \frac{b^2}{16}(1 - \lambda)(3 + \lambda). \quad (44)$$

This is less than the first-best surplus S^* .²³ Since B anticipates that, after the investments are made, he will be at the mercy of A, he invests less than the first-best level. The higher is λ , the less costly it is for A to fire B and the greater is the incentive problem of B. However, if $\lambda = 1$, there is full exploitation of B's output and he does not invest at all. If we think of λ as representing the extent of specialized skills, then economies

with greater skill intensity will suffer a smaller efficiency loss through this effect than those which only have generic labor input.

There are symmetric expressions if B is the owner. We now use $\mu \in [0, 1]$ to be the investment specificity parameter analogous to λ . By a similar analysis we find:

$$\hat{e}_A^B = \frac{a^2}{16}(1 - \mu^2) \quad \text{and} \quad \hat{e}_B^B = \frac{b^2}{4}. \quad (45)$$

Second-best surplus (also less than S^*) is:

$$\hat{S}^B = \frac{a^2}{16}(1 - \mu)(3 + \mu) + \frac{b^2}{4}. \quad (46)$$

As before, a larger value of μ induces a greater efficiency loss, all else equal.

We can now which party should own the asset to maximize economic efficiency (measured by total surplus) as a function of the key parameters: a , b , λ , and μ . Comparing Eqs. (44) and (46), we find that A should own the asset if $a^2(1 + \mu)^2 > b^2(1 + \lambda)^2$, while B should own it otherwise. We state this finding as:

Result 6. *If the marginal return of A's (B's) investment is greater than that of B's (A's) or his investment is more asset-specific than B's (A's), under the efficient assignment of property rights A (B) should own the asset.*

This theory of the “optimal” allocation of property rights can be thought of as reflecting two dimensions of the skill of the investors. The parameters a and b reflect their relative productivities as investors with a presumption that the most productive should own the asset. But the specificity of their skills matters too. If one investor has a very specialized skill so that replacing him would lead to a major loss in output, then it is best that he owns the asset. If not, the investment process is more prone to hold up. So if one party supplies generic effort which stays with the project whether or not he leaves, he will generally not optimally be the owner.

These ideas apply to thinking about ownership structures in agriculture in developing countries where landlords and tenants both have skills that can play a role in improving the land. The land should optimally be sold off to the tenant if the latter is more productive and has more land specific skills than the landlord. We now consider tenancy issues in more detail in light of this insight.

2.3.2 Role of tenancy

The model in the last section predicts that tenancy is an efficient arrangement when the landlord has high productivity and a high level of asset specificity. But in many contexts, the first of does not seem *prima facie* reasonable. The persistence of tenancy would then seem more plausibly due to the fact that credit market imperfections

prevent the transfer of the land to the tenant. There could also be other benefits to holding land such as linked benefits in the form of patronage or political power which make the land more valuable to the landlord and mean that he would always outbid the tenant for the land in an auction.

To explore this, we will suppose that A is the landlord, but $a = 0$ so that B should optimally be the owner. The value of the asset when A is the owner is $\hat{S}^A = (b^2/16)(1 - \lambda)(3 + \lambda)$ and since B's outside option is \bar{u}_B , A's payoff is $\hat{S}^A - \bar{u}_B$ and B's payoff is \bar{u}_B . If B is the owner, the value of the asset is $\hat{S}^B = b^2/4$. If B had the ability to make up-front payments, there are gains from trade. For example, if the transfer price p is set at $\hat{S}^B - \hat{S}^A$ then A is strictly better off and B is no worse off when ownership is transferred to B. But if B has no liquid funds, then this transfer will not take place. The arrangement that prevails will then resemble a share tenancy where B gets a 50% share of output.

It might be possible for a third-party (a bank) to enable B to buy the asset. To keep things simple, suppose that the interest rate is normalized to zero and B will simply have to pay back p to the lender. The problem now is that B will be in the same situation *vis-a-vis* the bank as he was previously *vis-a-vis* the landlord. So there is no gain in transferring ownership to a different unproductive party.

A land reform that transferred ownership to party B would now raise productivity. In fact, this is true for any land reform that dilutes the landlord's rights. To see this, suppose that with probability $\tau \in [0, 1]$ the tenant (party B) will acquire the land. This is similar to the way that we modeled attenuated property rights in Section 2.2.1. Now with probability $(1 - \tau)$, B's *ex post* payoff is as before, that is, $\frac{1}{2}(1 - \lambda)b\sqrt{e_B} + \frac{1}{2}\bar{u}_B$ but with probability τ it is $b\sqrt{e_B}$ as A has been expropriated via the land reform. Now

$$\hat{e}_B^A = \left\{ \frac{1}{2}(1 - \tau)(1 - \lambda) + \tau \right\} \frac{b^2}{4}. \quad (47)$$

When $\tau = 1$ this coincides with the outcome under pure B-ownership and when $\tau = 0$ it coincides with the outcome under A-ownership. Party B's investment is increasing in τ . We now have:

Result 7. *In the presence of frictions that prevent the efficient allocation of property rights, transferring property rights to the tenant will increase efficiency. Greater security of property rights for the initial owner now reduces efficiency.*

This result underpins the classic argument for forcible land redistribution toward tenants. That insecure property rights of one party (here the landlord) may enhance productivity is an application of the theory of the second best. Given that ownership is inefficient due to imperfect capital markets, a second distortion (imperfectly enforced property rights) can be efficiency enhancing. This result relates to the large literature on

tenancy showing that redistributive reforms such as land reform or tenancy reform might improve productivity and that the standard efficiency-equity trade off need not apply in all cases.²⁴

There is a link between this analysis of optimal property rights and the discussion of insecure property rights in the previous section. The producer in the benchmark model above could be thought of as the “rightful” owner (from the efficiency point of view) and the insecurity in the form of τ as therefore arising out of an inefficient allocation of property rights. Since the coercive authority cannot commit not to expropriate the producer, there is an *ex post* holdup problem and as a result of this the producer only gets a fraction of the share of the results of his investment. It is therefore inefficient if the coercive authority ends up owning the land. In Section 3 we will examine this issue in greater detail.²⁵

2.3.3 Communal property rights

The model developed so far looks solely at individualistic property rights. But it is often argued that communal property rights can, under some circumstances, be superior (see, e.g., Platteau, 2000). One way to think about communal property rights is that they maximize joint surplus because consumption is shared among members of the community. In that case, by assumption, communal tenure will always achieve the first best. This is not entirely plausible since the evidence of communal property rights does not provide unambiguous support to this view. For example, the well documented increases in agricultural productivity in China after switching to a household responsibility system seems to go against this finding.²⁶ A more promising approach would be to examine under what circumstances communal property might achieve greater efficiency than individual property rights.

In the above framework, communal property rights are best thought of as joint ownership in the sense that, if there is a disagreement at the bargaining stage, then production cannot go ahead. In other words, both parties have veto power (this is how joint ownership is modeled in Hart, 1995), that is, the disagreement payoffs of both parties are zero. In this case, it would seem likely that the holdup problem would be worse than with either party owning the land individually.

To examine this formally, observe that, using the symmetric Nash bargaining formula, the *ex post* payoff of both A and B is now

$$\frac{1}{2}(a\sqrt{e_A} + b\sqrt{e_B}). \quad (48)$$

Then the investment levels are $\hat{e}_A^J = a^2/16$ and $\hat{e}_B^J = b^2/16$, and the second-best net expected surplus is

$$\hat{S}^J = \frac{3a^2}{16} + \frac{3b^2}{16}. \quad (49)$$

It is straightforward to see that this level of surplus is less than both \hat{S}^A and \hat{S}^B . In other words, joint ownership is dominated by individual ownership.²⁷ As conjectured, joint ownership exacerbates the holdup problem.

But this negative conclusion on the merits of communal property rights depends on the output being a purely *private* good. To see this, suppose instead that the good produced is public so that, even if one of the parties is excluded by the owner, he is still able to enjoy some of the benefits. This might be, for example, because there are features of the asset that are enjoyed in common. More generally, any investments that improve the quality of an asset might spillover in part to neighbors.

The following extension of the model to public goods is based on Besley and Ghatak (2001) which extends the property-rights approach to the case of pure public goods. They show that joint ownership may dominate private ownership in this case.

To illustrate the argument, consider the following simplified version of the above model. Suppose that $a = 0$ implying that, if this was a private good, then B should be the owner. Suppose that the output $b\sqrt{e_B}$ is now a pure public good and θ_A and θ_B are the valuations of that good of parties A and B. The joint-surplus maximizing level of investment is now given by:

$$\hat{e}_B^* = \arg \max_{e_B} \{(\theta_A + \theta_B)b\sqrt{e_B} - e_B\} = \frac{b^2(\theta_A + \theta_B)^2}{4}. \quad (50)$$

A key distinction from the private good case is as follows. Since the output is a pure public good, then even if bargaining breaks down, the owner cannot exclude the other party from enjoying the benefit of it. Therefore, under B ownership $\bar{u}_A^B = \theta_A b\sqrt{e_B}$ and $\bar{u}_B^B = \theta_B b\sqrt{e_B}$.²⁸

As a consequence, the choice of investment is given by $\hat{e}_B^B = b^2\theta_B^2/4$. Under A-ownership $\bar{u}_A^B = \theta_A \lambda b\sqrt{e_B}$ and $\bar{u}_B^B = \theta_B \lambda b\sqrt{e_B}$. Hence, the choice of investment is given by

$$\hat{e}_B^A = \frac{b^2}{4} \left\{ \theta_B \frac{(1+\lambda)}{2} + \theta_A \frac{(1-\lambda)}{2} \right\}^2.$$

In contrast, under joint ownership $\bar{u}_A^J = \bar{u}_B^J = 0$ and the choice of investment is given by

$$\hat{e}_B^J = \frac{b^2}{4} \left(\frac{\theta_B + \theta_A}{2} \right)^2.$$

If $\theta_A > \theta_B$ then both ownership by A and joint ownership dominates ownership by B. This contrasts with the case of a purely private good. Moreover, if $\lambda > 0$, joint ownership dominates ownership by A.

We summarize this result as:

Result 8. *When the output produced with the asset is a public good, then communal property rights (joint ownership) may sometimes be optimal.*

The intuition behind this result is simple. Joint ownership “ties down” the two parties to the project and hence minimizes free-riding which is a problem for the provision of public goods.

Property rights allocation in the case of partly private and partly public goods has not been investigated much in the literature. However, it does seem relevant for understanding some forms of organization, especially in the context of communal assets such as condominium housing arrangements. This analysis suggests that, in general, the greater is the public good component in production, the more likely joint ownership will dominate individual ownership.

2.4 Evidence

This theoretical analysis naturally gives way to thinking about how property rights affect resource allocation in practice. There is now a significant literature which looks at this.²⁹ However, it is fairly rare to link the empirical analysis closely to the theoretical channels that we have analyzed so far.

One issue is what outcome to focus on. In a reduced form sense, all of the theoretical channels identified above would suggest a link between the level of output and property rights. In all cases, the level of investments, in the stylized model e , is (weakly) higher when property rights are more secure. However, as we showed in the example of guard labor, there can also be a reallocation of effort to or from more productive activities.

The two trade channels are quite specific in the way that they suggest that improved property rights will have an impact. In the first case, we should see a deepening in rental or sale markets for assets. In the second, we should see more use of credit among those whose property rights to collateralizable assets are improved. To investigate these ideas empirically requires going beyond looking solely at the effects on output, although we would expect output to be higher in both cases too.

One further issue concerns the level of aggregation. Our theoretical examples focused on a specific producer with fixed characteristics. These models mostly predict that the effect of improved property rights will be heterogeneous. To illustrate, consider the basic freedom from expropriation argument. In this case:

$$\frac{\partial e^*}{\partial \tau} = -\frac{(1-\tau)A^2}{2}. \quad (51)$$

This implies that factors that make A heterogeneous across producers such as wealth, access to other inputs and/or markets will tend to affect the marginal effect of an improvement in property rights. Such heterogeneous effects are a natural consequence of bringing theoretical considerations to bear on the analysis of the data.

We might also expect macroeconomic and microeconomic impact effects to be different in so far as the former capture general equilibrium responses to improvements in property rights. The overall macroeffect can mask many underlying mechanisms as emphasized here.

Another issue in bringing these ideas to the data concerns how to capture property rights. Our simple theoretical parameter, τ , masks a whole range of possibilities. In microdata, it is frequently possible to be quite precise about the claims that people have to their assets. For example, some asset ownership is backed by officially recognized and registered title deeds. However, other property is held more informally. A good example is the case of land rights in Ghana where land rights are granted by tribal authorities. Moreover, the rights to each plot of land are quite heterogeneous. In the data used in Besley (1995), rights can be decomposed into the different components—buying, selling, renting, leasing, and pledging.

The key issue whether in micro- or macrodata is how to identify the causal effect of changes in property rights on investment or productivity. Macroevidence tends to look at countries as units of analysis, sometimes regions within countries. Microevidence looks at the effect of property rights using data on firms and/or households. The core empirical approach is to run some kind of regression of the form:

$$\gamma_{it} = \alpha + \beta r_{it} + \gamma x_{it} + \varepsilon_{it}, \quad (52)$$

where γ_{it} is a measure of an outcome for cross-sectional unit i at date t , r_{it} is a measure of property rights and x_{it} are appropriate controls and ε_{it} is an error term.

In the basic case, there is no time dimension to this kind of analysis and the effect of property rights on outcomes is driven entirely by the fact that some firms or households appear to have better access to rights than others at a point in time. This raises quite difficult issues in estimating β . Omitted variables could be driving a simple correlation between the two: for example, better governance could be driving both secure property rights and a more investment-friendly environment. The other issue is that of reverse causality: investment itself could affect the nature of property rights.

In principle, either of these problems could be dealt with using instrumental variables, that is, finding a determinant of r_{it} which is not also a determinant of the decision of interest x_{it} . This is the approach of Acemoglu, Johnson, and Robinson (2001) in their study of cross-sectional country differences in property rights. They argue that settler mortality drives expropriation risk without having any direct impact on modern

day income *per capita*. In general, however, it is difficult to find convincing instruments even in microdata.

In some cases, there are changes in rights over time and space which allow researchers to explore the implications of changes in rights before and after with an explicit time dimension. Whether this succeeds in dealing with the issues of omitted variables and reverse causation is moot. This still depends on how the rights are allocated to households or firms. There may be scope for finding ways of explicitly modeling the political and economic forces that shape r_{it} .

Another route is to exploit variation between rights “within” firms or households. Thus, Besley (1995) exploits the fact that households in Ghana enjoy different rights on different plots of land that they farm and is able to look at how economic decisions vary across plots. This means that variation in household characteristics that affect the power of households to enhance their rights is not spuriously driving the relationship between economic outcomes and property rights.

Either way, this brief discussion emphasizes the need to understand the reasons why property rights differ in different times and places. This is something that we turn to in Section 3.

It is possible to take a bird’s eye view of the quality of property rights using cross-country data. To illustrate, we take two measures of property rights regimes using standard sources. The first is a measure of the security of property rights from the International Country Risk Guide (ICRG). It is measured on a scale between 0 and 10. A higher score corresponds to better protection of property rights. Figure 1 shows that this score is positively correlated with income *per capita* in the year 2000. In other words, countries with a higher risk of expropriation have lower levels of income *per capita*.

The second measure comes from the World Bank doing business project (<http://www.doingbusiness.org>). We focus on a measure of the ease with which individuals can register their property, specifically the country’s rank on this measure for 172 countries. This is a purely administrative dimension to property rights and follows the logic of the de Soto argument discussed in Section 2.2.2. Figure 2 shows that this too is strongly negatively correlated with income *per capita* in 2000. Thus, this more administrative dimension of property rights is weaker in low-income countries.

Together these figures illustrate the central proposition that improving property rights is associated with economic development. However, they say nothing about the direction of causation.

The correlation in Figure 1 is intriguing and forms the basis of the well-known empirical analysis of Acemoglu, Johnson, and Robinson (2001) who argue that this relationship is indeed causal. To this end, they use the mortality rates of colonial settlers as an instrument for property rights showing that the negative relationship between the ICRG expropriation risk measure and income *per capita* remains. Acemoglu and

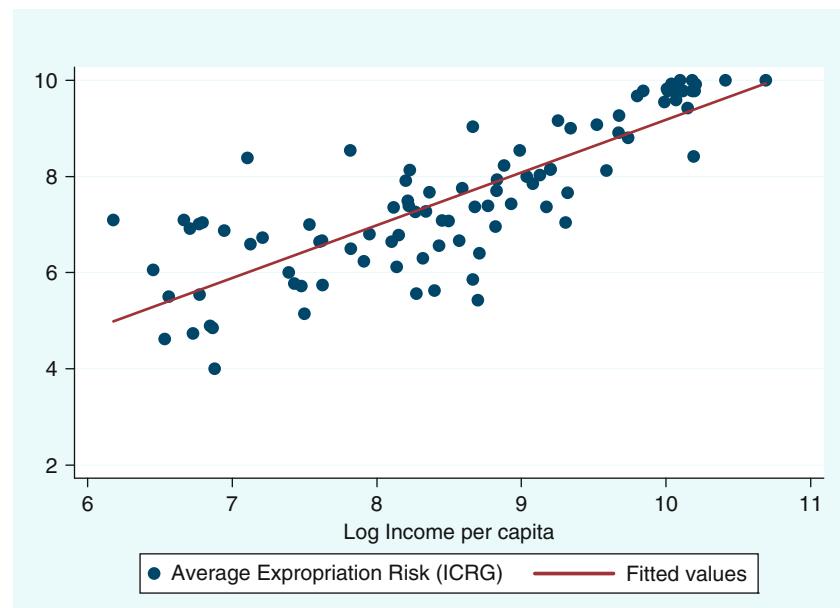


Figure 1 Expropriation risk and income *per capita*.

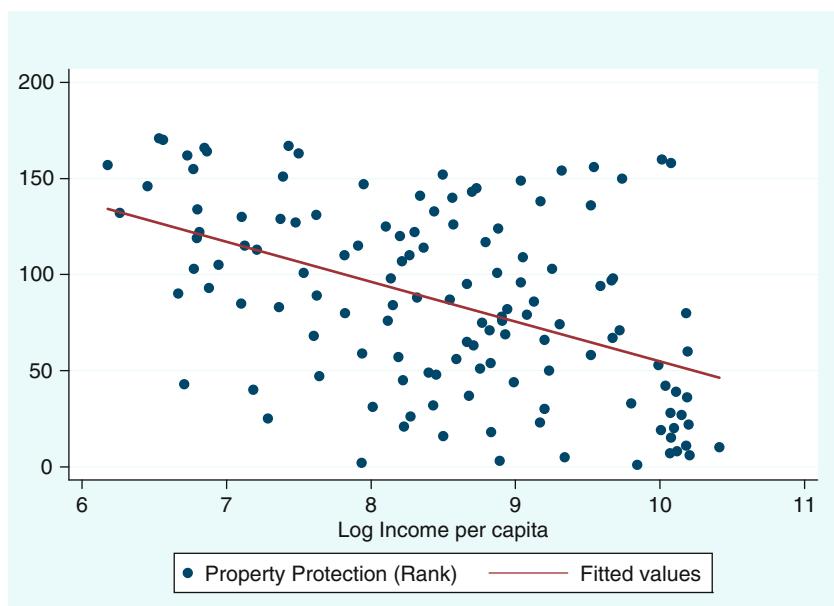


Figure 2 Property registration and income *per capita*.

Johnson (2005) look at two different dimensions of property rights and how they affect growth—expropriation risk and contract enforcement. Their aim is to assess which is more important in affecting aggregate output. They use the identity of the colonial power as an instrument for contracting institutions and settler mortality as the instrument for expropriation risk. On this base, they argue that only expropriation risk holds up as causal factor in affecting income *per capita* while the contracting environment affects the form of financial intermediation. In a related contribution, Glaeser, La Porta, Lopez-de-Silanes, and Shleifer (2004) argue that human capital could be a key missing variable in this kind of analysis, jointly determining both institutions and growth. However, given the scarcity of plausible instruments, it is clearly difficult to be able to identify between many competing potential causal pathways in cross-country data.

There are several microempirical studies that look directly at the question of whether secure property rights improve investment incentives. Here we provide a brief and selective review of the key findings. Besley (1995) in his study of property rights in Ghana, mentioned above, exploit the variation in the rights that individuals enjoy on different fields to test whether property rights matter for investment decisions. Ghana is in a transition between a traditional system of land rights (which emphasizes claims of the community) and a modern one (which emphasizes the claims of the individual and grants ability to transfer the land without needing a community sanction). The study focuses on self-reported transfer rights: whether each field owned and operated by a household has any of these rights is measured in the data, along with whether exercising this right requires lineage approval. In his study of the cocoa growing region, Wassa, in the west of the country, where the investment decision is the decision to plant trees, he finds that controlling for household fixed effects, investment is increased by better land rights. The study takes into account the potential problem of reverse causality: investment decisions (e.g., planting trees) could affect security of tenure as well. The basic result holds if land rights are instrumented with field level characteristics (soil quality, distance from house, investments already made when land acquired). As to which mechanisms linking property rights to investment are at work, this study is unable to find strong support for any particular mechanism, but on the whole, the support is the weakest for the collateral-based view.

In a more recent study on Ghana, Goldstein and Udry (2008) exploit the variation in security of tenure within the system of informal property rights administered by the local political system. They find that those cultivators without political power (e.g., those who do not hold any form of local political office) are less confident of their rights. Compared to those who hold political office, they leave their land fallow for significantly shorter duration (for fear that the land will be allocated to someone else), resulting in significant loss in profits per unit of land.

In a related study Field (2007) finds that property titles issued in Peru starting in the mid-nineties led to a significant increase in labor supply by urban slum dwellers. The

study looks at the effect of the program undertaken by Peruvian government that issued property titles to 1.2 million urban households during the 1990s on labor supply. While it does not directly look at the effect of investment, a key mechanism postulated in the chapter is that secure property rights reduced the need for guard labor and this freed up labor time that could be efficiently supplied in the labor market. In a related paper Field (2005) looks directly at investment and shows that residential investment also went up significantly, using a similar identification strategy and retrospective data on housing construction.

A more recent study (Hornbeck, 2008) shows that the introduction of barbed wire fencing to the American Plains in the late nineteenth century led to significant increases in the value of farmland, the productivity and production share of crops most in need of protection. Farmers were required to build fences to secure their land. From 1880 to 1900, the introduction and universal adoption of barbed wire reduced the cost of fencing, relative to wooden fences, most in counties with the least woodland. Over that period, counties with the least woodland experienced significant agricultural development and according to this study, this appears to reflect increased security of property rights due to barbed wire fencing.

Galiani and Schargrodsky's study (2005) is one of several studies that look at the collateral effect of property rights reform. It focuses on urban squatters in Argentina. Given that the allocation of titles was unlikely to have been random, they exploit a data-set which permits a cleaner identification strategy. They look at a group of squatters who occupied an area of wasteland in the outskirts of Buenos Aires more than 20 years ago from the time of the study. The area was composed of different tracts of land, each with a different legal owner. An expropriation law was subsequently passed, ordering the transfer of the land from the original owners to the state in exchange for a monetary compensation, with the purpose of entitling it to the squatters. However, only some of the original legal owners surrendered the land. The parcels located on the ceded tracts were transferred to the squatters with legal titles that secured the property of the parcels. Other original owners, instead, are still disputing the government compensation. As a result, a group of squatters obtained formal land rights, while others are currently living in the occupied parcels without paying rent, but without legal titles. Both groups share the same household pretreatment characteristics. Moreover, they live next to each other, and the parcels they inhabit are identical. Since the decision of the original owners of accepting or disputing the expropriation payment was orthogonal to the squatter characteristics, the allocation of property rights is exogenous in equations describing the behavior of the occupants. This assumes that this decision is orthogonal to land quality which seems reasonable in their context.³⁰

They find significant effects on housing investment, household size, and child education. The quality of the houses is substantially higher in the titled parcels. They only find modest effects on access to credit markets as a result of entitlement, and no

improvement in labor market performance. This is not surprising, as squatters could not transfer the property parcels for the first 10 years. They do compare early and late treatment households and find that 4% of the early treatment group received a mortgage loan. Their conjecture is that this small effect could be driven by difficulty of foreclosure on default.

Another study by Field and Torero (2006) looks at an urban land titling program in Peru. Their data allow them to directly observe whether loan applicants are requested to provide collateral. As a result they can isolate the effect of property titles on credit supply from their effect on demand by comparing loan approval rates when titles are requested to rates when they are not. Their results indicate that property titles are associated with increase in approval rates on public sector loans by as much as 12% when titles are requested by lenders. But they find no relationship between titles and approval decisions otherwise. In contrast, there is no evidence that titles increase the likelihood of receiving credit from private sector banks, although interest rates are significantly lower for titled applicants regardless of whether collateral was requested.

One explanation for this failure is that titling programs reduce banks' perceptions of their ability to foreclose. This is supported by data from Peru indicating that individuals with title have less fear of losing property in cases of default. Also, in Peru (and other comparable developing countries) even the middle-level propertied classes do not find it easy to receive credit. For example, in Peru a minimum of 2 years of tenure in a formal sector job and a high wage is a prerequisite for receiving loans from the formal sector. Therefore, it is not surprising that the urban squatters did not experience a huge increase in credit supply.

Another possibility is that de Soto essentially assumes that the binding constraint is always finance (which one can obtain by pledging collateral). But if a producer is in a low-return environment, because of either other shortcomings in the institutional environment or market failures, collateral is not going to do much good.³¹

However, more encouraging evidence is provided by Wang (2008) who looks at a housing reform in China that allowed state employees who were renting state-owned housing to buy their homes at subsidized prices. She finds that the reform increased the ability of individuals to finance entrepreneurial ventures by allowing them to capitalize on the value of the property.

The implications of weak property rights have been studied using microdata on firms. An interesting study along these lines is Johnson, McMillan, and Woodruff (2002) which uses a survey of firms in postcommunist countries. Their data exploit variation across firms and from different country institutional environments. They find that weak property rights do discourage firms from reinvesting their profits, even when bank loans are available. Where property rights are relatively strong, firms reinvest their profits. However, weak property rights appear to deter entrepreneurs from investing from their retained earnings.

3. ENDOGENOUS PROPERTY RIGHTS

So why might property rights protection be weak? While historically, nonstate actors have played a key role in the creation and enforcement of rights, in the modern world weak property rights boil down to problems in the way that the state functions. There are three types of state failure that are relevant to understanding this: predatory states, anarchic states, and ineffective states.

Predatory states are strong states that cannot find ways of limiting their own power. Anarchic states are those where there is no single authority—as when war lords and mafiosi retain coercive power.³² The power to enforce or violate property rights is therefore fragmented.

Ineffective states are those which, although they may have established a monopoly of force within a certain domain, have not invested sufficiently in relevant market supporting public goods such as courts and property registries. The first problem (predation) is an issue when the state is strong while the latter two (anarchy and ineffectiveness) are characteristic of weak states.

To date, most of the existing literature has focused on the first two problems. For example, Djankov, Glaeser, La Porta, Lopez-de-Silanes, and Shleifer (2003) pose the dilemma of effective government as finding the right balance between the problem of predation due to excessive state authority and anarchy due to weak state institutions. We will argue in Section 3.2 that understanding the forces that shape incentives to invest in market supporting property rights is also important.

3.1 Expropriation

There has been much discussion of aspects of state expropriation and arguments to the effect that limiting coercive power of the state is an important historical feature of state and market development (see, e.g., North, 1990). A classic reference is North and Weingast (1989) who argued that a decisive point in the history of state development in England came after the Glorious Revolution which limited the arbitrary power of the King subordinating his ability to raise taxes to Parliament. The need to limit state power and hence protect property rights is also at the heart of Acemoglu, Johnson, and Robinson's interpretation (2001) of why states with low rates of settler mortality built more effective states with more respect for private property.

Models of states' incentives to exploit arbitrary power include Grossman and Kim (1995), Grossman and Noh (1994), Moselle and Polak (2001), and Olson (1993).³³ It might be tempting to conclude that the problem of excessive state power is only a feature only of models where the government is controlled by a self-interested ruler intent on extracting resources from its citizens. However, as Kydland and Prescott (1977) shows, it may be optimal to limit state power even when the government is benevolent if it cannot commit to a future policy.

Expropriations by government are a fact of historical experience as illustrated in Table 1 which gives some examples over 700 years of human history. It begins with the expropriation of the Knights Templar by Philip IV of France in 1307. Thereafter, we find other regular examples of sovereign power being used to seize assets. Henry VIII dissolved the monasteries in England and took over the land beginning in 1536. The US government now widely regarded as paragon for upholding property rights

Table 1 Expropriations in selected countries

France, 1307-1312	From 1307 onward, Philip IV of France seized assets of the Knight Templars to alleviate France's serious financial problems. Five years later, Pope Clement V, pressured by the French King, abolished the order and transferred part of their assets to the Hospitallers. The value of the seized assets were likely significant as the Knights Templars operated the first known international banking network using their military strongholds
England, 1536-1541	Between 1536 and 1541 King Henry VIII expropriated monasteries in England and took over their land, amounting by some estimates to over 30% of the land holdings in England at that time. Some of the land and buildings went into the ownership of the crown, others were sold to the gentry. One of the unwanted side effects of the expropriation was the creation of a powerful upper class in England which became a serious restriction on the King's policies thereafter
United States, 1870-1910	Expropriation was commonly used as an instrument of public policy, designed to subsidize private enterprises in Railway construction, Milling and Mining. Expropriations and legislation supporting them were common in the United States at that time. Colorado's constitutional convention of 1875-1876, for example, stated that private property might be taken for "private ways of necessity, . . . reservoirs, drains, flumes, or ditches on or across lands of others, for mining, milling, domestic, or sanitary purposes"
Mexico, 1936-1938	As part of President Cardenas six year plan, the Mexican government expropriated direct investments in Agriculture, Railway and particularly Petroleum. Estimates of the Brookings institution put the value of expropriated properties belonging to US citizens to

Table 1 Expropriations in selected countries—Cont'd

	over \$ 300 million. The reaction was an Anglo-American boycott, decreasing oil exports and eventually a devaluation of the peso. Foreign direct investment contracted by more than two thirds between 1935 and 1940
Iran, 1951	British oil production facilities in Iran were expropriated under Prime Minister Mohammad Massadeq. In response, Britain boycotted Iranian oil depriving the country from its largest market. Supported by the United States and England, pro-monarchy forces toppled the government in 1953
Egypt, 1956	Following a withdrawal of American and British finance for the Aswan Dam, President Nasser nationalized the Suez Canal in 1956. The nationalization was the trigger to an armed conflict in the region including Isreal, France, and Britain and the occupation of Egypt
Cuba, 1959-1960	Following the Cuban Revolution in 1958, the Cuban government seized properties belonging to US nationals with an estimated value of \$1.8 billion. This was a higher amount than the total amount expropriated by all other Communist countries combined. The sectors most affected were public services, sugar growing and milling, and oil refining. The US government sanctions Cuba from the 1960s onward. Disruption in trade was enormous as trade links with the United States were close. The World Bank estimates that real GDP <i>per capita</i> was steadily decreasing for at least a decade after the expropriation
Algeria, 1971-1980	In 1971, President Boumediene inaugurated the “agrarian revolution”—a large-scale land reform which aimed at partial redistribution and nationalization of large land holdings. Absentee landowners were to be entirely expropriated. While a significant share of landowners avoided redistribution, over 1.3 million hectares had been distributed to nearly 100,000 beneficiaries by 1980. In the period following the reforms, productivity dropped drastically and agrarian production suffered. Partial reversal of reforms in the 1980s could not prevent heavy dependence on imports and rising food prices followed by social unrest

Continued

Table 1 Expropriations in selected countries—Cont'd

Chile, 1971	Led by President Allende, the Chilean government expropriated US copper mining companies of assets worth more than \$500 million. Expropriations were carried out through a constitutional amendment approved by the Chilean Congress in 1971. The case triggered conflict with the US companies and government and was followed by withdrawal of credit
Zimbabwe, 2000–2001	When the political mood seemed to swing against him in February 2000, Zimbabwe's president Mugabe launched a program of land redistribution. The relatively chaotic program was spearheaded by his party's paramilitary wing who began occupying white owned farms around the country. The land was taken, divided, and sold or given to peasants and party supporters. In the years following the expropriation, Zimbabwe's economy featured negative growth and rising levels of inflation. The country has increasingly become dependant on food aid

Sources: Baklanoff (1975), Barber (1994), Martin (2004), Rajan and Zingales (2003), Searingen (1990), Scheiber (1973), and World Development Indicators.

used expropriation as an instrument of public policy from 1870 onward to promote railway construction, milling, and mining.

Natural resources have also figured in government expropriations of the past. President Cardenas of Mexico expropriated, among other things, petroleum assets in the 1930s. Iran expropriated oil production facilities in 1951. The Cuban revolution in 1959–1960 also resulted in considerable expropriation of private assets by the state. Expropriation of land was the particular objective of the Algerian government between 1971 and 1980. In 1971, the Chilean government expropriated US copper mining assets reputed to be worth more than \$500 million at the time. Such expropriations continue to the present day as evidenced by the Zimbabwean government's program of forceful land redistribution since 2000.

All of these examples underline the contemporary and historical relevance of the material that follows.

3.1.1 Framework

We begin with a simple framework that includes the benchmark model of production from Section 2.

There are N identical producers. Each producer produces $x(\tau) \equiv \gamma(\tau) + z$ where $\gamma(\tau)$ is produced output, z is nonproduced output (e.g., natural resources), and τ is the rate of expropriation. The profit function of a producer is:

$$\pi(\tau) = \max_{e \geq 0} \{(1 - \tau)(A\sqrt{e} + z) - e\} \quad (53)$$

with optimal effort $e(\tau) = [(1 - \tau)A/2]^2$ and expected output $\gamma(\tau) = (1 - \tau)A^2/2$ as in Section 2.1.

We now suppose that there are M coercive authorities who together determine τ —the rate of expropriation. At this level of abstraction, we can think of such authorities in quite broad terms as states, powerful landlords, feudal barons, or roving bandits.

A coercive authority is distinguished by having access to a technology for expropriating the output of producers. Let T_j be the resources committed to expropriation by authority j . Collectively these actions determine the level of expropriation experienced by producers. We model this for the moment as a common resource problem with the aggregate expropriation rate being

$$\tau = \sum_{j=1}^M T_j, \quad (54)$$

the *sum* of the actions of the expropriating groups.³⁴ This is a very specific technology and one could consider others. But it serves to fix ideas about some of the basic issues that arise when studying the equilibrium level of expropriation.

Each coercive authority tries to capture a share of total output. We assume that the expropriation rate is common across both types of output and all producers, and hence expropriation cannot be targeted to specific production activities.

Aggregate output is simply N times output *per capita* and is denoted by:

$$X(\tau) = N[\gamma(\tau) + z]. \quad (55)$$

This is clearly decreasing in τ . For future reference, the aggregate produced output and natural resource output are $Y(\tau) \equiv N\gamma(\tau)$ and $Z \equiv Nz$, respectively. We now explore the determination of τ under different assumptions to investigate the kinds of factors that will lead to different levels of equilibrium expropriation.

3.1.2 Commitment

We start by assuming that the coercive authorities can commit to an expropriation level up front. This is built into the following timing structure:

1. Coercive authorities choose T_1, \dots, T_M
2. Producers put in their effort e
3. Output is realized and expropriation takes place

We now look at the equilibrium of expropriation as determined by a Nash equilibrium between the coercive authorities. These could be thought of as “roving bandits” in the sense of Olson (1993). The history of Europe is replete with marauding groups such as the Vikings and the Magyars that plundered from whatever sources they could find in the absence of a strong sovereign. We assume that expropriation is costly and let αN be the (constant) marginal cost of expropriation. We have made this increasing in N so that having a large group of producers to expropriate is more costly. We have also assumed that each producer has the same expropriation cost. This can be motivated by supposing that αN is equal to an outside wage determined by some kind of productive activity in which organizers of coercive authorities can otherwise engage. Having a more productive economy will then make expropriation more expensive.

The payoff of the j th coercive authority is:

$$T_j(X(\tau) - \alpha N) = NT_j(\gamma(\tau) + z - \alpha). \quad (56)$$

Thus the “profit” of a coercive authority comes from the outside wage rate that determines the opportunity cost of expropriation. The trade off for coercive authorities is quite standard. An increase in expropriation increases profits assuming that $\gamma(\tau) + z > \alpha$, but leads to each producer lowering his effort. The optimal rate of expropriation balances these two factors. A necessary condition for a positive rate of expropriation is that $(A^2/2) + z > \alpha$. This says that there have to be sufficient resources to plunder relative to the cost of expropriating. We assume that this is the case from now onward.

The Nash equilibrium in expropriation levels has all coercive authorities choosing T_j simultaneously. This yields first-order conditions, assuming an interior solution, of:

$$T_j \gamma'(\tau) + \gamma(\tau) + z = \frac{-T_j A^2 + (1 - \tau) A^2}{2} + z = \alpha. \quad (57)$$

Since all authorities have an identical expropriation technology, it is natural to focus on a symmetric outcome. It is straightforward to show that the overall expropriation rate τ , assuming an interior solution, is then given by:

$$\tau = \left(1 + \frac{2(z - \alpha)}{A^2}\right) \frac{M}{M + 1}. \quad (58)$$

There are three immediate comparative static results that are straightforwardly derived for this simple problem and will serve to organize our thinking about the organization

of coercion. These will look at how the level of expropriation varies with the number of coercive authorities (M), the level of natural resources per producer (z), and the cost of coercion (α).

Our first result comes from seeing how τ depends on M —the number of coercive authorities. Totally differentiating Eq. (58) yields

$$\frac{\partial \tau}{\partial M} = \left(1 + \frac{2(z - \alpha)}{A^2} \right) \frac{1}{(M + 1)^2} > 0. \quad (59)$$

This yields:

Result 9 (Monopoly of force). *Output is highest and expropriation lowest when there is a monopoly on coercive authority.*

This result follows from the observation that competitive determination of expropriation rates creates a commons problem. Each coercive authority fails to internalize the effect of its expropriation decision on others. States that are fragmented, that is, where coercive authority is wide spread will tend to be poorer according to this logic. This corresponds to the kind of fragmentation that is frequently referred to in the context of weak states—see, for example, the discussion in Acemoglu (2005).

This result goes back a long way. It underpinned Hobbes' concept of Leviathan and Weber's concept of the state in which a single state authority monopolizes the power to coerce. The simple model that we have setup shows that this has a rationale in terms of efficient organization of production. Olson (1993) puts this point colorfully as follows:

In a world of roving banditry there is little or no incentive to produce or accumulate anything that may be stolen and, thus, little for bandits to steal. Bandit rationality, accordingly induces the bandit leader to seize a given domain . . . and to provide a peaceful order . . . thereby obtaining more in tax theft than he could in migratory plunder. (p. 568)

Just how monopoly of coercion can be achieved is not clear. One could think in terms of a creating a state with the power to prevent all other actors in the economy from exercising coercive power. This certainly represents the situation that we see in many advanced states in the world. But the monopoly outcome could also be achieved by finding some kind of Coasian arrangement among those who possess the power to expropriate. These authorities could, in principle, bargain with one another to achieve the monopoly outcome and then use transfers among each other to achieve the cooperative outcome. However, practical experience suggests that states that have broken down find such cooperative outcomes quite difficult to sustain.

One way to achieve an outcome equivalent to monopoly expropriation is via a system of monopoly franchises (chieftains). A good example is the Zamindari system of land taxation in India where powerful landowners were given the power to expropriate

from particular tenants. The “franchise” arrangement can be thought as defining property rights by the coercive authorities. Let N^j be the group of producers assigned by such property rights to the j th coercive authority. Then the optimal expropriation decision maximizes:

$$T_j [N^j(\gamma(T_j) + z)] - \alpha N^j T_j. \quad (60)$$

This effectively achieves the Coasian outcome among the chieftains by assigning property rights to the coercive authorities, thereby overcoming the common pool problem.³⁵ But there is an issue of how such rights are defined and enforced in the absence of some kind of super coercive authority. In the case of the Zamindars in India, it was the British who used them as agents of colonial rule.

If there are many coercive authorities, then one issue is whether competition between them works as a further restraint on expropriation. This will happen only if there is mobility of producers across coercive domains. One feature of many low income economies is that such mobility is either naturally or artificially limited. Moreover, we would expect coercive authorities to strategically limit mobility in order to maintain the power to expropriate resources. Thus in many parts of Africa, systems of land tenure and passage of land across generations are set up to reduce mobility. This has a short-term logic for those who operate such systems. However, in a dynamic perspective, it clearly has a cost if expropriation levels are too high.

Our second result looks at how expropriation varies with the extent of nonproduced output (natural resources). Here, it is straightforward to see that:

$$\frac{\partial \tau}{\partial z} = \frac{2}{A^2 M + 1} M > 0. \quad (61)$$

This result can be interpreted as follows:

Result 10 (Resource curse). *A higher level of non-produced output leads to more expropriation and hence less output overall.*

This result is driven by the fact that such expropriation in this case does not create any disincentive effect. It mirrors a wide variety of empirical findings suggesting that resource richer countries find it difficult to establish regimes in which expropriation is limited (see, e.g., Mehlum, Moene, & Torvik, 2006 or Sachs & Warner, 2001).

Finally, we can look at the effect of an increase in α . This has at least two possible interpretations. One sees it as reflecting improvement in outside productive options among those who have the power to coerce. This would be the case in a more productive economy where wages are higher. The other is an improvement in systems of formal property rights protection, for example, by undertaking reforms of legal protection with an independent judiciary to protect the rights of producers. This would make

exercising private coercive power more costly. Differentiating Eq. (58) with respect to α yields:

$$\frac{\partial \tau}{\partial \alpha} = -\frac{2}{A^2} \frac{M}{M+1} < 0. \quad (62)$$

This yields:

Result 11. *An increase in the cost of coercion and/or the benefits to non-coercive activities increases produced output and reduces expropriation.*

This result ties our model to some general equilibrium approaches to rent-seeking in which coercive activity is affected by the level of economic development in general. Many authors have argued that a key role of institutions is to set the relative reward structures for different kinds of economic activities. Baumol (1990) and Murphy, Shleifer, and Vishny (1991) have argued that entrepreneurial talent can be reallocated toward rent-seeking and organized crime when the returns to such activities are high relative to producing. Even in advanced democracies, these authors emphasize that the legal system can be a device for organized rent-seeking which reduces production.³⁶

3.1.3 No commitment

The simplest way to capture the inability to commit in this model is to suppose that τ is chosen after the effort decision by the producers. Throughout, we study the case of a monopoly coercive authority and, for simplicity, set $\alpha = 0$. The timing of moves that we consider is:

1. Producers put in their effort e .
2. The coercive authority chooses τ .
3. Output is realized and expropriation takes place.

We consider a subgame perfect Nash equilibrium in this game. It is straightforward to see that in a one-shot setting we have:

$$\tau = 1 \quad \text{and} \quad e = 0. \quad (63)$$

This yields the obvious but important insight:

Result 12. *Without commitment in a static setting, the level of expropriation is one-hundred percent and produced output is zero.*

The logic is simple: for any $\gamma(\tau) \geq 0$, the coercive authority will set $\tau = 1$.³⁷ Anticipating this, the producers will commit no effort and produced output is zero.

This fits well the idea of a state that is “overstrong against thyself” following De Long (2000) who quotes the poet John Milton in this context. The outcome in this equilibrium is Pareto inefficient for low enough z . This is because there is a level of expropriation $\tau < 1$ which makes both producers and the coercive authority better

off from an *ex ante* point of view. The question is how to solve the commitment problem so that Pareto gains can be reaped. Unless noted otherwise, to make things as stark as possible we focus on the case where $z = 0$. We now explore five ways in which an outcome with $\tau < 1$ can be obtained: reputation, exit, secrecy, ownership, and voice.

3.1.4 Reputation

The fact that coercive authorities have an incentive to develop a reputation for restricting the use of their expropriative activity is a central theme of the literature on security of property. This has been applied to the problem of limiting state power by Grossman and Noh (1994) among others. We illustrate this in the conventional way—thinking of producers and the coercive authority as being in a long-run relationship. This means that the producers can “punish” the coercive authority for expropriating them excessively by ceasing to produce for some specified period. But for long-run relationships to work to secure property in the way that this suggests, it must be that there is some long-run entity called the state that can take a far-sighted view. Olson (1993) characterizes the state in such contexts as a “stationary bandit.” The fact that the state is stationary means that it is able to take a long-term view. He describes this idea as follows:

A stationary bandit will therefore reap the maximum harvest in taxes . . . only if he is taking an indefinitely long view and only if his subjects have total confidence that their “rights” to private property . . . will be permanently respected. (p. 571).

One feature that separates weakly and strongly institutionalized political systems is the extent to which they have long-lived political institutions that can be used to sustain reputational outcomes. For example, strongly institutionalized settings often have parties with long-term political ambitions and hence an incentive to build reputations. Olson (1993) emphasizes that this desire for the longest possible time-horizon was embodied in the familiar refrain “long live the King.” Just how relevant these ideas are in practice, is moot. Clearly forward-looking behavior has to apply across generations of politicians. Moreover, the data suggest that, if anything polities with long-lived politicians and parties holding office tend to have less secure property rights.

To study the role of reputation in the simplest possible way, suppose that there is an infinitely repeated interaction between the coercive authority and the producers. There is production at each date and the coercive authority chooses τ at each date. In such situations, the coercive authority can be punished by producers if it chooses to expropriate them more than promised.

We solve the ensuing infinitely repeated game by supposing that producers and the coercive authority use simple history-dependent stationary trigger strategies whereby producers set $e = 0$ after any history of play in which a coercive authority sets $\tau = 1$. Let the promised expropriation rate be $\hat{\tau}^*$. We assume that producers set $e = 0$ after any history of play in which the coercive authority chooses $\hat{\tau}^* < \tau \leq 1$. We assume

that the coercive authority discounts the future with discount factor $\beta \in [0, 1]$. This could be interpreted in the standard way as a part of preferences or it could be thought of as representing a “political” discount rate reflecting how likely will be the turnover of the current government. We will study strategies of expropriation that are credible in the sense that if a coercive authority promises such a rate, it will be in its interest to honor that promise. Hence, along the equilibrium path, there will not be any expropriation beyond the promised level and producers will commit effort levels consistent with this.

To see what levels of expropriation are sustainable, consider the “value function” of the coercive authority after it has deviated to maximal expropriation, that is, set $\tau = 1$, following a “promised” expropriation rate of $\hat{\tau}^*$. As noted before, for simplicity we set $z = 0$; since producers are assumed not to affect the flow of nonproduced revenue, their behavior cannot change depending on the actions of the coercive authority, on or off the equilibrium path. We will however comment later on how the results are affected when $z > 0$. The discounted expected payoff of the coercive authority following a deviation is:

$$\overline{V}(\hat{\tau}) = Y(\hat{\tau}). \quad (64)$$

It represents the fact that the coercive authority seizes all of the output and producers respond by setting $e = 0$ ever after. This represents a rather crude expropriation technology where the coercive authority is only able to expropriate everyone or no-one. We discuss what happens when producers can be treated differently below.

If the coercive authority has not “cheated” by reverting to $\tau = 1$, then its “value function” along the equilibrium path with an expropriation rate of $\hat{\tau}$ is:

$$\hat{V}(\hat{\tau}) = \frac{\hat{\tau}Y(\hat{\tau})}{1 - \beta}. \quad (65)$$

An expropriation level $\hat{\tau}$ is credible if it does not pay to deviate to $\tau = 1$. This will be the case if:

$$\hat{V}(\hat{\tau}) \geq \overline{V}(\hat{\tau}). \quad (66)$$

From this, we conclude that *an expropriation rate $\hat{\tau}^*$ is credible only if*

$$\hat{\tau}^* \geq (1 - \beta). \quad (67)$$

This expression says that promised expropriation has to be *high enough* to be credible. If the expropriation rate is low, then production will be high enough to tempt the coercive authority to maximally expropriate the producers. This result implies that a very

patient coercive authority (β close to one) can commit to any rate of expropriation and be credible whereas an impatient coercive authority (β close to zero) will only be able to commit to a high rate of expropriation.

Credibility is a binding issue for a coercive authority when it cannot promise an expropriation rate that maximizes its payoff. To see this, observe first that the payoff maximizing expropriation rate for the coercive authority in this context is:

$$\tau^* = \arg \max_{\tau \geq 0} \{\tau Y(\tau)\} = \frac{1}{2}. \quad (68)$$

The optimal *credible* expropriation rate maximizes the coercive authority's payoff subject to the credibility constraint, that is,

$$\begin{aligned} \hat{\tau}^* &= \arg \max_{\substack{\tau \geq 0 \\ \text{subject to}}} \{\tau X(\tau)\} \\ &\tau \geq (1 - \beta). \end{aligned} \quad (69)$$

If the credibility constraint is not binding, then $\hat{\tau}^* = 1/2$. Otherwise, $\hat{\tau}^* = (1 - \beta) > 1/2$. Thus the per-period payoff to the coercive authority in a credible equilibrium is:

$$\hat{R} = \begin{cases} (1 - \beta)Y(1 - \beta) & \text{if } \beta < \frac{1}{2} \\ \frac{1}{2}Y\left(\frac{1}{2}\right) & \text{otherwise.} \end{cases} \quad (70)$$

This logic associates higher output and lower levels of expropriation with long-lived forms of coercive authority. Thus, as claimed by Olson, reputation acquired by stable autocrats (such as hereditary monarchies) may perform better than short-lived leaders. The key insight from this analysis is summarized as:

Result 13. *The credible rate of expropriation supported by reputation is characterized by $\hat{\tau}^* = \max\{\frac{1}{2}, 1 - \beta\}$ and is therefore decreasing in the discount rate of the coercive authority.*

We have so far abstracted from the role of natural (unproduced) resources and their implications for the ability of a coercive authority to commit. However, it is straightforward to extend the above analysis to the case where $Z > 0$. In

this case, $\bar{V}(\hat{\tau}) = Y(\hat{\tau}) + Z + \frac{\beta}{1 - \beta}Z$ and $\hat{V}(\hat{\tau}) = \hat{\tau}(Y(\hat{\tau}) + Z)/(1 - \beta)$. This yields

$\hat{\tau} \geq (1 - \beta) + \beta \frac{z}{z + \gamma(\hat{\tau})}$ implying that the share of national income that is earned from natural resources, $z/(z + \gamma(\hat{\tau}))$, affects the ability to commit. The higher that share,

then the higher is the credible expropriation rate, that is, the more difficult it is for the government to commit to a low level of expropriation. This reinforces the resource curse finding from the previous section. The presence of natural resources also affects the optimal expropriation rate. This is now $\tau^* = \frac{A^2 + 2z}{2A^2} \geq \frac{1}{2}$ which is increasing in z , the value of “natural resources” per producer.

The study by Guriev, Kolotilin, and Sonin (2008) of nationalizations in the oil industry around the world over the period 1960–2002 resonates with our analysis. They find that nationalization is more likely to happen when oil prices are high and the quality of institutions is low even when controlling for country fixed effects. When oil prices are high the temptation to expropriate is high (in terms of our model, this can be thought of as a high value of Z).

This analysis assumes that there is a single expropriation rate for all production whether it requires producer effort or not. However, it should be noted that if the government can separate out property rights to produced output and natural resources it will wish to do so. Moreover, the reputational mechanism studied here cannot explain how property rights over natural resources would ever emerge in equilibrium. But given that we observe protection of such property rights in the real world when it is fairly transparent that there are pure rents to be earned by the state, we need an alternative explanation than the kind of dynamic reputational model studied here.

More generally, developing a reputation as a means of enforcing property rights best fits with situations where there is no institutionalized restriction on coercive power. Thus, if it applies at all, this analysis is probably most relevant to some weakly institutionalized polities where there are no workable formal rules to limit coercion. The main lesson from history is that, if government is to turn over on a regular basis, then there is a need to move beyond personal reputations as a means of sustaining property rights protection. Thus, we now move onto understanding other ways of trying to limit coercive authority.

3.1.5 Exit

Another way of preventing coercive power being abused is the possibility that producers can exercise an exit option. Exit could take the form of leaving to take an outside option denoted by a utility level \bar{u} or the ability to hide or leave with a fraction of output which we denote by μ .

This can make a difference to the expropriation level and hence output even when the coercive authority can commit. To see this, recall that in our benchmark static model where $z = 0$, then $y(\tau) = (1 - \tau)A^2/2$. Without any constraints, the coercive authority would choose $\tau = 1/2$. However, so long as the producer's payoff when $\tau = 1/2$ is less than \bar{u} this is no longer feasible. In particular, for $\bar{u} \geq [A/4]^2$ the exit option will be a binding constraint and so the maximum feasible level of τ will be

set by $[(1 - \tau)A/2]^2 = \bar{u}$. A similar result obtains when we allow the producer to leave with a fraction $1 - \mu$ of his output with $\mu > 1/2$.

More generally, we can study the no commitment case when the producer can hide, destroy, or carry away a fraction $1 - \mu$ of his output when threatened by expropriation. There is now a maximum tax rate denoted by $\bar{\tau} = 1 - \mu$. This affects the outcome described in the last section in two main ways. First, after a deviation from the promised level of expropriation, the government can only capture $\bar{\tau}Y(\hat{\tau})$ rather than all of the output. Second, the upper bound on expropriation may also apply to the expropriation rate along the equilibrium path which we have labeled $\hat{\tau}$. Hence, it also provides an upper bound on expropriation.

The condition for a credible equilibrium level of expropriation when exit imposes a constraint on expropriation is:

$$\bar{\tau} \geq \hat{\tau}^* \geq \bar{\tau}(1 - \beta). \quad (71)$$

This shows exactly how exit can permit the government to credibly commit to less expropriation. The maximum expropriation rate reduces the lower bound on $\hat{\tau}^*$. This makes it more likely that the coercive authority can commit to τ^* .

This analysis shows why a government might try to institutionalize exit options. One way to do this would be through decentralization where multiple governments compete and it is possible to leave one jurisdiction and move to another if expropriation is too high. Qian and Weingast (1997) refer to this as “market preserving federalism” which they argue has been relevant as a device to limit expropriation risk in the context of China.

From the perspective of the coercive authority, any exit rate which satisfies $\bar{\tau} \geq \hat{\tau}^* \geq \bar{\tau}(1 - \beta)$ would be optimal. Such exit options permit the authority to commit to the expropriation rate that maximizes its *ex ante* payoff.

However, national income and the welfare of producers would still be higher were it possible to increase exit above the level associated with τ^* . Nonetheless, a purely predatory government would not have an incentive to protect property rights above the level $1 - \tau^*$. But arguably there are governments around the world that protect property rights to a point where the state has more or less dispensed with predation. Thus, we need to consider other explanations of the behavior of such states.

3.1.6 Secrecy

The commitment problem that leads to full expropriation is based on the assumption that the coercive authority can costlessly observe output. If this assumption is relaxed, it is possible for the commitment problem to be mitigated. To make this point as simply as possible, we consider a variant of our benchmark model. Let e now be discrete: $e \in \{0, 1\}$. If $e = 0$ then output is y_0 . Otherwise, output is \bar{y} with probability p , and y

with probability $1 - p$, where $\bar{y} > \underline{y} > y_0$. Let $\hat{y} \equiv p\bar{y} + (1 - p)\underline{y}$ denote expected output when $e = 1$.

To set $e = 1$, the producer incurs a cost c . Both effort and output are unobservable to the authority. However, by incurring a cost of γ the authority can observe output. The producer, in contrast, knows the level of realized output, but only after the effort decision is taken. If $\gamma = 0$ then output is costlessly observable. Then *ex post* it is always worthwhile for the authority to observe output and as above, it will always set $\tau = 1$. As a result, the producer will select $e = 0$ and no output will be produced.

As in the previous section, suppose $1 - \mu$ is the part of the output that is lost to the authority because the producer has some margins of choice. Our interpretation here is that producers can actually carry away and not just destroy a fraction $1 - \mu$ of the output. For $\mu < 1$, the producer will set $e = 0$ under the assumption that c is sufficiently high:

$$c \geq (1 - \mu)(\hat{y} - y_0). \quad (72)$$

Therefore, the authority can gain an amount $R = \mu y_0$ from expropriation.

Suppose instead that $\gamma > 0$. Now whether to observe output is a matter of choice to the coercive authority. Suppose $\mu y_0 \geq \gamma$ so that for any output level it is worthwhile for the coercive authority to expropriate the producer if it wishes to do so. Would it be worthwhile for the authority to demand a flat amount t and observe and expropriate output only if the producer does not oblige, rather than always observing and expropriating output? We can interpret this as a fee in exchange for the promise not to expropriate. Is it possible that this would give the producers the incentive to choose $e = 1$ and the authority to partly capture the gains through the flat fee? There are two sets of incentive constraints. First, producers must prefer paying the fee to being audited and having their output expropriated. Second, the coercive authority must prefer accepting the fee from a willing producer to auditing and expropriating.

The producers' incentive constraints are:

$$\bar{y} - t \geq (1 - \mu)\bar{y}, \quad \underline{y} - t \geq (1 - \mu)\underline{y}. \quad (73)$$

The corresponding constraint for the coercive authority is:

$$t \geq \mu\bar{y} - \gamma, \quad t \geq \mu\underline{y} - \gamma. \quad (74)$$

As at the time when it decides whether or not to audit, the level of output is unknown to the coercive authority, the above two expressions can be replaced by a single one, namely:

$$t \geq \mu\hat{y} - \gamma. \quad (75)$$

Combining the relevant constraints yields the following condition for the incentive compatible fee:

$$\mu\hat{y} - \gamma \leq t \leq \mu\underline{y}. \quad (76)$$

Since the authority would prefer as high a fee as possible, then $t = \mu\underline{y}$.³⁸ Notice that if γ is small, the incentive constraint cannot be satisfied; the temptation to audit and expropriate is too high for the coercive authority.

We have to make sure that this fee is consistent with the producers' incentives to choose $e = 1$ which is the case so long as (using the fact that $t = \mu\underline{y}$):

$$\hat{y} - (1 - \mu)\gamma_0 - \mu\underline{y} \geq c. \quad (77)$$

Notice that now all parties are strictly better off.³⁹ Therefore, making it feasible for its citizens to legally conceal some of their output can be Pareto-improving in this setting.

Some aspects of secrecy are enshrined in the relations between citizens and government in most modern democracies. While such democracies may now have the ability to commit not to expropriate in other ways, this may also reflect historical circumstance in which the evolution of these values about personal freedom were, in the first instance, efficiency enhancing and allowed the flourishing of the market economy. A good example of a modern day autocratic regime that has embraced some aspects of this in its pursuit of developing a market economy is China. Bai, Li, Qian, and Wang (2004) have emphasized the use of anonymous banking laws in China as a means of limiting the possibility of expropriation.

3.1.7 Public ownership?

We have so far assumed that all production remains in private hands. However, both historical and contemporary experiences suggest that we should take seriously the possibility of public ownership. Thus the state could nationalize all production and employ the citizens to work as wage laborers. In this case, all the surplus in production accrues to the state. Indeed, this is how expropriation has worked in socialist economies.

This section asks whether public ownership can solve the problem of imperfect private property rights. In this case, since the coercive authority is also the owner and residual claimant then perhaps all the problems that we have studied so far go away. This view would be correct if all effort into production were also put in by the coercive authority. However, that is not realistic. So the interesting case to study is where labor power remains private while land and other fixed assets are owned by the state.

If a socialist government could enforce the level of effort that maximizes output, which in our example is $e = A^2/4$ then national income would be maximal and socialism would be more productive than private ownership. Indeed, before it was

discredited by the experience of Russia and Eastern Europe, this was a frequent claim made by its proponents. But, as we shall see, this ignores the standard problem of how workers are to be motivated when they are not residual claimants.

One thing is clear immediately from our simple model. The question of the distribution of the surplus is a separate issue from efficiency if the state has the means to set $e = A^2/4$. The government may be the residual claimant or it could choose to leave the fruits of production in the hands of the workers. In the case of predatory government, public ownership without compensation would constitute complete but efficient predation now and forever.

But the socialist approach to predation breaks down if there are limited means for the government to induce effort. This will turn out to generate parallel problems to those that we have studied so far. To see this, suppose instead that effort e is private information. Then there is a moral hazard problem and the state has to offer citizens an incentive to work. The wage, w , is paid only if output is realized as a kind of "incentive payment."⁴⁰ Assume that the outside option of all workers is zero. In this case, the workers will set their effort at the level:

$$e^* = \arg \max_{e \geq 0} \{ \sqrt{ew} - e \} = \frac{w^2}{4}. \quad (78)$$

The government which now owns the land on which output is produced earns all of the expected surplus which is now:

$$\frac{w}{2}(A - w). \quad (79)$$

It is easy then to see that the optimal wage is $w^* = A/2$. The state and the worker share output equally in the event of a successful project.

It is now easy to see that there is a formal equivalence between what happens under socialization when the producers are paid a contingent wage and the case of predation under private ownership where $\tau \in [0,1]$ was the share taken of privately owned output. In the private ownership case, the coercive authority optimally sets $\tau^* = 1/2$ making the government's payoff under state and private ownership identical. Thus, in this setting, there is no way for public ownership to create either higher returns than a predatory state with private production or to increase output.

This symmetry between public and private ownership economies is maintained if we now introduce limited commitment under public and private ownership. Under socialism, the government would have an incentive to offer a wage of $A/2$ which it would then be tempted to set equal to zero after output is realized. As with expropriation of privately owned output, this would create an incentive constraint which would

make it impossible for the government to commit to the wage that was optimal from the point of view of offering incentives to workers. In the absence of commitment, the workers would set $e = 0$. The sustainable level of incentives under commitment problems would then be $w = \beta A$ which is obtained by equating $(w/2)(A - w)/(1 - \beta)$ and $\frac{w}{2}A$.⁴¹ We would again have an identical outcome under public and private ownership.

So are there reasons to believe that retaining productive assets in private hands is more efficient as appears to be the case in practice? One possibility is simply that socialist economies have had trouble acknowledging the need for incentives at all. Indeed, in many socialist regimes, there has been aversion to incentives because they create *ex post* inequality when output is stochastic. This would imply that output would tend to be lower than under private ownership provided that private ownership economies can limit government expropriation. Of course, government might then appeal to non-monetary incentives such as working to support the motherland, etc.

3.1.8 Voice

By far the most important development in government in the past two centuries has been the development of systems of representative government in which citizens have a say in how private property rights are supported. In this section, we will discuss how institutional arrangements that enhance the power over decision making of producers may affect property rights protection.

To study this as simply as possible, suppose that the policy maker puts a weight of λ on the utility of producers. This sidesteps the issue of modeling the micro detail of political institutions. This objective function of government could represent the outcome of some kind of probabilistic voting model along the lines laid out in Persson and Tabellini (2000) or a lobbying model along the lines of Grossman and Helpman (1994).⁴² Either way, this is consistent with the idea of a “property owning democracy” where owners of assets have a say in government policy. That said, this formulation might also crudely capture how preferences are aggregated in an autocratic setting. The mechanism could also mirror what would happen if a country were ruled by a more benevolent autocrat who was either fearful of an uprising or genuinely interested in the well-being of its citizens. In the limiting case where $\lambda = 1$ whoever controls the government values its own rents and the utility of producers equally.

The payoff of the coercive authority is now given by:

$$R(\tau) + \lambda N\pi(\tau) = \frac{N}{2} \left(\tau(1 - \tau)A^2 + \lambda \frac{(1 - \tau)^2 A^2}{2} \right). \quad (80)$$

In a static model without commitment, the government would now choose a level of expropriation of private production after effort e has been sunk that maximizes:

$$\tau(\lambda) = \arg \max_{\tau \geq 0} \{ [\tau + \lambda(1 - \tau)] \sqrt{eA} - \lambda e \} = 1 \quad (81)$$

as long as $\lambda \leq 1$. Thus, this formulation still leads to full expropriation in a static setting.⁴³

Observe that, in this case, the coercive authority would *like to* commit to the expropriation rate:

$$\tau^* = \frac{1 - \lambda}{2 - \lambda}, \quad (82)$$

which is decreasing in λ and equal to 0 for $\lambda = 1$. For $\lambda = 0$, we get the solution that we studied in Section 3.1.4.

In a dynamic model, the value of λ affects the level of property rights protection that the coercive authority can credibly commit to. To see this, we now repeat the logic from above and consider what fixed rate of expropriation, $\hat{\tau}^*$, along the equilibrium path is credible. The condition for no expropriation beyond the promised level to be credible becomes:

$$\frac{R(\hat{\tau}) + \lambda N\pi(\hat{\tau})}{1 - \beta} \geq Y(\hat{\tau}) - \lambda Ne(\hat{\tau}) = N \left(\left(1 - \lambda \frac{(1 - \hat{\tau})}{2} \right) \frac{(1 - \hat{\tau})A^2}{2} \right). \quad (83)$$

Solving this equation, we find that the expropriation rate is credible only if:

$$\hat{\tau}^* \geq 1 - \frac{\beta}{1 - \lambda + \beta\lambda/2}. \quad (84)$$

Observe that if $\lambda = 0$, then we are back to the condition in Eq. 5. It is clear from this that for all $\lambda \geq \hat{\lambda} = (1 - \beta)/(1 - \frac{\beta}{2})$ the coercive authority can credibly commit to any expropriation rate. Observe that $\hat{\lambda} < 1$ —so a democracy in which the government values its own rents and the payoff to citizens equally, that is, $\lambda = 1$, will always fully protect their property rights. For any fixed β , a higher value of λ increases the range of expropriation rates that the coercive authority can commit to *ex ante*. Thus, modeling voice in this way does relax the expropriation incentive constraint of government.

This argument illustrates in a very simple way why institutions that force decision makers to weight the welfare of producers can lead to a lower expropriation threat and hence increase output in the economy. It explains why it could be in the interests of a powerful autocrat who cannot commit to invest in institutions that reduce its power.

These theoretical findings are consistent with the crude cross-sectional observation that democracies tend to be richer than autocracies on average. However, it is clear that

there is likely to be a two-way relationship. Indeed, there is a large and growing literature on this issue.⁴⁴

There are a number of historical episodes that resonate with this. It can, for example, explain the gradual and peaceful transition toward democracy in the United Kingdom which was initiated through concessions toward voting and establishing independent legal institutions charged with protecting property rights. Previous models, such as Acemoglu and Robinson (2000), have discussed the possible role of revolutionary threats in franchise extension. The current approach emphasizes that there can be a pure self-interest motive if the threat of expropriation is too great. Only if λ is higher can the state commit to lower rates of expropriation. This could be particularly important when other institutional changes increase political turnover and hence reduce β . The model does suggest that there will be a limit on this mechanism when it is controlled by the coercive authority. It will only have an interest in improving institutions encouraging producer voice up to the point where it can implement its own *ex ante* optimal expropriation rate. This is supported by institutions of voice, λ^* , such that:

$$\frac{1 - \lambda^*}{2 - \lambda^*} = 1 - \frac{\beta}{1 - \lambda^* + \beta\lambda^*/2}. \quad (85)$$

For all $\lambda < \lambda^*$, increasing producer voice is Pareto-improving. However, for $\lambda \geq \lambda^*$, there is a conflict of interest between producers and the coercive authority.⁴⁵

3.1.9 Heterogeneous producers

The analysis so far assumes all producers are treated symmetrically. However, it is possible that property rights are protected differentially across various social and economic groups. There are a number of possible sources of heterogeneity suggested by the model so far. The logic of the analysis above suggests, for example, that producers with greater access to exit opportunities (lower $\bar{\tau}$) and with more influence (higher λ) will suffer less from expropriation.

However, in the context of the reputational mechanisms that we have explored, an important issue is how far a coercive authority can selectively expropriate one group without undermining the trust of others. This depends in part on information flows across groups. In particular, the question is to what extent one group gets to see any expropriation of the other group. If one group can be secretly expropriated, then that should not undermine the confidence of the other. This suggests the possible role of a “divide and rule” strategy.⁴⁶

To illustrate the power of this mechanism to limit expropriation, suppose that there are two groups $J \in \{A, B\}$: those with high exit options (A) and those with low

exit options (B). If the state deals with each group separately, then the credibility constraint is:

$$\hat{\tau}_J \geq (1 - \beta)\bar{\tau}_J, \text{ for } J = A, B. \quad (86)$$

Suppose now that a deviation on either group results in *both* groups believing that the coercive authority has cheated on its promise, and that henceforth it will expropriate both groups at a common rate. This implies that if the coercive authority chooses to maximally expropriate one group, it will always maximally expropriate the other group as well, since it will be punished by both groups anyway. This can work only if there are good information flows between the groups (e.g., via the media). Then the incentive constraint becomes:

$$\hat{\tau}[Y^A(\hat{\tau}) + Y^B(\hat{\tau})] \geq (1 - \beta)[\bar{\tau}_A Y^A(\hat{\tau}) + \bar{\tau}_B Y^B(\hat{\tau})] \quad (87)$$

or

$$\hat{\tau} \geq (1 - \beta) \left[\bar{\tau}_A \frac{N_A}{N} + \bar{\tau}_B \frac{N_B}{N} \right]. \quad (88)$$

This relaxes the incentive constraint of the coercive authority.⁴⁷ The intuition is simple: if it cheats on its promise to group B, in addition to producers in group B, the producers in group A punish it as well, thereby increasing the cost of cheating.

If the goal of the political system is to reduce the aggregate level of expropriation, this analysis suggests the importance of two sets of institutions. First, as we already mentioned, those that permit free flow of information; and second, solidarity mechanisms in which *any* kind of expropriation by the coercive authority is treated as if it is an expropriation of everyone.

3.1.10 Taxation and expropriation

In our discussion so far, we have not talked about taxation and how it relates to expropriation. To libertarians taxation is a form of expropriation.⁴⁸ *Ex post*, it might seem it is a matter of semantics as to whether taking away money from private citizens is called taxation or expropriation. Also, even from an *ex ante* point of view, in our framework where everyone is risk neutral, whether τ is a tax rate, or the expected probability of expropriation, seems equivalent so long as these are known before e is chosen.⁴⁹

If we look across countries of the world, there actually seems to be a positive correlation between protection of property rights (as measured using the ICRG expropriation risk measure) and the share of taxation in GDP from Baunsgaard and Keen (2005).

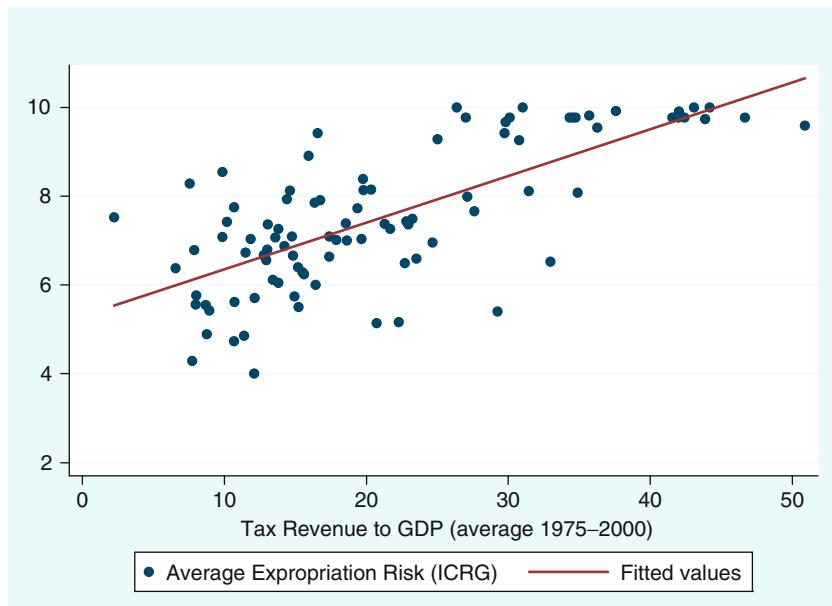


Figure 3 Expropriation risk and tax revenue.

This is illustrated in Figure 3, averaging tax revenue (as a percentage to GDP) between 1975 and 2000 and the ICRG expropriation risk score (scale 0–10) averaged between 1984 and 1997. As argued by Besley and Persson (2007), this type of pattern reflects the fact that countries with more developed fiscal systems tend to be richer and, on the whole, more market oriented. It brings into sharp relief the idea that expropriation of property (and not taxation) is symptomatic of a low level of development.

A key feature of tax systems is precisely that they lay down clearly defined rules and enable the citizens to determine their actions accordingly. Unlike expropriation, taxation is typically an organized and systematized form of extraction from citizens. Tax systems typically involve published codes according to which government tries to commit not to levy them *ex post* or to discriminate across producers on an *ad hoc* basis. A key issue is what enables the government to make this commitment. A natural starting point is the reputation-based story: we can equate expropriation with government seizing all of the output from the producers while taxation is the limited expropriation $\hat{\tau}^* < 1$. This distinction between expropriation and taxation is consistent with Weingast (1997) which argues that a feature of liberal political institutions (particularly democracy) is to create “fundamental transgressions”—lines in the sand which cannot be traversed without coordinated punishments. It emphasizes that there is need to have transparent access to information on expropriation by government to give the citizens a means to punish governments that violate their property rights.

There is another important difference between taxation and expropriation. Even if we allow for the possibility that taxes too can be subject to *ex post* changes, blurring the line between taxes and expropriation from the commitment point of view, an important issue is whether the producer can withdraw the asset from production. For example, in the case of land, it can lie fallow. However, a coercive authority that has the power to expropriate the owner of his assets, as well as any output that results from it, does not face this constraint. In principle, the coercive authority can force the asset to be brought into production. Using our previous arguments on reputation and exit options we can show that this means even with imperfect commitment, taxation will provide better incentives than when the coercive authority has the power to expropriate assets (in addition to output).

To see this formally, suppose there is a discrete decision denoted by $\delta \in \{0, 1\}$ which denotes whether the asset is used at all. We assume now that the production function is:

$$\delta[A\sqrt{e} + z]. \quad (89)$$

So if $\delta = 0$, nothing is produced, while if $\delta = 1$, then there is output of z even if $e = 0$. When the coercive authority cannot expropriate assets then, the “punishment” that the coercive authority faces if it takes away all the output from the producer is that output will be zero forever. However, if the coercive authority can expropriate assets, then output will be z each period. In the former case, the credible level of expropriation (of output) is $\hat{\tau} \geq 1 - \beta$. In the latter case, it is $\hat{\tau} \geq (1 - \beta) + \beta(z/(z + \gamma(\hat{\tau})))$, which is clearly higher than in the former case.

This argument is even stronger in modern economies given the importance of inalienable human capital (e.g., Grossman & Hart, 1986).⁵⁰ To illustrate this formally, suppose that in the event that the government sets $\tau = 1$, the producer can withdraw value equal to $1 - \mu$ from private production in the form of inalienable specific human capital. Then the maximum share of output that can be expropriated is equal to μ and, in the dynamic context, the incentive constraint for the government becomes more relaxed, namely, $\hat{\tau}^* \geq (1 - \beta)(1 - \mu)$.

3.1.11 Cross-sectional empirical regularities

The models that we have developed above give a crude sense of why, in cross-sectional regressions, we might find institutional and economic variables that can explain the extent of expropriation risk by government. Two things come rather directly out the analysis above. First, we should expect more stable governments with institutions that constrain government and enhance voice to have better property rights protection. Second, we should expect countries with high levels of natural resources to have weaker property rights protection.

Table 1 looks at these ideas using the ICRG expropriation risk variable averaged between 1984 and 1997 as the dependent variable. This was used to construct Figures 1 and 3 above. The variable is on a 0-10 scale with 10 denoting the highest level of property rights protection. Its mean is 7.3 and its standard deviation is 1.7.

We use three different institutional variables as independent variables in cross-country data for the year 1997.

Our first institutional measure is the extent of executive constraints as measured in the Polity IV data. This variable has a value of 7 when such constraints are strong. We create a dummy variable equal to one when countries score 7 and take the average value of this variable between 1945 and 1997. Its mean is 0.24 and its standard deviation is 0.29. The first column of Table 1 shows that this is negatively correlated with expropriation risk and is significant at 1%. A one standard deviation change in this measure of institutions leads to little less than half a standard deviation change in expropriation risk.

There is no direct measure of the government discount factor. However, Acemoglu, Johnson, and Robinson (2001) motivate their use of settler mortality to explain weak property rights use an argument in terms of incentives to set up long-term and short-term (extractive) institutions. As we have seen above, long-termism should lead to better property rights protection. The variable “settler mortality” has a mean of 245 and its standard deviation is 469. The second column of Table 1 shows the results for 64 former colonies for which this variable is available. As shown by Acemoglu, Johnson, and Robinson (2001), there is a negative correlation between settler mortality and expropriation risk (significant at 1%) with a one standard deviation change in settler mortality associated with a around one third of standard deviation change in expropriation risk in this subsample.⁵¹

Another likely factor that shapes short-termism in government behavior is the incidence of civil war. Civil war could also be a proxy for more fragmented political authority as war lords may gain control of some parts of a country. Here, we use a variable derived from the Correlates of War data base. It is the average number of years between 1945 and 1997 that a country has been engaged in a civil war. Its mean is 0.06 which says that in 6% of the country years on average in our sample are in civil conflict and its standard deviation is 0.12. Third column of Table 1 shows again there is a positive correlation between the prevalence of civil war and expropriation risk which is significant at 1%. In this case a one standard deviation change in the average civil war variable is associated with a around a third of a standard deviation change in expropriation risk.

To investigate the resource curse effect on property rights, we use a dummy variable which is equal to one if a country is an oil exporter. The mean is 0.12 and the standard deviation is 0.32. The fourth column of Table 1 shows that there is a positive correlation between expropriation risk and being an oil exporter which is significant at

10%. Being an oil exporter increases expropriation risk by half of one standard deviation.

The fifth column includes all these variables together on the sample of colonies for which settler mortality data is available. The correlations that we uncovered in the earlier columns remain significant with the exception of the oil exporter dummy variable.

While only suggestive, these findings support the relevance of the underlying theoretical reasoning that we explored throughout this section. The link between these reduced form correlations and specific theories is tenuous and one challenge for future empirical work in this area is to forge a closer link between theory and data.

3.2 Improving state effectiveness

Effective states in the current context are those that support institutions that allow households and firms to enjoy secure property rights. This constitutes a key investment in state legal capacity of the kind emphasized in recent work by Besley and Persson (2007, 2008). They formulate a dynamic model where such investments are forward looking and state capacity accumulates over time. For simplicity of exposition, we will focus here on a simpler static approach. This will emphasize how heterogeneous interests, and the way that these are mediated through political institutions, shape the decisions that states make to improve property rights protection.

To formalize this, we need a cost function which maps public expenditures on market supporting property rights into lower τ in the notation of this chapter. Concretely, these costs can be thought of as the resources needed to fund courts and property registries. We will black box this by simply writing this cost function as $L(1 - \tau)$ where $L(\cdot)$ is an increasing, convex function measuring the *per capita* cost of improving property rights at an economy wide level. This function could vary across countries due to different legal traditions and hence could constitute the way in which legal origins along the lines of La Porta., Lopez-de-Silanes, Shleifer, and Vishny (1998) enter the determination of market supporting property rights.⁵² For example, if common law countries are better at protecting asset owners from encroachment on their rights or in enforcing some kinds of contracts, then $\partial L(1 - \tau)/\partial \tau$ could be lower in these cases, that is, a lower marginal cost of delivering a given level of property rights protection.

Funding $L(1 - \tau)$ can be through either general taxation or a set of user fees, for example, paying to register property or to use the court system. Deciding on the method of finance is a societal decision which will have implications for the level of investment that is likely to take place. Also important for the decision to invest in the capacity to support effective property rights will be the way in which political institutions shape collective choice. This will matter most when there is significant heterogeneity in the population. The analysis so far has given many reasons why we would expect such heterogeneity in practice.

This section will lay out a rudimentary way of thinking about this and the insights that are aided by having the explicitly microfounded models from Section 2. However, it will be clear that improving state effectiveness in terms of property rights is not fundamentally different from any dimension of public spending that has an impact on productivity in the economy, such as extending education or building infrastructure. These also have to deal with issues that arise due to heterogeneous costs and benefits within the population.⁵³

Another theme in this section will be the interplay between formal and informal institutions in providing legal services needed to support property rights. One aspect of low income countries is the role of social networks in ensuring that property can be used as collateral or traded. For example, traditional land rights in Africa often require that the lineage or tribal authority has jurisdiction in this domain. This can be important since the incentives of the state to provide property rights could depend on such private informal alternatives. After all, if informal provision is effective then there may be no need for any kind of state investment. We can think of this in the way that we suggested in Section 2.2.2 where we supposed that there was value of τ_N that is specific to the network with a value of τ_F where “F” stands for “formal” which is relevant in the market. The latter could then be improved by collective investments represented by the function $L(1 - \tau_F)$.⁵⁴

In general, the case for state provision lies in extending the domain of trade. A legal system based on networks provides contract enforcement services only to members of that network creating a patchwork of different τ_N s. In principle, the formal legal system τ_F could be available to all households and firms in the economy creating a common basis for trade. To the extent that τ_F is lower this will tend to foster arms-length trade which will raise output.

3.2.1 A simple benchmark

To illustrate some of the issues involved in a simple and stylized way, we will work with the model of Section 2.2.1. While this was motivated as a model of expropriation risk rather than market supporting property rights we can use the interpretation that improving the court system allows better protection against claims by others to the fruits of the output. We will suppose that among the N producers there are different levels of productivity or land holding represented by A_i with $i = 1, \dots, N$. The payoff to typical producer is:

$$\pi(A_i, \tau) = \left[\frac{(1 - \tau)A_i}{2} \right]^2 + \bar{e}. \quad (90)$$

Suppose that each producer were faced with the *per capita* cost of sustaining property rights protection paid via a general lump-sum levy on all producers. Producer *i*'s preferred level of formal property rights protection would then be:

$$\tau^*(A_i) = \arg \max_{\tau \in [0,1]} \{\pi(A_i, \tau) - L(1 - \tau)\}. \quad (91)$$

The fact that A_i and τ are complements implies that $\tau^*(\cdot)$ will be a decreasing function, that is, more productive producers will prefer better property rights protection (lower τ).

To make a prediction about the level of property rights protection that would be chosen in this economy, we need to know whose preferences are decisive in collective choices over τ . As a benchmark, we will consider the outcome of a median voter model where those that benefit from weaker property rights are a negligible fraction of the population and hence do not influence the choice of policy.⁵⁵

Since preferences over τ are single-peaked, then a standard median voter model would predict that in a democracy, this will be the median value of A_i , denoted by A_m . Thus, according to this we will see $\tau^*(A_m)$.

In this simple case, the distribution of productivity in the economy (perhaps reflecting the underlying distribution of assets) would affect the location of the median voter and hence the level of formal property rights protection in the economy as a whole. This simple model could give a foundation as to why legal origins matter, if we think that it affects the function $L(\cdot)$. The median voter in a country with a lower marginal cost of property rights protection will tend to choose better property rights protection.

3.2.2 Extensions

Our median voter result is a useful benchmark result. However, there are reasons to be doubtful that it captures the full set of issues that are relevant to decisions to invest in property rights protection in reality. We will now discuss some departures from this benchmark and discuss how they may affect the result.

Many developing countries are not democracies and it is fanciful to think that the median producer could be most relevant for determining τ . Some form of elite rule where richer producers have more political power might seem more natural. If the elite were simply rich producers who use the formal legal system to protect their property rights, then we would expect giving them the right determining τ would actually increase property rights protection. So economies dominated by a producer elite may tend to toward higher output and greater income *per capita*.

However, another (and perhaps more plausible) interpretation of elite rule would be for the rentier class who live off the fruits of weak property rights have a say in political decision making. To make things simple, suppose that the rentiers act together in a unified way—a strong unified elite and do not face any cost from investing in the legal system (this remaining incident on the producers). Then, their payoff is:

$$\frac{\tau(1 - \tau)}{2} \sum_{i=1}^N (A_i)^2. \quad (92)$$

It is clear that their choice of property rights protection would be $\tau = 1/2$ whatever the cost of investing. Whether this is greater or less than would be desired by the median producer is not immediately clear. However, for small enough investment costs and a higher enough level of productivity for the median producer, we would expect there to be a conflict of interest with more producer oriented polities favoring stronger protection of property rights compared to those run by the rentier class. While not clear-cut, this does give some presumption for believing that property rights protection would be better in more democratic societies. These findings are in line with some of the discussions of Engerman and Sokoloff (2002) and their discussion of the role of inequality in shaping a variety of policy choices. However, as argued by Acemoglu and Robinson (2005), there is the possibility that fear of a revolution could lead to a policy choice by the elite that is closer to that of the wider citizens. Both of these observations are consistent with our earlier discussion of the impact of democracy on expropriation risk.

The existence of networks may also have implications for the political economy of property rights protection. Some groups, for example, may be able to enjoy good levels of property rights by trading in networks and hence without recourse to improving state effectiveness. Models of oligarchic property rights as developed by Acemoglu (2003) and Braguinsky and Myerson (2007) can be thought of in such terms. In a world where the ruling elite enjoy privileged access to a superior level of τ_N , there may be little incentive to improve property rights for the wider economy (see Sonin, 2003). However, Besley and Persson (2008) argue that the force of this argument depends on the development of the tax system, since the ruling elite would be better off maximizing production efficiency and then taxing the benefits for its own ends. This is essentially an application of the Diamond and Mirrlees (1971) efficiency theorem. This kind of network-based argument presents a somewhat different reason why fragmented authority can be damaging to property rights development.

3.2.3 Empirical regularities

The theoretical discussion gives a feel for why there should be difference in states' effectiveness in supporting property rights. We would expect this to depend on the level of economic development (measured via the A_s in terms of the theory), the political institutions and the legal history which affects the workings of legal institutions. To explore this in a very preliminary way, Table 2 presents some cross-sectional regressions based on data from the World Bank's Doing Business Web site (<http://www.doingbusiness.org>). The variable that we look at is the one of property registration which tries to get at how easy it is to register property under the law.⁵⁶ While this is only a partial perspective on the issues that interest us, it is certainly indicative and is available for 172 countries. The variable that we use is the summary ranking of each country's performance across three indicators: cost, number of procedures, and time.

Table 2 Cross-sectional determinants of expropriation risk

	(1)	(2)	(3)	(4)	(5)
High constraints on the executive (Polity IV)	−2.249*** (0.515)				−1.545** (0.644)
Settler mortality		−0.001*** (0.000)			−0.001** (0.000)
Average years in civil conflict 1945–1997 (COW)			−5.359*** (1.238)		−3.029** (1.409)
Oil exporter				−0.790* (0.453)	0.122 (0.494)
Observations	123	64	129	122	61
R-squared	0.154	0.105	0.149	0.022	0.265

Notes: Dependent variable is Average Expropriation Risk from the International Country Risk Guide for years 1984–1997. Robust standard errors in parentheses: *significant at 10%; **significant at 5%; ***significant at 1%.

In the first column of Table 2, we correlate this variable with the log of income *per capita* and find a strong negative correlation, that is, countries with higher income levels have better systems for registering property. Of course, the direction of causality is hard to know. A one standard deviation change in log income *per capita* is associated with around 20 places in the country ranking.

In the second column, we look at the correlation with legal origins, with French legal origin being the omitted category. There are strong correlations with Socialist; Scandinavian and German legal origin all of which have better rankings compared to French legal origin. These findings are in line with La Porta et al. (1998).

The third column of Table 2 looks at the correlation with democracy using the Polity data for the period from 1945 to 2000. The variable that we use measures the proportion of years in the year 2000 for which the country has had a polity score greater than zero over this period. Moving from being continuously autocratic to continuously democratic over this period is worth around 45 places in the ranking suggesting that policies that enable property registration are superior in democracies (Table 3).

Finally, in the fourth column, we include all of these sets of variables. Although the magnitudes of the correlations change, the basic findings are robust.

Table 3 Cross-sectional determinants of property registration

	(1)	(2)	(3)	(4)
Log income <i>per capita</i>	−20.790*** (3.303)			−10.648*** (4.023)
English legal origin		−11.387 (8.953)		−0.953 (8.982)
Socialist legal origin		−32.951*** (9.835)		−40.314*** (10.606)
Scandinavian legal origin		−87.751*** (7.332)		−46.793*** (13.234)
German legal origin		−66.984*** (8.955)		−32.330*** (11.619)
Proportion of years in democracy 1944–2000 (Polity)			−44.550*** (10.729)	−28.369** (12.621)
Observations	130	169	172	130
R-squared	0.225	0.162	0.094	0.349

Notes: Dependent variable is a country's rank (1–172) on the World Bank Doing Business Web site for time, number of procedures and cost of registering property. Legal origin omitted category is French legal origin. Robust standard errors in parentheses: *significant at 10%; **significant at 5%; ***significant at 1%.

As with the earlier results, these are only suggestive correlations and they in no sense test the theoretical models that we have put forward. However, they hint at the possible empirical relevance of the ideas that we have laid out and breathe life into the theory. But much remains to be done to refine our understanding of these issues at an empirical level in a way that is linked to theory.

4. CONCLUDING COMMENTS

Market economies rely on the creation and enforcement of property rights. In this chapter, we have reviewed the ways in which property rights affect economic decisions, and the incentives for creating effective institutions for the protection of private property.

Economists have often approached the problem of designing public policies by taking the starting point of market failure—typically where a competitive market fails to internalize externalities or there is a lack of competition. But the study of markets rests

often on assumptions about effective property rights which cannot be taken for granted. The role of competitive markets in allocating resources is then quite limited. This observation is especially relevant in the context of financial markets given the importance of assets in supporting trade.

The issues studied in this chapter are, we believe, of first-order importance in studying the development process and there is now a wide variety of supporting evidence. Looking at the evolution of property rights also means integrating insights from historical experience. This chapter has, we hope, given a sense of the richness of the issues that are involved in studying the interplay between property rights and economic development. It also has created a unified analytical framework drawing from the literature on development economics, political economy, and economics of contracts and organizations. However, it has not been possible to cover things in as much depth as we would have liked to, nor have we been able to provide a thorough survey of the literature.

One key message that stands out is how one should be cautious in thinking about property rights extension in a monolithic way. The creation of effective property rights is heterogeneous in its impact and there are many potential mechanisms that can sustain property rights. This suggests that there should not be a “one size fits all” mantra of extension of private property rights, nor a blind faith that this is a magic bullet that will cure all economic ills.

End Notes

*. Chapter for Volume V of the *Handbook of Development Economics* edited by Dani Rodrik and Mark Rosenzweig. We are grateful to Hannes Mueller for research assistance. We would also like to thank Madhav Aney, Oriana Bandiera, Avinash Dixit, Erica Field, Dilip Mookherjee, Gerard Padró i Miquel, Dani Rodrik, T. Paul Schultz, Ragnar Torvik, Chris Udry, and especially, Oliver Denk and Rocco Macchiavello for helpful comments. The usual disclaimer applies. Besley is grateful to the ESRC and CIFAR for financial support.

1. See Barzel (1997) and Alchian and Demsetz (1972).
2. See, for example, Acemoglu (2003).
3. One of the earliest advocates of private property rights was Aristotle who thought that property would be more likely to be looked after if someone owned it or profited from it than if it were treated as common (Robbins, 2000 p. 18).
4. Changes in technology or demand that lead to a rise in the value of the asset are argued to be key drivers of emergence of private property rights (see North & Thomas, 1973).
5. See de Soto (1989, pp. 158–163) for a detailed discussion.
6. There is an overlap with the issues covered here and Baland, Moene, and Robinson (2008).
7. In principle, this could even apply to labor, for example, “forced” labor.
8. The natural question is, given this deadweight loss why does this form of imperfection in property rights exist? We will examine this question in detail when we endogenize τ .

9. A necessary condition for the existence of labor markets is property rights in one's own labor, that is, absence of slavery or other forms of coercive use of labor. In this chapter, we will focus on property rights in nonhuman assets, such as land.
10. This is just an application of the separation property of agricultural household models: with complete markets, individual preferences do not affect production decisions (see Singh, Squire, & Strauss, 1986). This is taking prices as exogenous. Otherwise, household preferences will affect production decisions when prices are endogenous: in economies where people value leisure a lot, wages will be high and this will clearly affect production decisions.
11. Since $A \leq 2$, $\tau \in [0, 1]$, and $\gamma \in [0, 1]$ these solutions for effort are both interior.
12. For the proof of (i), observe that the sign of the derivative $\partial\sqrt{e_1}/\partial\tau$ depends on the sign on $\gamma^2 A^2 \tau (2 - \tau) - 4$. Now $\tau(2 - \tau)$ is increasing in τ and the maximum value that it can take is 1. Therefore, the maximum possible value of $\gamma^2 A^2 \tau (2 - \tau) - 4$ is $\gamma^2 A^2 - 4$ but by assumption $4 - (\tau\gamma A)^2 > 0$ for all τ , A , and γ . For (ii) observe that the sign of the derivative $\partial\sqrt{e_2}/\partial\tau$ depends on the sign on $4 + \tau^2 \gamma^2 A^2 - 8\tau$. Clearly, this is positive for low values of τ and negative for high values of τ . The proof of (iii) follows directly by applying the envelope theorem.
13. This follows from the assumptions that the asset that is subject to insecure property rights is also one where productive labor is used to generate income and the resource constraints are not binding.
14. See de Meza and Gould (1992) and Dixit (2004).
15. For example, landlords in India often leave their land fallow rather than leasing them out for fear of rights and control to tenants due the presence of tenancy laws that provide security of tenure to tenants and regulate rents. This prevents the land-poor from accessing land through tenancy and is viewed as an unintended negative consequence of the existing tenancy laws. See Hanstad, Haque, and Nielsen (2008).
16. There is now a large literature that focuses on the implications of credit constraints for the path of economic development. See, for example, Aghion and Bolton (1997) and Banerjee and Newman (1993).
17. Note that together these assumptions imply $\rho < 1$. Given that we have normalized the cost of effort to one, this is an assumption in the relative price of the input x relative to effort e .
18. In Besley and Ghatak (2008), we provide a more thorough microfoundation to this story using a costly state verification model.
19. In this case:

$$\pi = \sqrt{\frac{A^2}{4} + w(1 - \tau)} \left(A(1 + \Delta) - 2\sqrt{\frac{A^2}{4} + w(1 - \tau)} \right) + w(1 - \tau) - \rho.$$

20. Formally, the result follows by taking the derivative of r with respect to τ and observing that $(A^2/4) + w(1 - \tau) = e \leq 1$.
21. See Besley and Ghatak (2008) for further discussion of this.
22. This is standard in the literature following Grossman and Hart (1986). The transfers solve:
$$t^* = \arg \max \{ (a\sqrt{e_A} + b\sqrt{e_B} - \bar{u}_A^\Delta - t)(a\sqrt{e_A} + b\sqrt{e_B} - \bar{u}_B^\Delta + t) \}.$$
23. Observe that $(1 - \lambda)(3 + \lambda)$ is strictly decreasing in λ . Also, it takes the value 3 for $\lambda = 0$ and the value 0 for $\lambda = 1$. Hence, $(b^2/16)(1 - \lambda)(3 + \lambda) \leq (3b^2/16) < (b^2/4)$ implying that \hat{S}^A is less than S^* .
24. See Banerjee, Gertler, and Ghatak (2002) and Mookherjee (1997).

25. Here we assume symmetric Nash-bargaining and therefore $\tau = 1/2$. However, if we use asymmetric Nash-bargaining then τ can be any nonnegative number between 0 and 1 and will reflect the relative bargaining power of A .
26. See, for example, Lin (1988, 1992).
27. This result holds more generally, for example, when the investments are complementary (see Hart, 1995).
28. We have set $\bar{u}_B = 0$.
29. See Pande and Udry (2005) for a comprehensive review of this literature.
30. Di Tella, Galiani, and Schargrodsky (2007) study the formation of beliefs using the same data set and find that lucky squatters who end up with legal titles report beliefs closer to those that favor the workings of a free market. To the extent these beliefs encourage effort and enterprise, this could be an additional channel through which property rights might affect productivity.
31. Our assumption (1) rules out the possibility that *vis-a-vis* projects with low-expected returns, collateral is not a binding constraint.
32. There is a large literature on private enforcement, that is, when protection is provided by profit-maximizing organizations (see Dixit, 2004), and also, some research on private institutions for property-rights, including ones that fight against predation by the government (see Dixit, 2009).
33. Discussions of private property rights enforcement include Grief (2005) and Skaperdas (1992).
34. We will make assumptions to ensure that $\tau < 1$ in any equilibrium.
35. Note however, that this is not the first-best outcome: there is still the standard distortion of an output-based transfer scheme.
36. A general equilibrium model of rent-seeking has also been developed by Acemoglu and Verdier (1998).
37. Observe that if $\alpha > 0$ and $z = 0$, then the coercive authority can commit to $\tilde{\tau}$ such that:

$$Y(\tilde{\tau}) = N\alpha.$$

- If $z > \alpha$, then we get the same equilibrium as with $\alpha = 0$.
38. We are assuming here that the coercive authority can credibly commit to audit and expropriate if producers offer to pay some $t' < t$ such that $\mu\hat{y} - \gamma \leq t'$. If this is not possible, then t will have to be set at the lowest level that is consistent with the constraint of the coercive authority, namely, $\mu\hat{y} - \gamma$.
39. If $\gamma = \gamma_0$ then the argument goes through but the authority would have been exactly as well off as before, not strictly better off. This is why we chose this particular formulation.
40. There could also be a fixed payment which is paid to workers regardless of whether output is realized, but we have set this to zero as we are assuming that the aim of the government is to extract as much surplus as possible from the workers under socialism.
41. In the language of contract theory, the coercive authority is now using a relational contract. See Baker, Gibbons, and Murphy (2002) for a discussion in the context of firms and workers.
42. Garfinkel and Lee (2000) and Marceau and Smart (2003) have applied these ideas to issues related to those studied in this section. See also the historical discussion in Rajan and Zingales (2003).
43. In fact, we know from Kydland and Prescott (1977) that even a welfare maximizing government may have an incentive to announce time inconsistent taxation.
44. See Persson and Tabellini (2008) for a review and discussion of the issues.
45. The above condition holds with equality for $\beta \leq 1/2$. If $\beta > 1/2$, $\tau^* = 1/2$, and $\lambda^* = 0$.
46. See Acemoglu, Robinson, and Verdier (2004) for a model along these lines.
47. This is an example of a collective reputation mechanism, similar to the one studied by Greif (1993) for mediaeval traders in the Mediterranean. We have assumed that, even if there is no collective

punishment, the government commits to a common expropriation rate. Otherwise, the incentive constraint for the group with low exit options would be relaxed, while the one for the group with high exit options would be strengthened.

48. See Nozick (1974).
49. The fact that taxation is typically deterministic and expropriation stochastic is not the key distinction in our framework of risk-neutral producers. Indeed, here a mean-preserving spread in the tax rate or the rate of expropriation would raise efficiency so long as the producer is told which way the uncertainty is resolved before he undertakes effort. Consider a mean-preserving spread in τ such that it is $\bar{\tau}$ with probability p and $\bar{\tau}$ with probability $1 - p$. As the producer's expected payoff is $\frac{1}{4}(1 - \tau)^2 A^2$ which is convex in $1 - \tau$ it follows that he is better off with the random tax schedule. This is a well-known result in public finance—see Weiss (1976).
50. Haber, Maurer, and Razo (2003) argue that there is considerable proprietary knowledge of markets and technology in the Mexican oil industry. They argue that this limited the amount of expropriation that the government could undertake.
51. In an ingenious paper, DeLong and Shleifer (1993) look indirectly at property rights protection in Medieval Europe and find that absolutist monarchs presided over slower growth in urbanization.
52. See Glaeser and Shleifer (2002) for an interesting discussion of the historical political economy behind the development of different legal codes.

References

- Acemoglu, D. (2003). *The form of property rights: Oligarchic vs. democratic societies*. NBER Working Paper No. 10037.
- Acemoglu, D. (2005). Politics and economics in weak and strong states. *Journal of Monetary Economics*, 52(7), 1199–1226.
- Acemoglu, D., & Johnson, S. (2005). Unbundling institutions. *Journal of Political Economy*, 113(5), 949–995.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401.
- Acemoglu, D., & Robinson, J. A. (2000). Why did the west extend the franchise? Democracy, inequality, and growth in historical perspective. *Quarterly Journal of Economics*, 115(4), 1167–1199.
- Acemoglu, D., & Robinson, J. A. (2005). *Economic origins of dictatorship and democracy*. Cambridge: Cambridge University Press.
- Acemoglu, D., Robinson, J. A., & Verdier, T. (2004). Kleptocracy and divide-and-rule: A model of personal rule. *Journal of the European Economic Association*, 2(2–3), 162–192.
- Acemoglu, D., & Verdier, T. (1998). Property rights, corruption and the allocation of talent: A general equilibrium approach. *Economic Journal*, 108(450), 1381–1403.
- Aghion, P., & Bolton, P. (1997). A theory of trickle-down growth and development. *Review of Economic Studies*, 64(2), 151–172.
- Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, 62(5), 777–795.
- Bai, C. E., Li, D. D., Qian, Y., & Wang, Y. (2004). *Commitment, incentives, and information: The case of anonymous banking*. Available at <http://elsa.berkeley.edu/~yqian/anonymous%20banking.pdf>
- Baker, G., Gibbons, R., & Murphy, K. J. (2002). Relational contracts and the theory of the firm. *Quarterly Journal of Economics*, 117(1), 39–84.
- Baklanoff, E. (1975). *Expropriation of US investments in Cuba, Mexico, and Chile*. New York: Praeger Publishers.
- Baland, J. M., Moene, K. O., & Robinson, J. A. (2009). “Governance and Development” draft chapter for the Volume V of the *Handbook of Development Economics*. Amsterdam: North-Holland.

- Banerjee, A. V., Gertler, P. J., & Ghatak, M. (2002). Empowerment and efficiency: Tenancy reform in West Bengal. *Journal of Political Economy*, 110(2), 239–280.
- Banerjee, A. V., & Newman, A. F. (1993). Occupational choice and the process of development. *Journal of Political Economy*, 101(2), 274–298.
- Barber, M. (1994). *The new knighthood: A history of the order of the temple*. London: Cambridge University Press.
- Barzel, Y. (1997). *Economic analysis of property rights* (2nd ed., pp. 1–2). Cambridge: Cambridge University Press.
- Bauer, P. (1954). *West African trade*. Cambridge: Cambridge University Press.
- Baumol, J. (1990). Entrepreneurship: Productive, unproductive, and destructive. *Journal of Political Economy*, 98(5), 893–921.
- Baunsgaard, T., & Keen, M. (2005). *Tax revenue and (or?) trade liberalization*. IMF.
- Besley, T. (1995). Property rights and investment incentives: Theory and evidence from Ghana. *Journal of Political Economy*, 103(5), 903–937.
- Besley, T., & Ghatak, M. (2001). Government versus private ownership of public goods. *Quarterly Journal of Economics*, 116(4), 1343–1372.
- Besley, T., & Ghatak, M. (2008). *Creating collateral: The de Soto effect and the political economy of legal reform*. LSE.
- Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, (forthcoming).
- Besley, T., & Persson, T. (2008). *State capacity, conflict and development*. Paper to support Persson's Econometric Society Presidential Address.
- Braginsky, S., & Myerson, R. B. (2007). *Capital and growth with oligarchic property rights*. Available at <http://home.uchicago.edu/~rmyerson/research/oligarch.pdf>
- Coase, H. (1960). The problem of social cost. *Journal of Law and Economics*, 3(1), 1–44.
- DeLong, J. B. (2000). Overstrong against thyself: War, the state, and growth in Europe on the eve of the industrial revolution. In M. Olson & S. Kahkohnen (Eds.), *A not-so-dismal science: A broader view of economies and societies*. Oxford: Oxford University Press.
- DeLong, J. B., & Shleifer, A. (1993). Princes and merchants: European city growth before the industrial revolution. *Journal of Law and Economics*, 36(2), 671–702.
- De Meza, D., & Gould, J. R. (1992). The social efficiency of private decisions to enforce property rights. *Journal of Political Economy*, 100(3), 561–580.
- De Soto, H. (1989). *The other path: The economic answer to terrorism*. New York: Basic Books.
- De Soto, H. (2000). *The mystery of capital: Why capitalism triumphs in the west and fails everywhere else*. New York: Basic Books.
- De Soto, H. (2001). *The mystery of capital. Finance and development*. IMF. Available at <http://www.imf.org/external/pubs/ft/fandd/2001/03/desoto.htm>
- Diamond, P., & Mirrlees, J. (1971). Optimal taxation and public production: I production efficiency. *American Economic Review*, 61(1), 8–27.
- Di Tella, R., Galiani, S., & Schargrodsky, E. (2007). The formation of beliefs: Evidence from the allocation of land titles to squatters. *Quarterly Journal of Economics*, 122(1), 209–241.
- Dixit, A. K. (2004). *Lawlessness and economics: Alternative modes of governance*. Princeton: Princeton University Press.
- Dixit, A. K. (2009). *Governance institutions and economic activity*. Presidential Address to the American Economic Association, Princeton University.
- Djankov, S., Glaeser, E., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2003). The new comparative economics. *Journal of Comparative Economics*, 31(4), 595–619.
- Engerman, S. L., & Sokoloff, K. L. (2002, Fall). Factor endowments, inequality, and paths of development among new world economies. *Economia*, 41–109.
- Fafchamps, M. (2005). Spontaneous market emergence and social networks. In G. Demange & M. Wooders (Eds.), *Group formation in economics: Networks, clubs and coalitions*. Cambridge: Cambridge University Press.
- Field, E. (2005). Property rights and investment in urban slums. *Journal of the European Economic Association, Papers and Proceedings*, 3(2–3), 279–290.

- Field, E. (2007). Entitled to work: Urban property rights and labor supply in Peru. *Quarterly Journal of Economics*, 122(4), 1561–1602.
- Field, E., & Torero, M. (2006). *Do property titles increase credit access among the urban poor? Evidence from a nationwide titling program*. Harvard University. Working Paper Available at <http://www.economics.harvard.edu/faculty/field/files/FieldToreros.pdf>
- Galiani, S., & Schargrodsky, E. (2005). *Property rights for the poor: Effects of land titling*. Business School Working Paper No. 06/2005. Universidad Torcuato Di Tella.
- Garfinkel, M. R., & Lee, J. (2000). Political influence and the dynamic consistency of policy. *American Economic Review*, 90(3), 649–666.
- Glaeser, E., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth*, 9(3), 271–303.
- Glaeser, E. I., & Shleifer, A. (2002). Legal origins. *Quarterly Journal of Economics*, 107(4), 1193–1229.
- Goldstein, M., & Udry, C. (2008). The profits of power: Land rights and agricultural investment in Ghana. *Journal of Political Economy*, 116(6), 981–1022.
- Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *American Economic Review*, 83(3), 525–548.
- Greif, A. (2005). Commitment, coercion, and markets: The nature and dynamics of institutions supporting exchange. In C. Menard & M. M. Shirley (Eds.), *Handbook of new institutional economics*. Berlin: Springer.
- Grossman, S. J., & Hart, O. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4), 691–719.
- Grossman, G., & Helpman, E. (1994). Protection for sale. *American Economic Review*, 84(4), 833–850.
- Grossman, H. I., & Kim, M. (1995). Swords or plowshares? A theory of the security of claims to property. *Journal of Political Economy*, 103(6), 1275–1288.
- Grossman, H. I., & Noh, S. J. (1994). Proprietary public finance and economic welfare. *Journal of Public Economics*, 53(2), 187–204.
- Guriev, S. M., Kolotilin, A., & Sonin, K. (2008). *Determinants of expropriation in the oil sector: A theory and evidence from panel data*. Available at SSRN: <http://ssrn.com/abstract=1103019>
- Haber, S., Maurer, N., & Razo, A. (2003). When the law does not matter: The rise and decline of the Mexican oil industry. *Journal of Economic History*, 63(1), 1–32.
- Hanstad, T., Haque, T., & Nielsen, R. (2008). Improving land access for India's rural poor. *Economic and Political Weekly of India*, 43(10), 49–56.
- Hart, O. (1995). *Firms, contracts, and financial structure*. New York: Clarendon Press.
- Hart, O., & Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98(6), 1119–1158.
- Hornbeck, R. (2008). *Good fences make good neighbors: Evidence on the effects of property rights*. New York: MIT.
- Kydland, F. E., & Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3), 473–491.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106(6), 1113–1155.
- Lin, J. Y. (1988). The household responsibility system in China's agricultural reform: A theoretical and empirical study. *Economic Development and Cultural Change*, 36(3), 199–224.
- Lin, J. Y. (1992). Rural reforms and agricultural growth in China. *American Economic Review*, 82(1), 34–51.
- Johnson, S., McMillan, J., & Woodruff, C. (2002). Property rights and finance. *American Economic Review*, 92(5), 1335–1356.
- Marceau, N., & Smart, M. (2003). Corporate lobbying and commitment failure in capital taxation. *American Economic Review*, 93(1), 241–251.
- Martin, S. (2004). *The knights templar*. Harpenden: Barber Pocket Essentials.
- Mehlum, H., Moene, K., & Torvik, R. (2006). Institutions and the resource curse. *Economic Journal*, 116 (508), 1–20.
- Mookherjee, D. (1997). Informational rents and property rights in land. In J. Roemer (Ed.), *Property relations, incentives and welfare*. London: Macmillan Press.

- Moselle, B., & Polak, B. (2001). A model of a predatory state. *Journal of Law, Economics, and Organization*, 17(1), 1–33.
- Murphy, K. M., Shleifer, A., & Vishny, R. W. (1991). The allocation of talent: Implications for growth. *Quarterly Journal of Economics*, 106(2), 503–530.
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.
- North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. New York: Cambridge University Press.
- North, D. C., & Weingast, B. R. (1989). Constitutions and commitment: The evolution of institutional governing public choice in seventeenth-century England. *Journal of Economic History*, 44(4), 803–832.
- Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.
- Olson, M. (1993). Dictatorship, democracy, and development. *American Political Science Review*, 87(3), 567–576.
- Persson, T., & Tabellini, G. (2000). *Political economics: Explaining economic policy*. Cambridge: MIT Press.
- Persson, T., & Tabellini, G. (2008). The growth effects of democracy: Is it heterogenous and how can it be estimated? In E. Helpman (Ed.), *Institutions and economic performance*. Cambridge MA: Harvard University Press.
- Platteau, J. P. (2000). *Institutions, social norms, and economic development*. Amsterdam: Harwood Academic Publishers.
- Qian, Y., & Weingast, B. R. (1997). Federalism as a commitment to preserving market incentives. *Journal of Economic Perspectives*, 11(4), 83–92.
- Pande, R., & Udry, C. (2005). *Institutions and development: A view from below*. Economic Growth Center Discussion Paper No. 928. Yale University.
- Rajan, R. G., & Zingales, L. (1998). Which capitalism? Lessons from the East Asian crisis. *Journal of Applied Corporate Finance*, 11, 40–48.
- Rajan, R. G., & Zingales, L. (2003). The emergence of strong property rights: Speculations from history. NBER Working Paper No. 9478.
- Robbins, L. (2000). Lecture 2: Plato and Aristotle. In S. G. Medema & W. J. Samuels (Eds.), *A history of economic thought: The LSE lectures*. Princeton: Princeton University Press.
- Sachs, J. D., & Warner, A. M. (2001). The curse of natural resources. *European Economic Review*, 45(4–6), 827–838.
- Scheiber, H. (1973). Property law, expropriation, and resource allocation by government: The United States, 1789–1910. *The Journal of Economic History*, 33(1), 232–251.
- Searingen, W. (1990). Algeria's food security crisis. *Middle East Report*, 166, 21–25.
- Singh, I., Squire, L., & Strauss, J. (1986). *Agricultural household models: Extensions, applications, and policy*. Baltimore: Johns Hopkins University Press for the World Bank.
- Skaperdas, S. (1992). Cooperation, conflict, and power in the absence of property rights. *American Economic Review*, 82(4), 720–739.
- Sonin, K. (2003). Why the rich may favor poor protection of property rights. *Journal of Comparative Economics*, 31(4), 715–731.
- Wang, S. Y. (2008). *Credit constraints, job mobility and entrepreneurship: Evidence from a property reform in China*. NYU.
- Weingast, B. R. (1997). The political foundations of democracy and the rule of law. *American Political Science Review*, 91(2), 245–263.
- Weiss, L. (1976). The desirability of cheating incentives and randomness in the optimal income tax. *Journal of Political Economy*, 84(6), 1343–1352.

This page intentionally left blank

Governance and Development*

Jean-Marie Baland

C.R.E.D., Facultes Universitaires Notre-Dame de la Paix, 8 Rempart de la Vierge, B-5000 Namur, Belgium.
email: jean-marie.baland@fundp.ac.be.

Karl Ove Moene

Department of Economics and Centre of Equality, Social Organization, and Performance, University of Oslo, Box 1095, NO-0317 Oslo, Norway. email: k.o.moene@econ.uio.no

James A. Robinson

Department of Government and IQSS, Harvard University, Cambridge, MA 02138. email: jrobinson@gov.harvard.edu.

Contents

1. Introduction	4598
2. What is Governance	4602
3. Measurement	4603
3.1 Political institutions	4603
3.2 State capacity and effectiveness	4603
4. Theoretical Mechanisms	4608
4.1 The political economy of governance in Sierra Leone	4610
5. Consequences and Causes of Variation in Governance	4615
5.1 Correlates of governance	4615
5.2 The first stage	4620
5.3 Origins of bad governance in Sierra Leone	4625
6. Improving Governance	4629
6.1 Persistence of power and incentives—The See-Saw Effect	4631
6.2 Persistence of <i>de facto</i> power	4633
6.3 The Iron Law of Oligarchy	4637
6.4 Successful reform	4640
6.5 Governance and equity—The Scandinavian model	4641
6.6 Promoting reform	4647
7. Conclusions	4648
End Notes	4649
References	4650

Abstract

In this chapter, we discuss whether or not “governance” is an important source of variation in development experiences. We draw four main conclusions. First, governance is best thought of a subset of “institutions” and as such emphasis on governance is consistent with much recent academic work. Nevertheless, governance is a quite vague rubric which is difficult to unbundle. Second, the

governance of a society is the outcome of a political process and as such is closely related to the literature on the political economy of development. Third, improving governance necessitates understanding the nature of the entire political equilibrium. Finally, an important research frontier is understanding the forces that create or impeded endogenous changes in governance.

JEL classifications: O1, D7

Keywords

collective decision-making
cooperation
authoritarianism
developing countries
developing country
development

1. INTRODUCTION

Like it or not there are fashions in social science and in beliefs about which paradigm is the correct one for understanding the world. Such fashions are highly prevalent in the study of economic development. When Max Planck famously remarked that “science advances one funeral at a time” he seriously underestimated the creativity of the scholars, policymakers, and practitioners in this field and their willingness to accept new ideas. A current paradigm puts issues of “governance” at the heart of an understanding of development. Poor countries are poor because they have bad governance and the countries that grow or are rich are those that improved their governance. Such a view of development is now enshrined as a key mission of the World Bank whose former President Paul Wolfowitz (2006) summed up this conventional wisdom in the following way during a speech in Indonesia

In the last half-century we have developed a better understanding of what helps governments function effectively and achieve economic progress. In the development community, we have a phrase for it. We call it good governance. It is essentially the combination of transparent and accountable institutions, strong skills and competence, and a fundamental willingness to do the right thing. Those are the things that enable a government to deliver services to its people efficiently . . . An independent judiciary, a free press, and a vibrant civil society are important components of good governance. They balance the power of governments, and they hold them accountable for delivering better services, creating jobs, and improving living standards.

In practice, bad governance is often associated with corruption and reflecting this President Wolfowitz continued

Today one of the biggest threats to development in many countries, including I think here, is corruption. It weakens fundamental systems, it distorts markets, and it encourages people to apply their skills and energies in nonproductive

ways. In the end, governments and citizens will pay a price, a price in lower incomes, in lower investment, and in more volatile economic fluctuations.

One can easily find a plethora of examples from the development experience of the past 50 years which loosely support the notion that bad governance is the root of underdevelopment. The *locus classicus* of how poor governance derailed development in Africa is Tony Killick's *Development Economics in Action*. Killick reports in detail the sad litany of governance failures in Ghana in the 1960s. A startling example is the construction of a fruit canning factory "for the production of mango products, for which there was recognized to be no local market, [and] which was said to exceed by some multiple the total world trade in such items" (Killick, 1978, p. 229). The governments own report on this factory is worth quoting at some length (Killick, 1978, p. 233)

Project A factory is to be erected at Wenchi, Brong Ahafo, to produce 7,000 tons of mangoes and 5,300 tons of tomatoes per annum. If average yields of crops in that area will be 5 tons per acre per annum for mangoes and 5 tons per acre for tomatoes, there should be 1,400 acres of mangoes and 1,060 acres of tomatoes in the field to supply the factory.

The Problem The present supply of mangoes in the area is from a few trees scattered in the bush and tomatoes are not grown on commercial scale, and so the production of these crops will have to start from scratch. Mangoes take 5–7 years from planting to start fruiting. How to obtain sufficient planting materials and to organize production of raw materials quickly become the major problems of this project.

Killick's acerbic comment is that "it is difficult to imagine a more damning commentary on the efficiency of project planning" stated a whole year before the factory was constructed. This is certainly an example of dreadful governance leading to a huge mis-allocation of resources and helping to perpetuate the poverty of Ghana. Nevertheless, there is a big jump from examples of bad governance like this to an explanation for comparative development and useful policy advice. On one hand, one needs to explain why on earth this happened. On the other hand, one needs to carefully establish the causal effect of bad governance in circumstances where governance is endogenous and there are many omitted variables.

In this chapter, we aim to assess the state of knowledge about the role of governance in development. Of course, not all scholars agree with the view that poor development is caused by poor governance and there are many approaches to comparative development. Some scholars dismiss governance out of hand, for instance, Sachs et al. (2004) assert that poor countries cannot afford good governance and after regressing various measures of governance on GDP *per capita* use the residuals to assert that many African countries have surprisingly good governance! Others, such as Hausman, Rodrik, and Velasco (2008) recognize that governance may be an issue, but list it as only one among many other factors which may be the "binding constraint" on growth in any particular society.

Still others would criticize the emphasis of Paul Wolfowitz on “public governance” and the role of the state. Most of the governance literature considers how society is governed in isolation from how it is organized. To apply an old-fashioned distinction, most attention is devoted to *the superstructure* of society, very little to *the base* of the economy. Governance, however, is not only about those who govern, but also about those who are governed. The basic habits of those who are governed are essential in determining public governance. All this was clearly seen by Alexis de Tocqueville. After his 9 months visit to America in 1830 he wrote:

Among the new objects that attracted my attention during my stay in the United States none struck my eye more vividly than the equality of conditions. I discovered without difficulty the enormous influence that this primary fact exerts on the course of society; it gives a certain direction to public spirit, a certain turn to the laws, new maxims to those who govern, and particular habits to the governed. (2000, p. 3)

Though most of the work on governance and development ignores these issues, there are many who would argue that forms of “private governance” are crucial for comparative development. An example of this work might be the work of La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998) on corporate governance and its relation to legal systems. While the importance of the legal system for explaining development (and even corporate governance) is moot (Acemoglu & Johnson, 2005; Acemoglu, Johnson, & Robinson, 2001; Lamoreaux & Rosenthal, 2005), there is no doubt that many aspects of private governance differ across countries (see Bloom & van Reenan, 2007, for important evidence) and that this may be an important source of variation in development experiences. Though one may ultimately be able to treat many aspects of private governance as related to the state,¹ this may not be true of other aspects of the form of organizations and private governance, such as what Kreps called “corporate culture” (Kreps, 1986). This notion relates more basically to the nature of social equilibria which generate trust or cooperation and which may differ radically across societies (as emphasized by Dixit, 2004; Fafchamps, 2004; Greif, 2006; or Platteau, 2000). Though we regard the issue of “private governance” as important and we introduce some elements of it into our discussion, for example, in our discussion of social capital, *de facto* political power, and the Scandinavian experience, the many potential dimensions of this issue mean that it is beyond the scope of our essay. Following de Tocqueville, we therefore introduce aspects of private governance to the extent that they influence the nature of public governance.

We draw the following conclusions. First, the literature on governance is most usefully thought of as part of the larger literature on the relationship between institutions and economic development. Following in the footsteps of North and Thomas (1973), a recent empirical literature, particularly the papers of Hall and Jones (1999) and Acemoglu, Johnson and Robinson (2001, 2002, 2005b), has tried to estimate the causal

effect of economic institutions on development. They find that the preponderance of income differences between poor and rich countries can be explained by differences in these institutions.² Though this empirical literature emphasizes security of property rights, not governance, as being the most crucial institution, it also recognizes that what matters for development is a “cluster of institutions.” The right way to think about governance is as a potentially important part of this cluster of institutions. So far, governance has not been “unbundled” (to use the terminology of Acemoglu & Johnson, 2005), so we do not have definitive evidence for the importance of governance, but there is evidence that some parts of what people mean by this is surely important for development. We note, however, that since governance is really a vague rhetoric it may be intrinsically impossible to unbundle it.

Second, to understand governance and why it varies one has to study the political economy of development. Poor governance is not exogenously assigned and is the outcome of political decisions and as a result will reflect the political institutions and sources of power in society which mould the political process. Nevertheless, there is still a great deal of uncertainty about what political institutions or circumstances lead to good policies and institutions. Some variables, such as measures of constraints on the executive and measures of checks and balances are robustly correlated with development (Acemoglu, Johnson, & Robinson 2005a; Henisz, 2000). But such measures are often equilibrium outcomes, rather than related to more specific institutions such as the form of the constitution. Focusing on these, we know that there is little evidence that the form of the constitution, such as whether the system is presidential or parliamentary, matters for growth, and other aspects of the form of government, such as whether or not it is democratic, do not seem to correlate with development either. As yet, we do not have a satisfactory characterization of the circumstances which lead to a political equilibrium which is conducive to economic growth in a society and it appears that there are many ways in which property rights security, for example, can be achieved. We emphasize these issues by examining the political economy of the “Scandinavian model” of governance and development. If the nature of the political forces that generate stable property rights in early twenty-first century China are very different from those that do so in the United States, so the economic institutions and patterns of governance which generated such rapid growth in Scandinavia during the twentieth century are likewise very dissimilar from those of the United States.

Third, from a policy point of view, the key issue is to understand the entire political equilibrium. Governance reform is unlikely to be successful unless we understand the political forces that generate bad governance in the first place. In lieu of such an understanding, policy reforms to improve governance will often be ineffective. We suggest that, to be effective reform has to change the political equilibrium of a society. Though it is possible that small changes may do this, it is more likely that reform has to take place simultaneously in many dimensions.

Finally, a research priority is to try to understand salient cases of endogenous transitions from bad to good governance (or vice versa, though this is much less common) to identify some generalizations about the factors or circumstances that lead to improved institutions, and by implication, better governance.

The chapter proceeds as follows. In the next section, we ask what is governance? Section 3 then asks how governance can be measured. Section 4 discusses what theoretical mechanisms might link governance with development and illustrates the political economy of bad governance with an extended case study of Sierra Leone. In Section 5, we then discuss the correlates of governance and evidence which suggests that governance has a causal effect on economic development. We include here a discussion on the “first stage” or what causes variation in governance. In Section 6, we move to discussing how to improve governance. We, here, emphasize that the political economy approach to governance suggests that there will be many pitfalls in attempts to reform governance. Section 7 then examines some successful instances of institutional and governance reform and draws some lessons. In this section, we particularly focus on a great success story of the endogenous emergence of good governance—the Scandinavian model. Section 8 concludes this chapter.

2. WHAT IS GOVERNANCE

The first thing is to decide what governance is. The Oxford English dictionary defines governance as “government, control, or authority” or alternatively as “the action, manner, or system of governing.” The organizers of the World Bank’s governance project and website define governance to be,

the traditions and institutions by which authority in a country is exercised. This includes (1) the process by which governments are selected, monitored and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions amongst them.

Kaufman, Kraay, and Zoido-Lobatón (1999, p. 1)

These definitions of governance, as with that proposed by Wolfowitz, are obviously very broad and include not just an emphasis on state capacity and the ability of the government to implement autonomous decisions, which many associate with governance, but also on the form of political institutions which determine who governs and presumably what choices they make. For example, both definitions also stress that accountability is a part of governance. Wolfowitz stresses this directly, while Kaufman, Kraay, and Zoido-Lobatón do so indirectly through their first criterion.

These definitions of governance are so broad that they encompass nearly all aspects of the political institutions of a society in addition to the structure and capacity of the state. Indeed, the only thing they do not directly refer to are the objectives of the

government, though ex-President Wolfowitz mentions a “fundamental willingness to do the right thing.”

In this chapter, by the governance of a society, we shall mean two things. First, the nexus of political institutions in society construed broadly. We, therefore, mean not only the factors such as the form of the constitution and electoral system, whether or not there is democracy, whether the state is federal, and the system of political parties, but also other factors which influence the political power of different groups and individuals in society, such as whether or not they can solve the collective action problem, exert influence, exploit connections, and maybe even fire guns. Such institutions determine the aggregation of preferences in society and thus what institutions and policies arise in equilibrium. Second, governance also refers to the implementation of such collective choices once made. This includes such issues as state capacity and the ability of the state to coherently and efficiently implement policies and the extent and impact of corruption.

It should be clear that with respect to the first meaning, the organization of the economy and perhaps the nature of society more broadly, “private governance,” will be important even if they do not in themselves have a causal effect on development outcomes.

3. MEASUREMENT

3.1 Political institutions

We cannot aspire here to overview the available sources of information on political institutions. Most of these, such as those compiled by Freedom House, the Polity Project, or the book by Persson and Tabellini (2003), are well known and we discuss these when they come up in our analysis. There are other sources of political power of course, such as the ability of groups to solve the collective action problem, and whether or not the central state has a monopoly of violence. These have been much less systematically measured, but they may be just as important in poor countries for the nature of the political equilibrium.

3.2 State capacity and effectiveness

There are several sources for cross-country comparisons of this aspect of governance. The main one, or at least the main one which uses this terminology, is the World Bank’s Governance project. For the past decade or so, they have been collecting information on many institutional variables which are closely related to different notions of governance. These are (1) voice and accountability, (2) political stability and absence of violence, (3) government effectiveness, (4) regulatory quality, (5) rule of law, and (6) control of corruption. Table 1 reports data from the last four of these variables for a variety of countries which pertain to the second sense in which governance is

Table 1 The world's most poorly governed countries

Country	Government effectiveness 2005	Country	Regulatory quality 2005	Country	Rule of law 2005	Country	Control of corruption 2005	Country	Corruption perception index 2006
Somalia	-2.21	Somalia	-2.35	Somalia	-2.36	Equatorial Guinea	-1.79	Haiti	1.8
North Korea	-1.82	North Korea	-2.31	Iraq	-1.81	Somalia	-1.74	Guinea	1.9
Congo, Dem. Rep.	-1.64	Zimbabwe	-2.2	Congo, Dem. Rep.	-1.76	Haiti	-1.45	Myanmar	1.9
Iraq	-1.64	Myanmar	-2.19	Afghanistan	-1.68	Myanmar	-1.44	Iraq	1.9
Comoros	-1.63	Turkmenistan	-1.95	Haiti	-1.62	Sudan	-1.4	Chad	2
Myanmar	-1.61	Eritrea	-1.84	Liberia	-1.6	Afghanistan	-1.37	Bangladesh	2
Turkmenistan	-1.57	Cuba	-1.75	Myanmar	-1.56	Congo, Dem. Rep.	-1.34	Sudan	2
Central African Republic	-1.47	Uzbekistan	-1.71	Sudan	-1.48	North Korea	-1.32	Congo, Dem. Rep.	2
Guinea-Bissau	-1.46	Liberia	-1.7	Cote d'Ivoire	-1.47	Turkmenistan	-1.3	Equatorial Guinea	2.1
Zimbabwe	-1.42	Congo, Dem. Rep.	-1.66	Zimbabwe	-1.47	Tonga	-1.28	Cote d'Ivoire	2.1

Equatorial Guinea	-1.42	Afghanistan	-1.63	Congo, Rep of	-1.42	Iraq	-1.27	Cambodia	2.1
Haiti	-1.39	Comoros	-1.63	Turkmenistan	-1.41	Zimbabwe	-1.24	Uzbekistan	2.1
Togo	-1.38	Iraq	-1.61	Nigeria	-1.38	Cote d'Ivoire	-1.23	Belarus	2.1
Cote d'Ivoire	-1.38	Belarus	-1.53	Guinea-Bissau	-1.33	Nigeria	-1.22	Kenya	2.2
Liberia	-1.36	Iran	-1.49	Equatorial Guinea	-1.33	Chad	-1.22	Kyrgyzstan	2.2
Burundi	-1.34	Libya	-1.44	Uzbekistan	-1.31	Paraguay	-1.19	Nigeria	2.2
Rep. of Congo	-1.31	Equatorial Guinea	-1.31	Central African Republic	-1.29	Bangladesh	-1.18		
Sudan	-1.3	Sudan	-1.29	Angola	-1.28	Cameroon	-1.15		
Syria	-1.23	Angola	-1.24	Chad	-1.23	Cambodia	-1.12		
Afghanistan	-1.2	Central African Republic	-1.23	Venezuela	-1.22	Laos	-1.1		

defined in the previous section. We use the latest data received from the World Bank, which is 2005, and the first four columns show the bottom decile of the distribution of countries. All the series are normalized to lie between 2.5 and -2.5, with -2.5 being the lowest possible score.

Table 1 shows that there is a fair amount of agreement among these different dimensions of governance about which countries in the world have the worst governance. The usual suspects include Somalia, Haiti, Sudan, North Korea, the Democratic Republic of the Congo, and Myanmar (Burma). Needless to say, these same countries would also be right at the bottom of any ranking of countries based on voice and accountability or political stability and absence of violence.

The World Bank defines government effectiveness as “the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government’s commitment to such policies.” Regulatory quality is defined as “the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.” Rule of law is defined as “the extent to which agents have confidence in and abide by the rules of society, and in particular of contract enforcement, the policy, and the courts, as well as the likelihood of crime and violence.” Finally, control of corruption is “the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as ‘capture’ of the state by elites and private interests” (World Bank, 2006, pp. 2-3).

These indices are constructed by unobserved components from 31 different datasets collected by 25 different entities (see World Bank, 2006, p. 17). Some of these datasets are surveys, while some of them represent the perceptions of experts in different areas. Data based on expert perceptions have been much used in the recent literature, for example, the influential data on the security of property rights constructed by Political Risk Services. Obviously, data based on perceptions are problematic, in the absence of a valid instrument, in that one tends to see low corruption in a country that is doing well and vice versa. For instance, Kang (2002) specifically argues that the notion that corruption fell in South Korea in the 1960s is false and he claims that the growth of the economy simply led people to believe that corruption must have been lower. Information can be based on the perceptions of only a small number of informants. It may also be true that a free press, by denouncing more regularly corrupt practices, may give the image of more corruption, while the actual amount is or becomes smaller. In addition, to the extent that this data represent data about a country that is relevant to international businessmen (the usual market for the firms collecting this data) there may also be an issue of whether governance for international investors is the same as for domestic investors.

Nevertheless, one should be not overly skeptical about this data. Every source of data has problems and many different and independent strategies for measuring

governance come up with similar rankings of countries and it seems unlikely that a country in which foreign investors had to deal with corruption would be one where its own citizens did not have to also deal with corruption, or that countries use one bureaucracy for foreigners and another one for their domestic residents. Though it may be true that President Mobutu's Zairianization program was focused on expropriating the assets of foreign nationals (Turner and Young, 1985, chapter 11), the assets of Zairians were not safe either.

An example of one of the surveys used in constructing the World Bank's governance indicators in the BEEPS II Interactive Dataset: Enterprise Survey in Transition, 2002.³ This is based on surveys of firms in transition economies and asks a series of questions about governance and corruption. For instance, it asks "When firms in your industry do business with the government, how much of the contract value would be typically paid in additional or unofficial payments/gifts to secure the contract?" Other important surveys which the World Bank uses for its governance indices include the Latin Barometer⁴ which, for example, asks respondents to report actual percentages of corrupt officials or actual number of times they witnessed the acts of corruption, and the Afrobarometer which also has questions about governance and corruption.

Another important source of information on corruption comes from Transparency International. Their corruption perceptions index is widely used in empirical work (for instance, Persson, Tabellini and Trebbi, 2003; Treisman, 2000) and the final column of Table 1 shows the most corrupt countries in the world according to Transparency International. Transparency International also has a bribes index where they survey firms and enquire regarding bribing.⁵

There are, however, other sources of information about governance and corruption. One of the most important and influential studies of governance is that of Putnam (1993). He focused on the factors that determined the comparative performance of government in Italian regions. To do this, he collected original data on 12 measures of governance and tried to explain why these varied so much within Italy. Several of the series he collected are very interesting and are worth understanding in detail. One measure, which he called "budget promptness" was the average delay in approving the state budget over the period 1979–1985. The date at which this happened was on average January 27 in Friuli-Venezia Giulia to August 7 in Calabria. Another measure very relevant for state capacity is whether or not the state had created any statistical or information facilities. Six regions, Abruzzi, Calabria, Campania, Marche, Molise, Puglia, and Sicily, had none. Other regions, such as Emilia-Romagna and Lombardia were well equipped with such services. Putnam (1993) also collected information on legislative innovation. To do this, he examined 12 different areas on which similar laws appeared in different regions, for example, air and water pollution, consumer protection, strip mining regulation, etc. He then coded which regions led in implementing these laws measuring on average across the 12 domains how soon after the first

appearance was the law implemented in a particular region. The pioneering region was given a score of 100 with a region that never adopted being given a 0. This data were aggregated into an index of legislative innovation. A final index of interest was that which attempted to measure bureaucratic responsiveness. To construct this, in January 1983, researchers contacted regional bureaucracies requesting information on various issues, for instance, the agricultural department was asked, on behalf of a “farmer friend” for information about loans and subsidies for experimental crops. The initial requests were by mail, and replies were evaluated for their promptness, clarity, and comprehensiveness. If no reply came then follow up telephone calls were tried and even personal visits made. From all this information, Putnam constructed an index of bureaucratic responsiveness. Fascinatingly, in Emilia-Romagna and Valle d’Aosta two of the three requests received detailed replies within 1 week, while in Calabria, Campania, and Sardinia none of the mailed requests received a reply at all.

With respect to data on corruption, an important study by Ferraz and Finan (2008) used data on a random corruption audit of municipal governments carried out by the central government in Brazil. They construct an index of corruption which is simply the number of violations that mayors are found to have done. Another innovative study is by Olken (2007) who used audit data on 600 road projects in Indonesia to investigate the misallocation of funds.

4. THEORETICAL MECHANISMS

Given the different ways that one can think of good governance, there are obviously a huge number of different potential theoretical mechanisms linking governance to development. Indeed, given that we shall emphasize the political roots of governance, this topic encompasses almost the entire literature on political economy and we cannot hope to survey this literature here. In this section, therefore, we do the following, we first emphasize a few key ideas and present what we feel is a basic approach to thinking about governance and the politics of governance. We then illustrate the political forces that lead to bad governance with an in-depth case study of Sierra Leone since the political economy of decline here is very representative of the political economy of governance in much of Sub-Saharan Africa.⁶

Development is caused by a society adopting institutions and policies that create incentives for its citizens to save, invest, and innovate. Though institutions such as secure property rights create huge potential Pareto improvements in society, in general, growth enhancing institutions have important distributional effects. To refer to an obvious example, though the insecurity of human and property rights in Darfur is a disaster for economic incentives, it is in the interests of the current Sudanese government in Khartoum. These interests may not be immediately economic, but can also be political. In the Sudanese case, the policy of the government toward Darfur is

focused on maintaining power. Referring Table 1, this approach implies that while the rule of law, for example, may be poor in Sudan, and even though this is costly for growth, the governing elite in Khartoum is better off without the rule of law (at least if that is the only way they can retain their rents).⁷

Economic institutions and policies are endogenous and are determined as collective choices of the society. Governance refers both to these choices, such as whether to build an effective bureaucracy or establish the rule of law, and also refers to parts of the institutional nexus which lead to these choices by influencing who has power and how it can be exercised. Clearly, there is no guarantee that all individuals and groups will prefer the same set of economic institutions because, as noted earlier, different economic institutions and policies lead to different distributions of incomes and power. Consequently, there will be a conflict of interest over the choice of economic institutions. In such a situation, it will be the inherited distribution of political power in society that determines what institution has been chosen. The group with more political power will tend to secure the set of economic institutions and policies that it prefers.

The distribution of political power in society is also endogenous, however. Following Acemoglu and Robinson (2006), we distinguish between two components of political power: *de jure* and *de facto* political power. Here, *de jure* political power refers to all types of power that originates from the political institutions in society. Political institutions determine the constraints on and the incentives of key actors in the political sphere. Examples of political institutions include the form of government, for example, democracy versus dictatorship or autocracy, and the extent of constraints on politicians and political elites. However, a group of individuals, even if they are not allocated power by political institutions, for example, as specified in the constitution, may nonetheless possess political power. Namely, they can engage in collective action, revolt, use arms, hire mercenaries, or use economically costly but largely peaceful protests in order to impose their wishes on society. We refer to this type of political power as *de facto* political power.

It will be this composition of *de facto* and *de jure* power in society that determines the actual power of a group or set of interests and this will determine which economic institutions and policies arise. This vision emphasizes that those with power today take decisions not just to maximize their income today, but also to maintain their grip on power. These goals are often in contradiction. This can be for the simple reason that economic policies which increase even the incomes of elites today may increase the incomes of opponents even more, thus influencing the future distribution of *de facto* power. It may also be that, as pointed out in the seminal study by Bates (1981), good economic policies are often not good politics. In particular, though providing public goods may increase the incomes of the elite, staying in power may be better achieved by using redistributive instruments which can be targeted at supporters and withheld from opponents.

How does governance fit into this scheme of things? Used in the first sense, of the nexus of political institutions, governance plays an important role in determining the

distribution of *de jure* power in society. This matters for the aggregation of preferences. For example, if we consider the Glorious Revolution of 1688 in England, we would identify the change in political institutions which endowed Parliament with more power (see North & Weingast, 1989) as an improvement of governance. It had important implications for the types of policies and economic institutions that arose because it changed the distribution of *de jure* power in the direction of those who had an interest in socially more desirable policies. The Glorious Revolution was a change in governance not just because it changed political institutions, but also because it led to a revolution in the nature of the state and how it functioned. For instance, Brewer (1988) studied the creation of the excise tax bureaucracy in eighteenth century England. Brewer (chapter 4) shows how a fiscal-military state emerged after 1688 which needed a strengthened revenue base. Despite the fact that many positions in the bureaucracy were venal and there was a lot of corruption, both Whig and Tory governments built a highly efficient and meritocratic excise bureaucracy because they needed new taxes to fund naval expansion. Figure 1 shows the excise rounds of Supervisor George Cowperthwaite between June 12 and July 5, 1710. Brewer records that during this trip Supervisor Cowperthwaite traveled 290 miles, visited 263 victuallers, 71 maltsters, 20 chandlers, and one common brewer. In all, he took 81 different measurements of production and checked the work of nine different excisemen who worked for him. The government stopped corruption by collecting a huge amount of information, creating incentives and punishments, and also regularly rotated excise officials in order to stop them colluding with local employers.⁸ So the Glorious Revolution led to changes in both senses of governance as we discussed earlier.⁹

Our theoretical approach to governance is not the approach of many studies. For instance, much of the work on corruption sees it as an inevitable by-product of bureaucracy and the principal agent problems inherent in having a state (e.g., see Bertrand, Djankov, Hanna, & Mullainathan, 2007). This literature often focuses on the way corruption can be mitigated or eliminated, as well as its consequences for private incentives and resource allocation.¹⁰ Our emphasis instead is on the political equilibrium of which corruption may be one consequence. Though corruption in bureaucracies may be the inevitable result of principal agent problems, and no doubt takes place in every country of the world, we are dubious that variation in this type of corruption is sufficient to account for first-order patterns of development. The type of corruption that one sees in countries such as Sierra Leone, which we examine next, may be sufficient, but as we shall argue this corruption has to be seen as part of a vector of governance pathologies which stemmed from the political strategy of the government of Sierra Leone.

4.1 The Political Economy of Governance in Sierra Leone

In 2002, Sierra Leone emerged from a 9-year civil war as possibly the poorest country in the World in terms of *per capita* income and ranked last in terms of the United

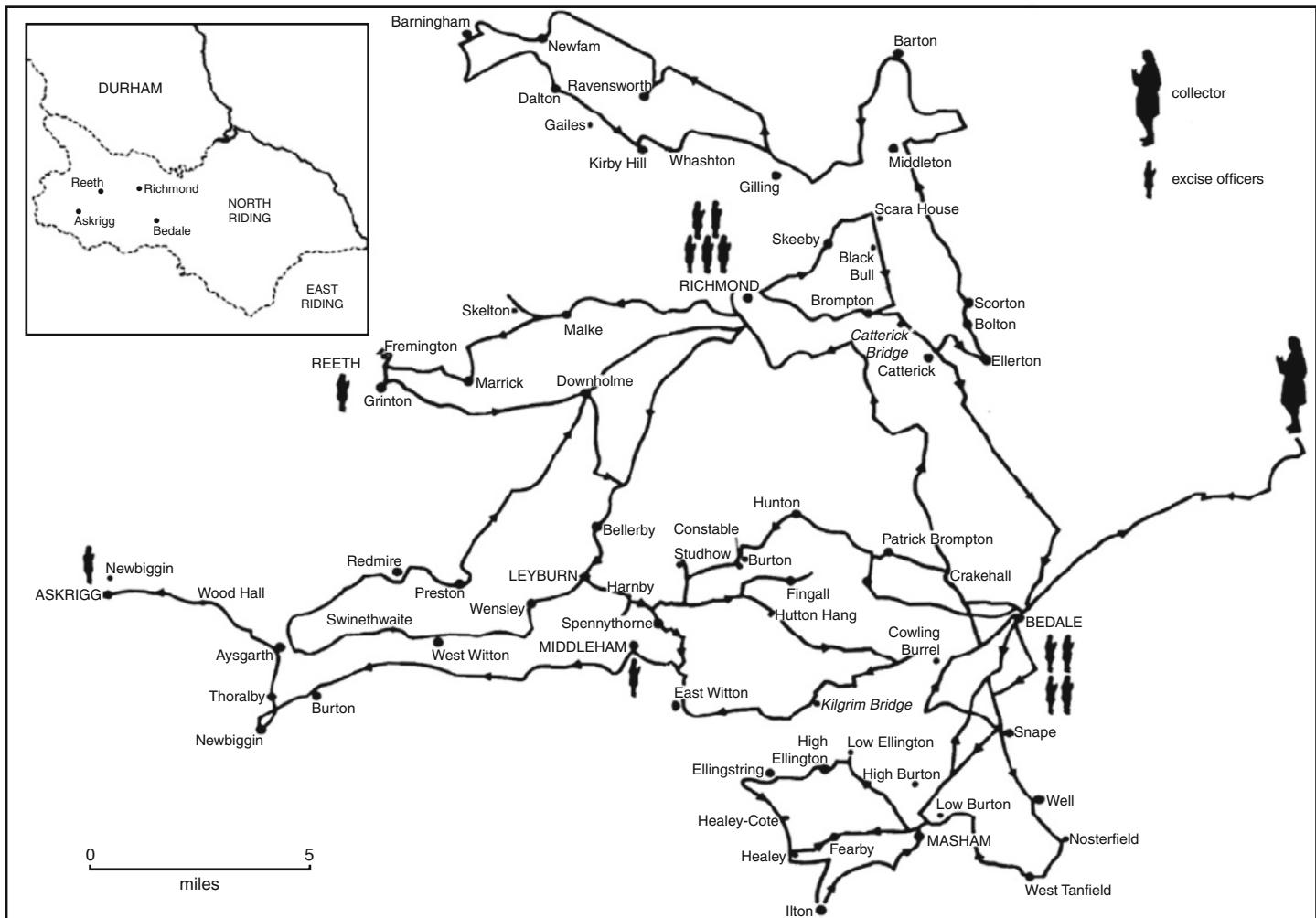


Figure 1 Supervisor Cowperthwaite's excise round, June 12-July 5, 1710, Richmond, Yorkshire. Source: Brewer (1988).

Nation's Human Development Index. The political and economy history of the country since independence in 1961 can be easily summarized. After the early governments of Sir Milton Margai and his brother Sir Albert Margai, successive leaders of the Sierra Leone People's Party (SLPP), the country was ruled from 1967 by the All People's Congress Party (APC) until it was ejected from power by a military coup in 1992. During this period, under the presidencies of Siaka Stevens until 1985 and subsequently Joseph Momoh, the economy declined almost monotonically and state institutions collapsed. The coup, led by a group of young officers who formed the National Provisional Ruling Council (NPRC), was in response to the widening civil war which began with the first incursion across the border from Liberia of the Revolutionary United Front (RUF) in March 1991.

It is clear that Sierra Leone is a poor country because it has had terrible governance. Though the civil war which blighted the country between 1991 and 2002 caused further economic hardship and distress, the main reason that Sierra Leone emerged from war as the world's poorest country, is that it entered the war with that status in 1991.¹¹ Sierra Leone is not intrinsically poor, it has diamonds and great agricultural potential and is close to European markets. Had governance been better Sierra Leone may not have become South Korea or Taiwan, but at worst it would have become Botswana.

The British colonial administration seems to have had little interest in developing the country and even appears to have actively undermined business interests of Krios (the descendants of freed slaves who formed something of an elite during the colonial period) from Freetown, for example, by promoting Lebanese commercial interests. The British also designed political institutions in the 1950s to guarantee that the more educated Krios would not dominate the politics of a newly independent Sierra Leone. Since independence, a sequence of politicians has had no interest or incentive in providing the most elementary things which can lead a society to prosper. The main reason for this is that postindependent politicians ruled the country using the political strategy of patrimonialism, and the general unaccountability of the political class led to a high degree of kleptocracy.¹²

Patrimonialism, also called "neo-patrimonialism," "personal rule," or more simply "clientelism," is a style of governance where politicians control power through a system of personal relationships where policies/favors are distributed in exchange for political support. There are several reasons that this is disastrous for economic policy and performance and they hinge mostly on how the exchange is structured in order to maximize the control and bargaining power of those running such regimes.

First, the form in which patrimonial or clientelistic exchanges have to take place is highly inefficient. For instance, patrons will find it politically desirable to use private goods which can be targeted to supporters and withheld from opponents. Public goods are not politically attractive ways to generate support and are thus generically undersupplied under patrimonialism.

Second, patrimonial rulers need to make people reliant on them for their future success or failure. To do this, they create insecurity and uncertainty which only they can resolve. To quote a famous example, Rafael Trujillo, who ruled the Dominican Republic for 31 years forced all politicians to write a resignation letter which he kept on his desk. One parliamentarian was served with his resignation during his own speech in parliament (Turits, 2003, for this story). One robust consequence of patrimonialism is that property rights are insecure. People only have property because patrons allow them to have it, but such rights are always conditional and can be withdrawn. This creates terrible incentives to invest in assets. Moreover, laws are selectively applied with no concept of the rule of law or equality before the law which of course are completely inconsistent with how clientelism is dispensed. In a patrimonial regime, you have rights if you are a client of the patron and otherwise you do not. The application of uniform rules or criteria to allocate resources impedes the ability of patrons to use discretion.

Third, as Bates (1981) first pointed out, patrimonial regimes create distortions in market prices to create rents which can then be politically allocated. When supply is not equal to demand, something is in short supply and this is a great political resource to those who can allocate it. This creates massive economic distortions, but it can be good politics.

Finally, patrimonialism undermines the coherence of the bureaucracy. This is because the bureaucracy represents a potential source of political opposition to patrons, and a consequence is that bureaucrats are continually “shuffled” so that they cannot conspire against rulers—a rather different type of shuffling from that experienced by eighteenth century British excise tax collectors. Another reason seems to be that bureaucrats in patrimonial regimes are even encouraged to be corrupt and steal, perhaps because this gives patrons more leverage over them. As Mobutu Sésé Seko famously said in this context

If you want to steal, steal a little in a nice way. But if you steal too much to become rich overnight, you'll be caught. (quoted in Gould, 1980, p. 485)

This strategy of patrimonialism makes it very difficult for a central state to really establish and institutionalize its capacity and control over its territory. At some level, as we shall shortly discuss, patrimonialism is an attempt to create a national political order, but it does so not by eliminating alternative sources of authority in society, but by co-opting them. It does so not by trying to create a national identity, but by attempting to disarticulate potential sources of opposition or alternative identities. This leaves many sources of potential challenges to patrimonial rule simmering close to the surface and is why many of them break down. This problem is exacerbated by the fact that civilian control of the military seems to be inconsistent with patrimonialism, possibly because the regimes lack legitimacy and also because by their nature highly personalized

regimes are very easy for even small groups of dissident soldiers to overthrow.¹³ The reaction to this is that patrimonial regimes tend to keep the military very weak so as not to mount a credible threat. Instead, they often privatize security with presidents having bodyguards, often consisting of foreign nationals.

This somewhat stylized vision of how a patrimonial regime functions seems to fit quite well with accounts of how Stevens and Momoh ran the country. They used the control and discretion that this gave them to loot the nation's wealth. This description also seems to help understand the utter lack of provision of public services and the increasing inability of the central state to impose its writ on the country. The characterization fits right down to the decision by Stevens shortly after assuming power to cut the army to 1000 men since it was initially a coup that had stopped him assuming power in 1967. Stevens also privatized violence creating a private security force initially named the Internal Security Unit (the ISU—which was apparently referred to by his long-suffering people as “I Shoot U”) and afterwards the Special Security Division (SSD—or “Siaka Steven’s Dogs,” Jackson, 2004, p. 63; Keen, 2005, p. 17 on these acronyms). In the end it was a group of only 30 soldiers led by Captain Valentine Strasser which pitched the APC regime from power on April 30, 1992.

The APC came to power on the basis of strong support in the North of the country, particularly from the Temne and Limba ethnic groups. Patrimonialism was perfected to a fine art by Stevens, who was known as “Pa Siakie” (father of the nation—Kpundeh, 1995, p. 23). Though Stevens built his patrimonialism on the social networks of the APC and much of the academic literature on this topic emphasizes the importance of informal relations, the strategy was greatly facilitated by changes in political institutions which began soon after he assumed power. Importantly, these involved making himself president and concentrating power in the executive, suspending democracy and creating a one-party state, first *de facto* in 1973 and then *de jure* in 1977. Stevens also seriously manipulated traditional political institutions such as Chiefship, buying support through the distribution of patronage and jobs, privatizing and personalizing state finances, and looting the diamond wealth of the country (Barrows, 1976; Reno, 1995; Tangri, 1978).¹⁴ Violence and the coercion of political opponents, if not on the scale of Omar Al Bashir, Idi Amin, Samuel Doe, Mengistu Haile Miriam, or Charles Taylor, were a regular feature of life and criticism was not tolerated. A notorious example being the murder in 1980 by defenestration from a top floor office window of Sam Bangura, the governor of the central bank (Reno, 1995, pp. 137–141). This was “politics of the belly” writ large (Bayart, 1993). The regime also featured the dominance of Freetown over the rest of the country and an urban bias (Bates, 1981; Lipton, 1977) which involved the reversal of decentralization and the abolition of district councils in 1972.

Casual empiricism suggests that barely any public goods were provided in the country in the 40 years prior to the end of the civil war and re-democratization in 2002.

After coming to power in 1967, Stevens, a man who apparently without irony used to enjoy quoting the Krio aphorism “the cow eats where it is tethered,”¹⁵ famously pulled up the railway to Bo, Kenema, and Pendembu and sold off all the track and rolling stock to make the change as irreversible as possible. Though interpretations of this event differ, a salient one is that he did this to isolate Mendeland which was the area which most strongly supported the SLPP (for instance, Abharam & Sesay, 1993, p. 120; Davies, 2007, pp. 684–685; Richards, 1996, pp. 42–43). The roads fell to pieces and schools disintegrated. National television broadcasts stopped in 1987 when the transmitter was sold by the Minister of Information and in 1989 a radio tower which relayed radio signals outside Freetown fell down ending transmissions outside the capital (Reno, 2003, p. 48 for these stories). Other aspects of economic policy also fit very well with the classical analyses of patrimonial regimes. For example, the Sierra Leone Produce Marketing Board, inherited from the British, had a monopsony over all export crops, and paid farmers very low prices (Bates, 1981) as low as 40% of the world level (Davies, 2007). The exchange rate was massively overvalued creating a black market and a scarce resource which Stevens’ allocated through what Reno (1995) dubbed the “shadow economy.” According to Maddison’s data, GDP *per capita* fell almost monotonically from the early 1970s onward and declined to about 40% of the level recorded at independence by the end of the civil war. As if to add insult to injury, Stevens (1984) left a highly disingenuous autobiography!

There is a simple lesson from this section. Most of the things that people refer to as “bad governance” are the consequences of a particular strategy of rule or power consolidation. They are not an inevitable outcome of principal agent problems. No doubt there was bureaucratic corruption in Sierra Leone along the lines studied in much of the literature on corruption, but this was second order compared to the institutionalized bad governance emanating from the state itself.

5. CONSEQUENCES AND CAUSES OF VARIATION IN GOVERNANCE

5.1 Correlates of governance

There are several well-established facts about the correlations between measures of governance and corruption and various socioeconomic and institutional outcomes. We produce some of these in a few figures. Figures 2–5 show the correlation between the World Bank’s four main measures of governance we discussed earlier and GDP *per capita*. Figure 6 shows the relationship between average income and the Transparency International corruption index. It is clear that these measures of governance are all highly correlated with income, and also obviously highly correlated with each other. Generally, one does not find a country with good rule of law and endemic corruption. Though these indices of governance heavily use subjective ratings by experts, the

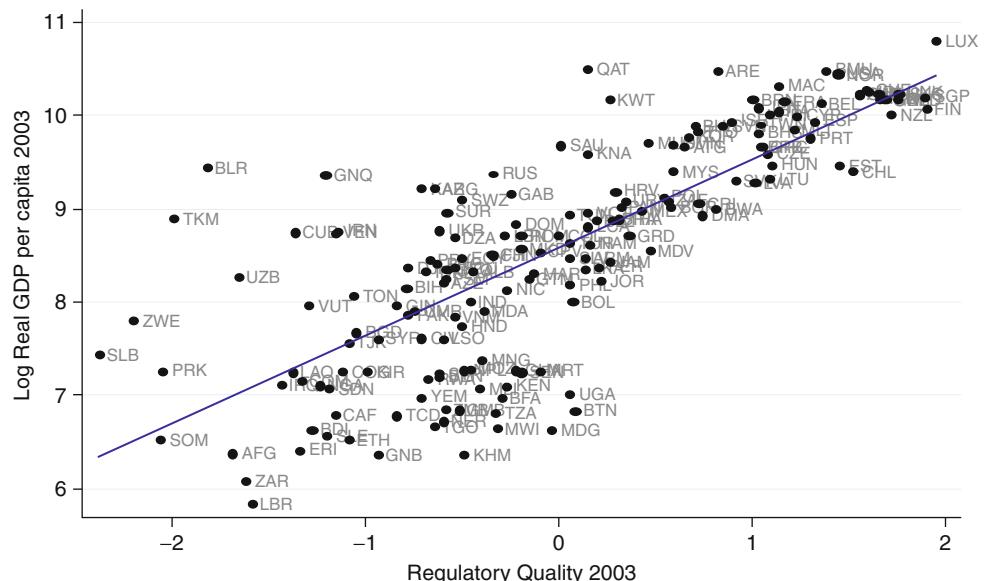


Figure 2 Income and government effectiveness.

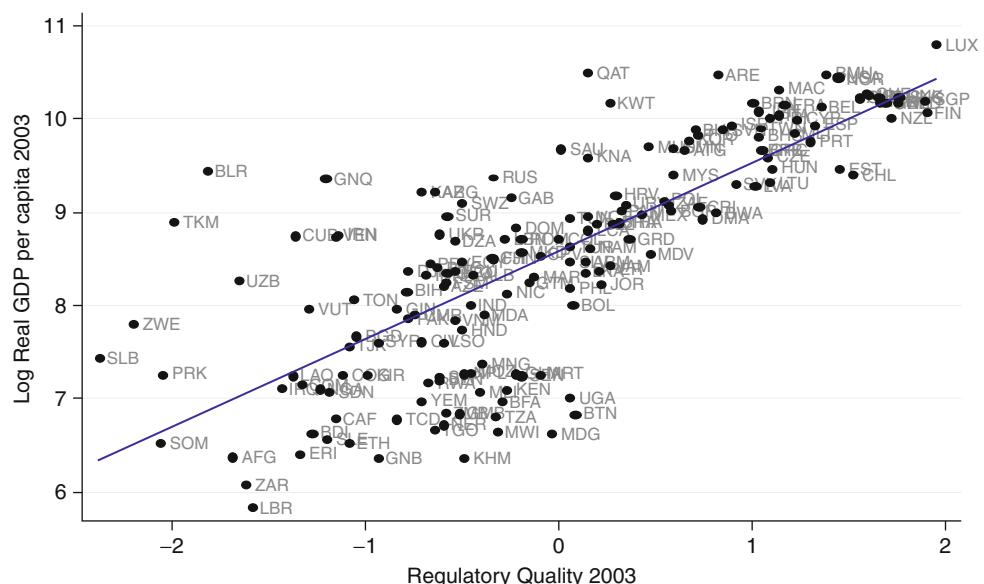


Figure 3 Income and regulatory quality.

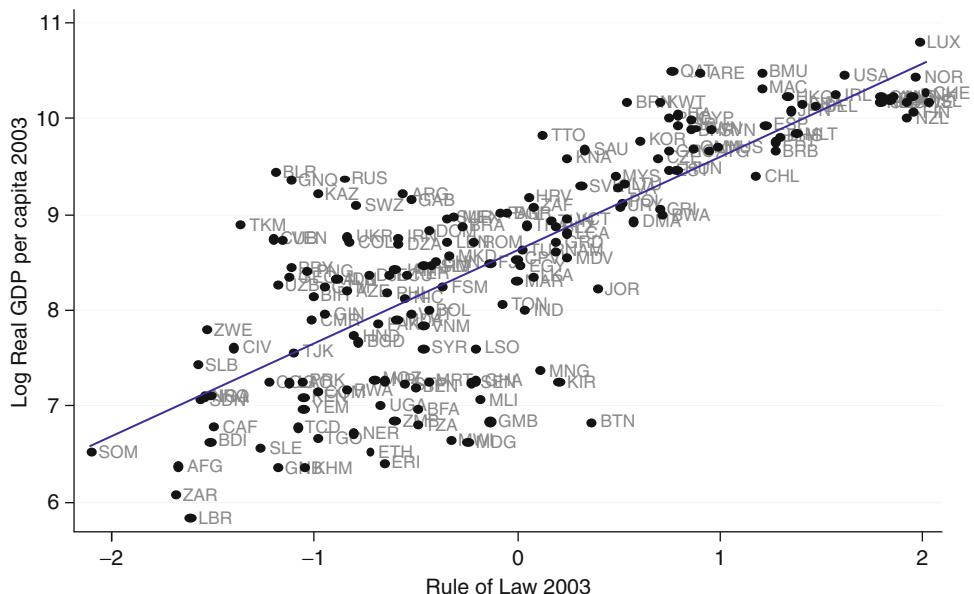


Figure 4 Income and the rule of law.

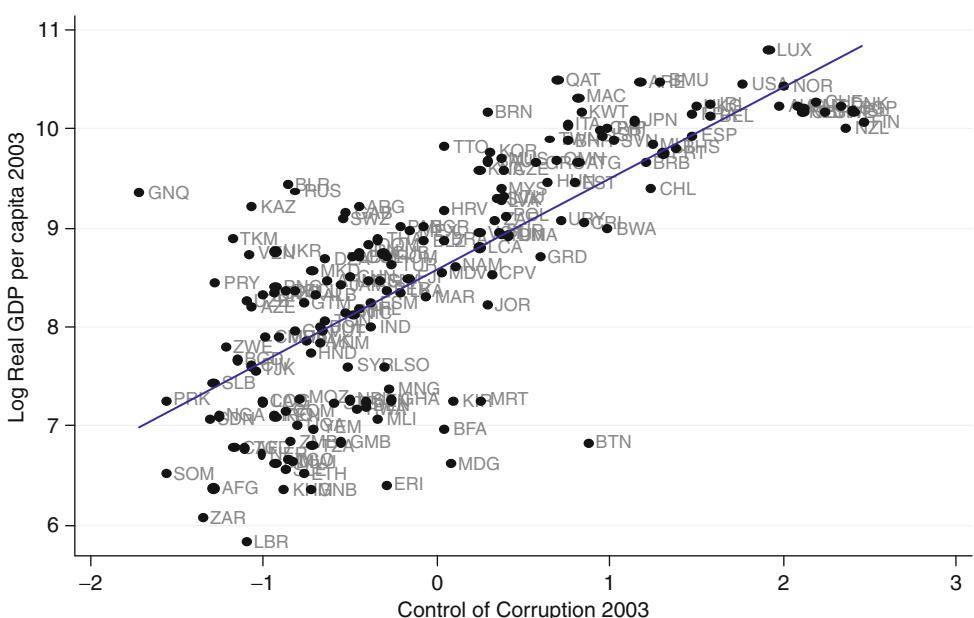


Figure 5 Income and control of corruption.

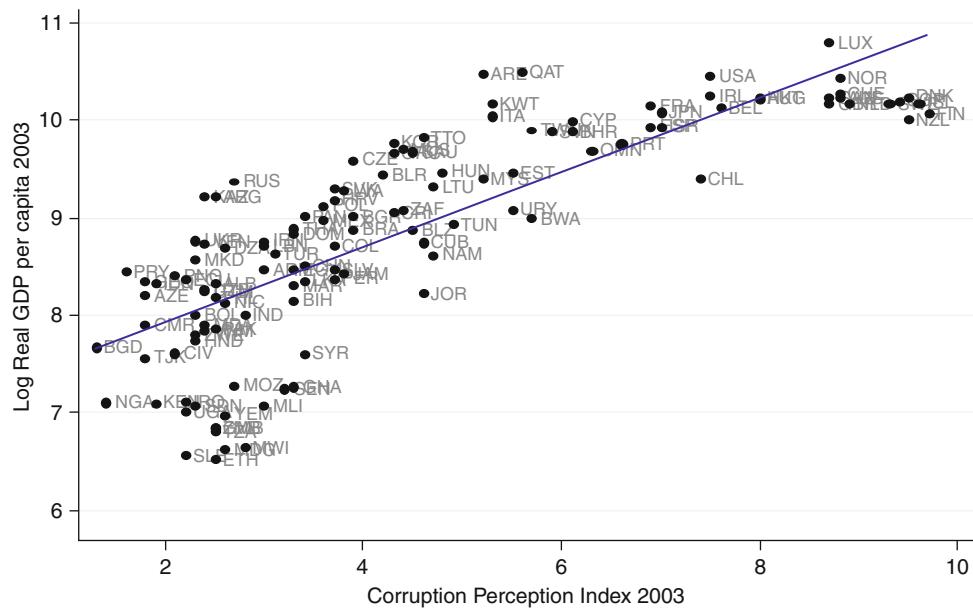


Figure 6 Income and corruption perceptions.

association between income and governance is real. This is clear from Putnam's data on Italy where his measures of governance are highly correlated with regional income *per capita* (see Putnam, 1993, figure 4.2, p. 85) and from Figure 7 which shows the correlation from Ferraz and Finan (2008) between the number of corrupt violations permunicipality in Brazil and the income *per capita* of the municipality. Clearly, more prosperous municipalities have fewer corruption violations.

Obviously, however, these figures say nothing about any type of causal relationship. One simple strategy to examine this might be to use panel data and country fixed effects. To think about what this approach might yield, Figures 8 and 9 plot the change in the World Bank's measures of government effectiveness and control of corruption 1996–2003 against the change in the log of GDP *per capita* of the country over the same period. These changes are uncorrelated in the case of control of corruption even though in quite a few cases the value of the index changes quite a lot. On the other hand, Figure 8 shows that there is some evidence of a positive association between changes in income *per capita* and government effectiveness over this period.¹⁶

Our perspective in this chapter, however, is that one should think of governance as something intimately related to the wider literature on institutions and growth. From this perspective, one can interpret the evidence in Hall and Jones (1999) and Acemoglu et al. (2001, 2002) as demonstrating that governance has a causal effect on economic growth. Nevertheless, one should perhaps be cautious about this. First, these papers propose an instrument for economic institutions and though the argument behind

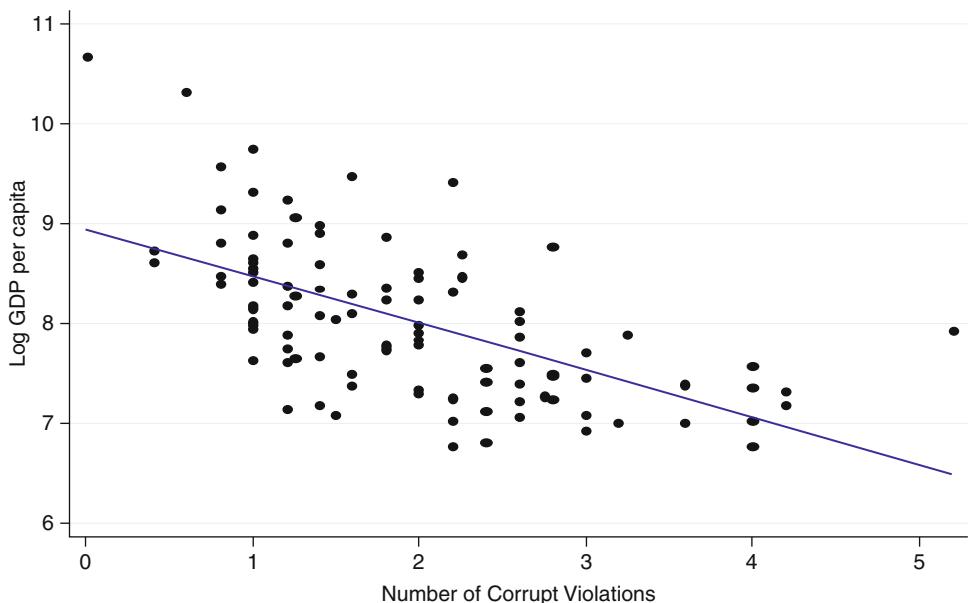


Figure 7 Corruption and income across Brazilian municipalities.

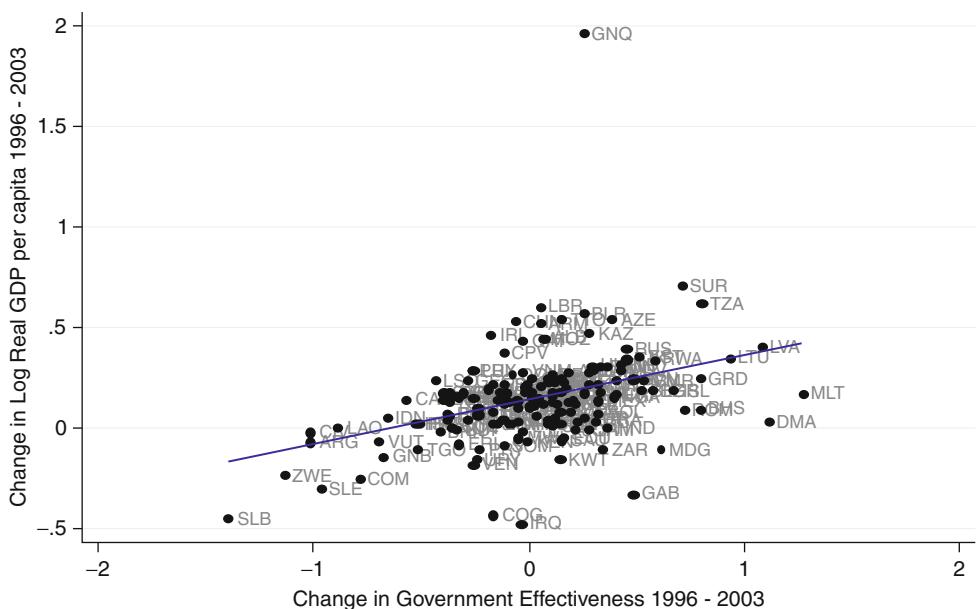


Figure 8 The “within variation”—government effectiveness.

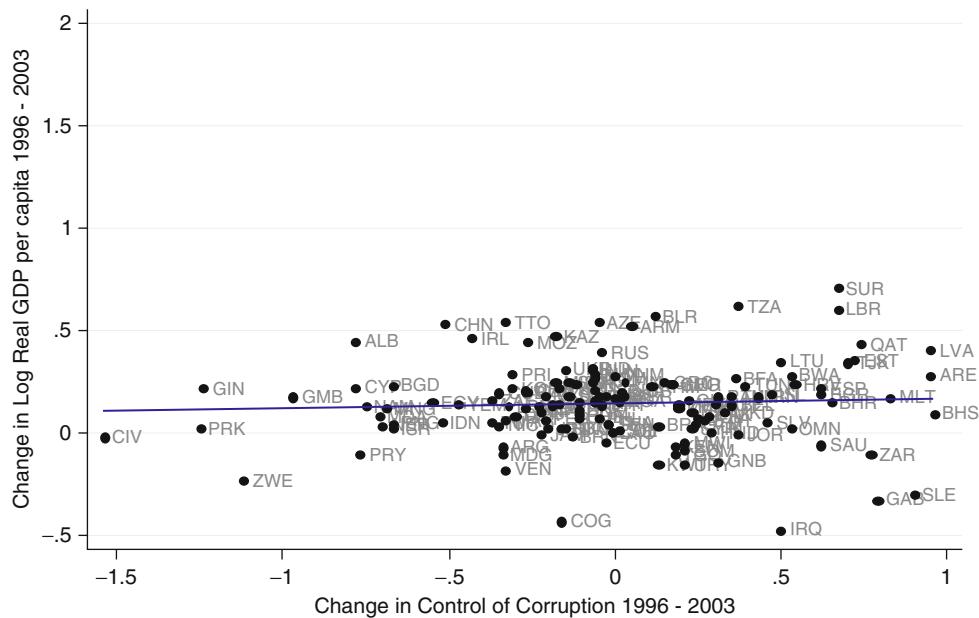


Figure 9 “The within variation”—control of corruption.

the instrument in Acemoglu et al. involves elements of governance, they do not directly test a hypothesis about governance. Other work, such as that of Mauro (1995) which attempted to develop an instrument for corruption, is not very successful. Mauro suggested that ethnolinguistic fragmentation influences growth only via its impact on corruption, which seems rather unlikely (see Acemoglu, 2005a for a discussion of this issue). Other attempts to instrument measures of governance by variables such as settler mortality (e.g., Kaufman & Kraay, 2002) have exclusion restrictions which are equally as implausible.

We would conclude, therefore, that while there is evidence that a “cluster of institutions” does have a causal effect on economic development and it is plausible that governance is connected to this, as yet governance has not been unbundled. This is not very surprising since the word is used in many different ways and quite imprecisely at that.

5.2 The first stage

One of the most important aspects of the empirical approach of Hall and Jones (1999) and Acemoglu et al. (2001, 2002) and the more historical agenda developed by Engerman and Sokoloff (1997) was the focus on the sources of variation of institutions.¹⁷ Acemoglu et al. used variation in institutions within former European colonies arguing that historical mortality faced by Europeans and density of indigenous population were important sources of variation in the types of institutions that emerged.

Acemoglu, Johnson, Robinson, and Yared (2008) show that these same historical variables also predict long-run democracy in addition to economic growth.

The general picture of institutional variation developed in these papers is that the institutional organization of society has a strong tendency to persist. It is implicit in the discussion, so far, that there are large forces which generate persistence in the political economy equilibrium. If a certain group is empowered by the existing structure of political power then they will choose economic institutions that favor them. This will increase their wealth and be one channel through which their *de facto* power will persist or increase. Moreover, those who hold power today will not only be able to determine economic institutions today. They will also be able to determine political institutions in the future which tends to cement the *de jure* power of such a group. Hence, the distribution of political power and thus economic institutions is naturally highly persistent over time.

The tendency for institutions and patterns of governance to persist can also be illustrated by returning to the observations of de Tocqueville we quoted in the introduction. Let us use “inequality” as a catch-all word for a wide range of economic and social disparities. What Tocqueville hinted at can then be visualized as a political and economic system where (A) initial inequality affects individual and organizational behaviors—the selection of preferences of those who govern, and the possibility to implement different policies toward those who are governed; (B) the selected behaviors and policies affect the outcome and the social and economic inequality which form the initial conditions (A') for a new round of political and economic behavior. The dynamics of the system loosely described by (A)–(B)–(A') can clearly be path dependent. While a high level of inequality can generate policies and governance that induce or maintain inequality, a more egalitarian starting point can generate policies and governance that maintain this low inequality. The system is in a long-run equilibrium whenever initial inequalities are reproduced.

Nevertheless, institutions and governance can change. There may be intrinsic dynamics or shocks to the system that lead to redistributions of power and thus changes in economic institutions. It may also be the case that the incentives of those in power are changed, perhaps because of technical innovation or new market opportunities, and this may change economic institutions as well.

This perspective suggests, however, that while small changes may cumulate into systemic changes in institutions, the more usual case seems to be that at large “critical junctures” the institutions of society are moulded and after that tend to be highly inertial. This was why the current institutions of Latin America, for example, look so different from those of North America. It was the apparent importance of such critical junctures which led Acemoglu et al. (2001) to focus on the institutional consequences of European colonialism.

An interesting way to see the importance of the emphasis on critical junctures is to consider more broadly the consequences of factor endowments for development.

Engerman and Sokoloff (1997) and Acemoglu et al. (2001) show how factor endowments at the time of colonization had large effects on institutions. But what about the impact of subsequent changes in factor endowments, such as the role of oil in the twentieth century? Some authors, such as Sachs and Warner (1995) blame bad economic performance on the abundance of such resources. Nevertheless, as Mehlum, Moene, and Torvik (2006) show, what matters is the combination of resource abundance and institutional quality. The growth effect of resource abundance varies with institutional quality. In countries with good institutions, resource abundance attracts entrepreneurs into production. In countries with weak institutions, however, entrepreneurs are diverted away from production and into rent appropriation. The historically determined institutional path that a society is on, therefore, subsequently conditions what happens when shocks, such as the discovery of oil, arise.¹⁸

There are many potential sources of institutional variation in the world and to understand why Western Europe became so successful economically compared to Eastern or Southern Europe one would need to appeal to very different historical events. Scholars who have focused on the emergence of capitalist institutions in Western Europe, have focused on such systematic factors as the rise of the mercantile economy (Pirenne, 1937), the differential response of institutions to the population collapse of the Black Death (Postan, 1944), different patterns of social conflict (Brenner, 1976), the great shock of 1492 and the expansion of European powers into the world (Acemoglu et al., 2005a; Pomeranz, 2000; Williams, 1944), or the French Revolution (Acemoglu, Cantoni, Johnson, & Robinson, 2008). Institutions may also evolve in ways which depend on technology, as emphasized by White (1962), or even because of religious conversion (Pirenne, 1939).

So understanding institutional variation necessitates an explicitly historical approach. Two good examples of this come from the work by Putnam (1993) and Evans and Rauch (1999, 2000). Putnam (1993) famously proposed that what determined the measures of government performance discussed earlier was the relative amount of social capital of different Italian regions. In Figures 10 and 11, we use his data on the number of inhabitants per cultural or recreational association in 1985. These associations include choral societies, hiking clubs, and bird watching groups. We plot this data against two of the variables we introduced earlier. Figure 10 shows the density of associational life to be negatively associated with the mean budget delay while Figure 11 shows it to be positively associated with the quality of interaction with the bureaucracy.

Why does social capital vary? Here, Putnam argued that this had deep historical roots related to the imposition of feudalism in the south of Italy in the eleventh and twelfth centuries after it was invaded by the Normans compared to the development of self-governing communes in the north as the power of the Holy Roman Empire declined (see Guiso, Sapienza, & Zingales, 2008, for an attempt to test this idea).

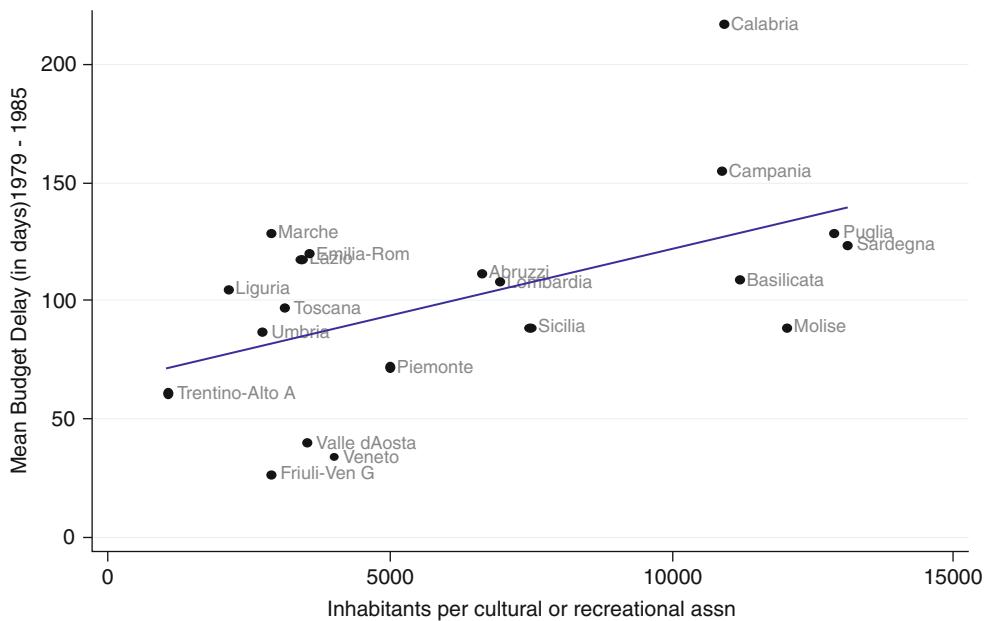


Figure 10 Social capital and budget delay in Italy.

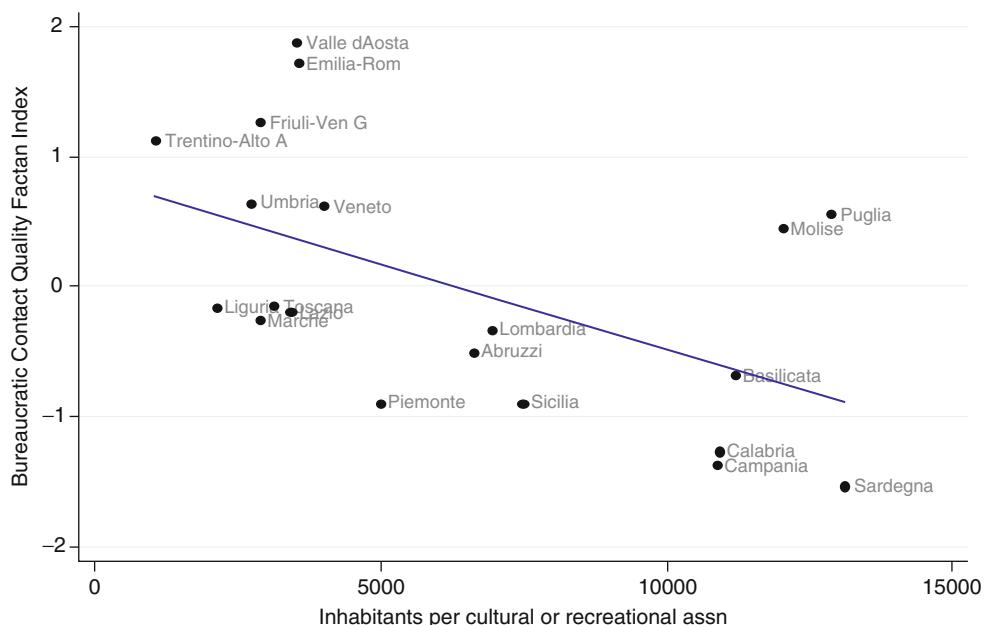


Figure 11 Social capital and bureaucratic quality in Italy.

Another important example both of measurement and the historical evolution of institutions comes from the research of Evans and Rauch (1999, 2000). They conducted a survey of experts to collect information on the structure of the bureaucracy in 35 countries. This project was an attempt to measure the extent to which the bureaucracy was “weberian.” Figure 12 plots the data they collected on how meritocratic the bureaucracy was against the World Bank’s government effectiveness index. This shows that the greater the extent to which meritocratic procedures are used, the more effective is the government. The variable was constructed by averaging the replies to the following two questions: (Evans & Rauch, 1999, pp. 55–56) “Approximately what proportion of the higher officials in these agencies enter the civil service via a formal examination system?” and “Of those that do not enter via examinations, what proportion have university or postgraduate degrees?” Figure 13 instead plots against government effectiveness an index of how “weberian” the state is from Evans and Rauch (2000) which uses information from 10 different questions. Again the weberianess of the state is positively correlated with the effectiveness of the government.

Why is it that some countries have weberian states while others so not? Evans (1995) suggests that this is the outcome of deep historical processes. For instance, he argued that the capable bureaucracy in South Korea was related to the historical organization of the bureaucracy and examination system.

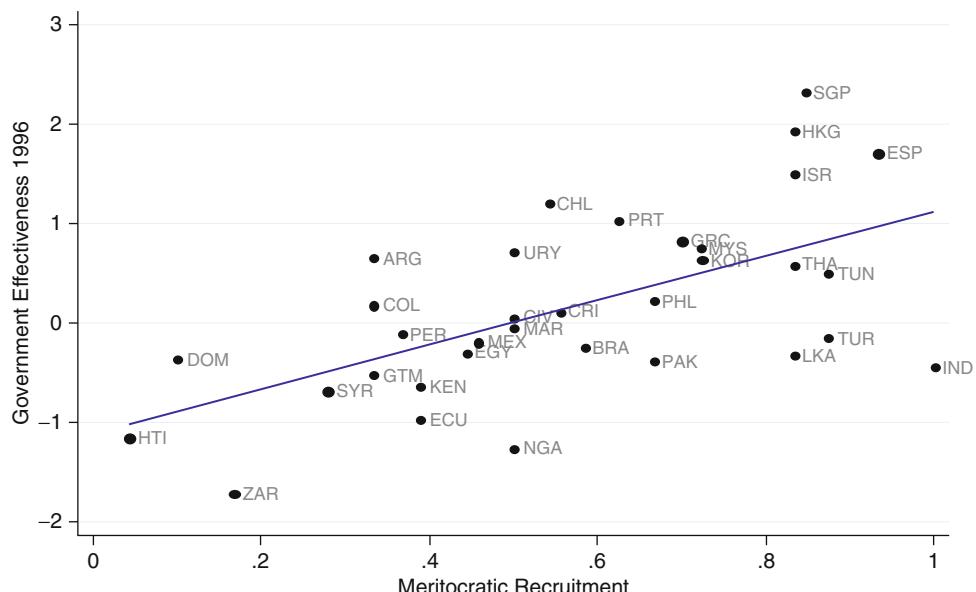


Figure 12 Meritocratic recruitment and government effectiveness.

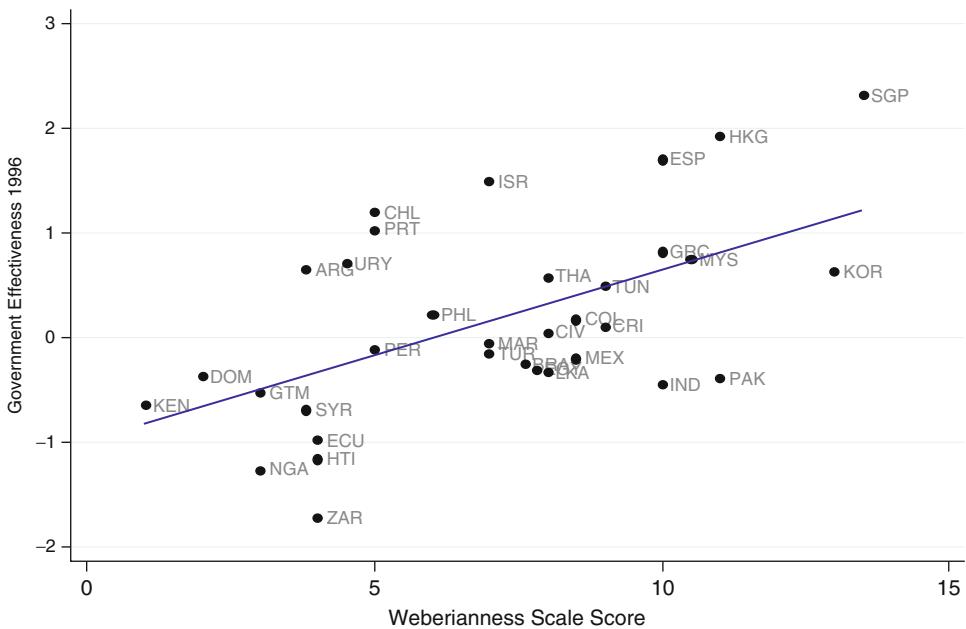


Figure 13 “Weberianness” and government effectiveness.

Our emphasis here is not on the path dependence of particular institutions, though we would not deny this can be important, but rather on the path dependence and historical roots of the entire political equilibrium of which bureaucratic capacity and governance more generally are two parts. In Sierra Leone, for example, we emphasized that the poor performance of the bureaucracy was closely tied to the “strategy of rule” adopted after independence. More generally there is a close relationship between clientelism and bureaucratic capacity. Indeed, the connection between clientelism and poor governance emerges from Putnam’s data. One such example is given in Figure 14. Here, we plot the incidence of innovative laws, discussed earlier, against the % of preference votes cast as a % of total possible preference votes in the 1976 Italian national election. Voters in Italian elections at that time voted for a single national party list with seats being allocated in the legislature by proportional representation. However, if people wished, they could also indicate a preference for a particular candidate on that list. Such preference voting is associated with clientelistic exchanges. Figure 14 shows that clientelism thus measured is inversely related with legislative innovativeness.

5.3 Origins of Bad Governance in Sierra Leone

5.3.1 General ideas about patrimonialism

We now return to the Sierra Leone experience and ask: what do we know about what causes patrimonialism? We first do this at a more general level and in the next subsection focus on how these general arguments apply in Sierra Leone. There are two sides

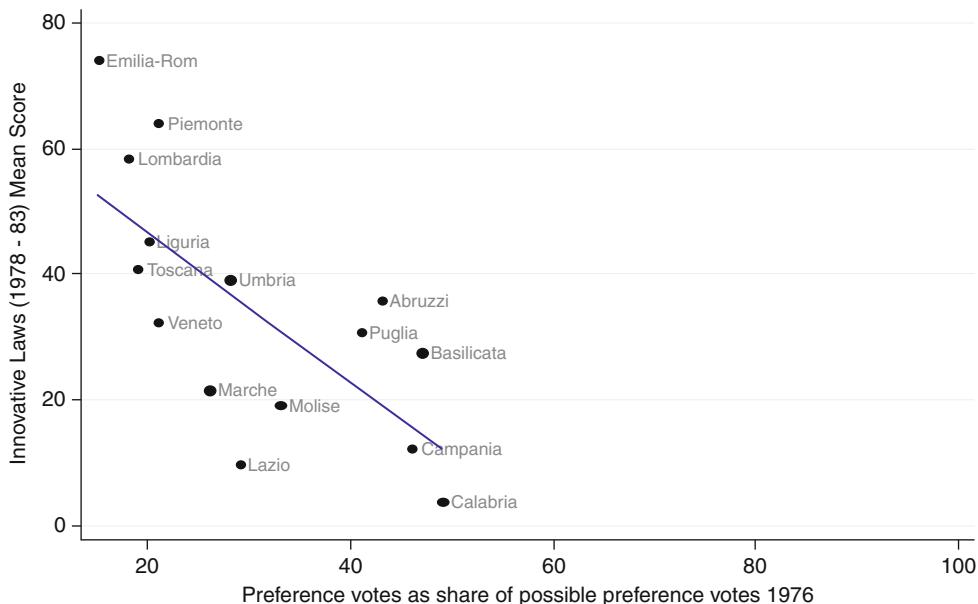


Figure 14 Clientelism and innovative laws in Italy.

to this question, the supply of patrimonialism and the demand. Almost the entire academic literature focuses on the supply side, namely the circumstances under which patrimonialism is an attractive strategy for politicians to use to stay in power. The demand side, the circumstances under which citizens demand patronage instead of, say, public goods, is neglected though may be important.

There are some basic ideas about what causes the supply of patrimonialism. First, it seems to be attractive in circumstances where national identities are lacking and there is no settled social contract or legitimate structure of authority in the country. This was obviously quite a common circumstance in Africa after independence where countries inherited the arbitrary boundaries created by European colonialism.¹⁹ Though we do not have good analytical models of this, Zolberg's (1966) insights about "creating political order" still seem germane today. Patrimonialism may have great social costs but it is a very effective way of governing in a fissiparous nation.

Second, patrimonialism seems to be encouraged by heterogeneity of identities. There seem to be a few reasons for this. One is that patrimonialism heavily uses various forms of divide and rule and this strategy appears to be much more feasible in societies have distinct ascriptive identities (Padró-i-Miquel, 2007). Another stems from the demand side. In a very heterogeneous society there may be much less demand for public goods, which reduces the opportunity cost of the massive undersupply of public goods which is the natural by-product of patrimonialism. Another facilitating factor

on the demand side may stem from the fact that in a society divided by ascriptive differences it is easy to conceive of a polarization between different groups which may lead people to be disposed to reject the application of universal rules, so undermining principles which would impede the creation of patrimonialism. Heterogeneity of identities also seems to make it more difficult to establish civilian control of the military. When the military cannot be relied on, it becomes very difficult to establish such key state functions as monopolizing violence and this again makes patrimonialism more attractive as a method of political control (see Horowitz, 1985, on how civilian control of the military is fraught in ethnically divided societies).

The third potent source of patrimonialism seems to be a type of path dependence from the institutional structure and operation of the colonial state. A large amount of recent empirical and historical work has emphasized the path-dependent influence of colonial institutions (e.g., Acemoglu et al., 2001; Engerman & Sokoloff, 1997), and this work resonates with arguments made by Africanists. Cooper (2002), for example, coins the term “gate-keeper state” to refer to the institutions the Europeans created. They built governing centers on the coasts and constructed infrastructure and institutions only to the extent that this was needed to extract valuable resources or provide the thinnest veneer of order. Cooper’s argument suggests that center-periphery relations in postindependent African states were heavily influenced by the structure of the colonial state. Young (1994) specifically argues that the political strategies adopted by post-colonial African regimes were directly modeled on the way that colonial states functioned. For example, indirect rule practiced in British colonies precipitated indirect rule by postindependent governments.²⁰

It is also possible to make path-dependent arguments about the creation of state institutions more generally. The institution of indirect rule was the antithesis of the construction of effective central state institutions and the development of what Mann (1986) calls “infrastructural power.” Indirect rule was not a strategy for building state institutions, rather it was a minimum cost way of creating order which gave up the desire of “penetrating” society so as to provide public goods or structure public life, in order to focus on resource extraction. As Fanthorpe (2005, p. 4) points out in the case of Sierra Leone,

even today, the vast majority of rural Sierra Leoneans obtain primary rights of residence, land use, and political/legal representation as “natives” of chiefdoms rather than as citizens of the state.

Here, the academic literature builds firmly on the work of those like Hintze (1975) and Tilly (1975, 1990) who studied the development of central state institutions in Europe. These scholars emphasized how fiscal and tax institutions, bureaucracy and political institutions such as parliaments evolved out of the desire of states to raise resources to fight interstate wars. Herbst (2000) and Bates (2001) specifically attribute

the poor capacity of African states to the absence of the historical factors that led to them in Europe. To these scholars, as to Max Weber, patrimonialism was the natural state of affairs before the creation of states and “rational-legal authority” and they see no natural tendency for patrimonialism to change without external challenges.²¹

Though the deep roots of the phenomenon are therefore complex, what is clear is that the strategy is complementary to certain types of political institutions. Throughout Africa, the creation of patrimonial rule went hand in hand with changes in political institutions. This involved a strengthening of the executive which was achieved by a move away from parliamentary to presidential constitutions. One can think of several reasons why being a president was more attractive than being a prime minister for politicians such as Stevens, but there seems to be a natural affinity between presidentialism and patrimonialism. It also involved a reduction in accountability which was implemented by the suspension of democracy. During the consolidation of patrimonialism, power was also concentrated in the central state and capital city.

5.3.2 The Sierra Leonean context

Some of these ideas obviously apply to Sierra Leone. Like most Sub-Saharan African countries, Sierra Leone was “arbitrary” and, the Western Areas aside, made up of a patchwork of different precolonial states and polities with different languages, histories and cultures. At independence, there was probably little notion of a national identity or agreed rules for structuring the contest for power and this no doubt led different groups to perceive that the stakes were very high.

Sierra Leone also seems to fit well with ideas of the gate-keeper state’ centered on Freetown and of course the Protectorate was ruled indirectly through a system of chiefs institutionalized by the British in the 1890s.²²

Civilian control of the military was also highly problematic and Sierra Leone even appears as a case study in Horowitz (1985) of a society where the colonial military recruitment policies gave postindependent civilian governments severe problems. For instance, the British recruited primarily from the South of the country and this is one of the reasons why the Mende dominated officer corps opposed the election of the APC in 1967 and mounted a coup to block the transfer of power. Though Stevens quickly moved to replace Mende officers with northerners, his primary strategy was to emasculate the military, something which proved a disaster when the RUF invaded.

This being said, the intensity of patrimonialism in Sierra Leone might not have been anticipated in 1961. For one, unlike most Sub-Saharan African countries it actually had a national language, Krio, something which is associated with political stability in Tanzania, or with very good development outcomes in Botswana.²³ Second, though there were distinct ethnicities in Sierra Leone, these were not based on the type of socioeconomic differences seen in Rwanda or Burundi and at no point does the civil war seem to have degenerated into anything approximating an “ethnic conflict.”²⁴

Herbst (2000) even classifies Sierra Leone into one of the rare African countries with “easy” political geography based on its small size and distribution of population.

Despite these advantages, there seem to have been some other factors that may have exacerbated the intensity of patrimonialism in the country. Most commonly cited is not the extent of natural resources but the *form*. It is widely observed that the alluvial nature of diamonds and their widespread presence across the country makes it very difficult for the state to control diamond mining. This breeds illegality, smuggling, and evasion and as Clapham (2003, p. 12) puts it

a situation highly unconducive to the maintenance of legal norms and bureaucratic state structures.

It seems quite likely that the failure of the British colonial state and postindependence governments to bring the diamond mining economy under the control of formal state institutions helped to stimulate different ways to control it and extract wealth from it. The contrast with the deep mined capital intensive diamonds of Botswana is interesting.²⁵

Another factor often quoted in favor of Sierra Leonean distinctiveness is an undertow of socioeconomic and political conflicts possibly missing in most African countries. For example, Richards (1996), Sawyer (2004), and Fanthorpe (2005) see historical roots of the civil war stemming from inequitable patterns of access to land and resources and many have written about the extent of “gerontocracy” in Sierra Leone and the grievances this created among young men. It is possible that the existence of a large strata of disillusioned young men also facilitated the creation of patrimonialism since excluded from local resources, they may have valued very highly what rents were on offer from the political patrons in Freetown.

It may also have been that an added incentive to undermine the bureaucracy and create a shadow state' in the postindependence period came from the fact that the state inherited from the colonial powers was dominated by Krios, while politics was dominated by people and interests from the former protectorate (Clapham, 2003, p. 12).

Finally, one could also argue that the fact that the Mende and the Temne, each with about 30% of the population, formed the basis of the two main political parties gave politics a type of “polarized structure” in the sense of Esteban and Ray (1994) which may have been very conducive to conflict.

6. IMPROVING GOVERNANCE

Our discussion in this chapter has embedded the discussion of governance into the political economy of institutions and development. Our basic argument is that bad governance arises because it distributes income and or political power in particular ways. To improve governance, we need to understand the political forces that lead to bad governance in the first place and we also need to understand how and why

governance improves. These questions are closely related to more general ones about the processes that lead to institutional persistence and change. Institutions persist when the incentives and structures of power that led to them also persist. It is clear that even though there is a great deal of persistence of institutions in the world, institutions do change. The approach we have taken in this chapter suggests that this will happen when the political equilibrium changes.

Nevertheless, at present, we do not have a satisfactory understanding of the circumstances under which dysfunctional political equilibria arise and sustain themselves. A natural idea would be to focus on specific political institutions such as democracy. Yet we know that democracy *per se* is not necessarily associated with better development outcomes and we all know the famous examples of “developmental dictatorships” such as in South Korea or Taiwan. However, as yet, we do not understand why some dictatorships are developmental and others not or why, for instance, there has never been a developmental dictatorship in Sub-Saharan Africa or Latin America.

Though we cannot yet say under what circumstances political equilibria which lead to economic growth will arise, we can illustrate the utility of the ideas we have developed by examining the issue of governance and institutional reform. If governance is poor then a natural approach is to directly try to reform governance. If there is corruption, then promote an anticorruption law, set up an anticorruption agency, make aid conditional on eliminating corruption. The potential problems facing such an approach highlight the first set of *pitfalls of reform*. Our perspective emphasizes that one should not try to understand or manipulate governance without thinking about the political forces that created the particular patterns of governance that we observe.

Indeed, we shall argue that direct governance reform in itself is unlikely to be effective and that instead it might be more useful to focus on understanding and reforming the forces that make governance bad. It is, therefore, important to focus on political institutions and the distribution of political power in thinking about reform. This raises the second potential pitfall of governance reform; while we have recognized the importance of political institutions, we are still at the beginning of understanding the complex relationship between political institutions and the political equilibrium and thus governance. Sometimes changing political institutions may be insufficient, or even counterproductive, in leading to better economic outcomes.

The pitfalls of reform are related to the fact that patterns of relative economic performance are very persistent. Examining the pitfalls of reform is one way of approaching this issue. We then move to examining successful change.

We begin our discussion by focusing at more length on whether reforming governance without thinking about politics is likely to be effective. We argue that such reforms may not work if they do not change the political equilibrium. We then examine if these pitfalls of reform can be solved by reforming political institutions (thus, altering the distribution of *de jure* power in society). We argue that this may not work

either because *de facto* power may persist or may override the effects of reforms to political institutions. Thus, it might seem to follow that a successful reform necessitates changes in both *de jure* and *de facto* power. We show that simultaneously changing both may not achieve real reform either because the political equilibrium may be path dependent.

6.1 Persistence of power and incentives—The See-Saw Effect

Many poor growth experiences are accompanied by a system of dysfunctional laws and regulations and other aspects of governance. An obvious idea might be to directly intervene in these components of governance and promote change in laws and regulations. This was the sort of reasoning that led to the famous Washington Consensus some of whose components, for example, privatization of state enterprises, deregulation, and legal security for property rights all seem related to governance.

The first pitfall of reform is that directly reforming specific institutions, policies, or aspects of governance may not be sufficient, and may even backfire. The reason why such reforms may be ineffective is that it is usually not a coincidence that some aspect of governance is bad. Bad governance is probably fulfilling some political objective. But there are many different ways and a multitude of instruments to achieve a specific goal. Taking away one instrument without altering the balance of power in society or the basic political equilibrium can simply lead to the replacement of one instrument by another with little net effect of the ultimate goal—economic performance. This phenomenon was dubbed the *See-Saw Effect* by Acemoglu, Johnson, Robinson, and Thaïcharoen (2003).

6.1.1 Case study: Central bank independence

Acemoglu, Johnson, Querubín, and Robinson (2008) use the case of central bank independence (CBI) to illustrate these ideas about reform. CBI is a canonical example of a reform in governance which was proposed as a way to improve monetary policy. It is a particularly interesting type of reform to study empirically since it has a clearly delineated target—*inflation*—and one can thus judge the success of CBI by investigating whether or not it has reduced inflation. Acemoglu et al., in the spirit of the argument earlier, argue that the impact of reform will be conditional on the initial political equilibrium that generated the need for reform. For example, only policy reforms that the groups with political power cannot easily override, circumvent, or ignore are likely to achieve their objectives. In consequence, in many countries, such as those in Sub-Saharan Africa, where there are only few constraints and checks on politicians and on politically powerful groups, policy reform is unlikely to be very effective. Case study evidence for this is presented in van de Walle (2001, see also his 1993 paper) who illustrates that for African politicians (2001, p. 13)

restoring economic stability and growth has often taken a back seat in government motivations to preserving political power.

In the context of structural adjustment he argues (2001, p. 76)

Often, the policies have changed on paper, but in practice, something resembling the status quo ante continues to prevail. In some cases, the old policies were reinstated under a new name or with some new policy objective... In other cases, governments ignore the spirit of their own liberalization efforts by continuing to interfere in officially deregulated markets.

Put starkly, who would expect policy reform to have significant effects in Zimbabwe as long as Robert Mugabe is in power? As an illustration of this, Figure 15 plots the inflation rate in Zimbabwe and plots a vertical line in 1995 when the central bank's act was modified in order to grant the Reserve Bank of Zimbabwe more independence. Clearly, the independence of the central bank in Zimbabwe did little to restrain the subsequent monetary policy of the government. Therefore, a major reason why policy reform will often fail is because of the absence of a functioning accountability system and lack of constraints and checks on politicians.

Acemoglu et al. (2008) point out, however, that this does not imply that better political institutions and transparency always increase the impact of reforms. In particular, we would not expect societies with a functioning system of accountability and with checks on politicians to be those pursuing such highly distortionary policies. For example, inflation was already low in the United Kingdom in the 1990s, before

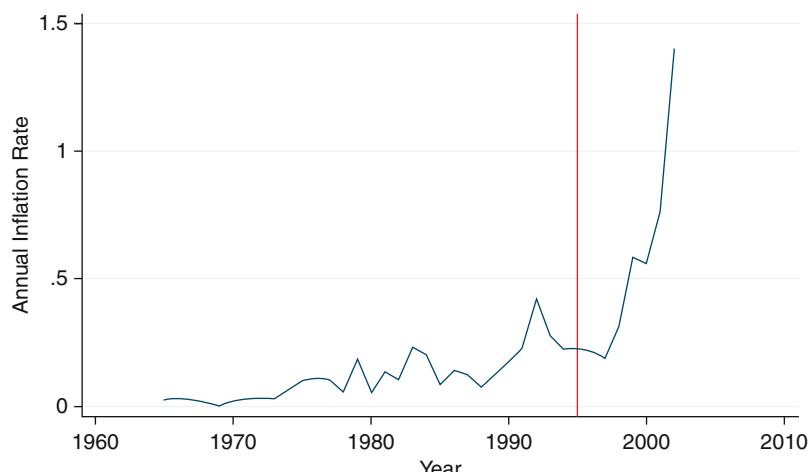


Figure 15 Central bank independence and inflation in Zimbabwe. Note: Annual inflation rate corresponds to the annual variation in the consumer price index reported in the International Financial Statistics (IMF). Vertical line shows the year of Central Bank reform taken from Polillo and Guillen (2005).

the central bank became independent in 1998 and in consequence the potential for a large effect from CBI reform is limited.

The arguments in the previous two paragraphs suggest that once we take a political economy perspective, there should be a nonmonotonic relationship between the extent of constraints on politicians and the impact of policy reform. Acemoglu et al. (2008) then test these ideas using cross-national panel data. They find that CBI has little effect on inflation in countries with either high or low values of constraints on the executive, but it does significantly reduce inflation in countries with intermediate values of constraints. However, they also hypothesize that if policy reform does not change the political equilibrium, reform in one dimension can lead to simultaneous dis-reform in another dimension. They test this idea by examining whether or not CBI leads to deterioration in fiscal policy. They indeed find some evidence that while in countries with high or low constraints on the executive CBI has no impact on fiscal policy, for countries with intermediate levels of constraints it tends to increase the size of the government relative to GDP. This is a precise instance of the *See-Saw Effect*. We illustrate this idea in Figure 16 with two figures from their paper which plot for Argentina and Colombia inflation and government expenditure as a % of GDP. Again the introduction of CBI is indicated by a vertical line. The figures suggest that while CBI reduced inflation, it also coincided with a large expansion in the size of the government relative to the economy.

6.1.2 General lessons

Making or imposing specific reforms may have little impact on the general structure of governance or performance if they leave untouched the underlying political equilibrium. Of course, it is possible that a reform in governance may induce a change in power and ultimately in the broader political equilibrium. Nevertheless, as the above example make clear, this is far from certain.

Despite all of the Washington Consensus reforms that took place in Argentina, for example, there was little change in the way politics worked. The political genius of Menem and the Peronist party after 1989 was to recognize that the policies of the Washington Consensus could be bent to function as “politics as usual.” In consequence, there was little change in the underlying political equilibrium though the instruments which the Peronists used after 1989 were different. Though it is possible that such reforms could change the political equilibrium, it did not happen and this is why the Washington Consensus led to such disappointing results.

6.2 Persistence of *de facto* power

The last section illustrated that reforming particular parts of governance without perturbing the underlying political equilibrium may not lead to improved economic performance. Moreover, we shall now argue that even reforming *de jure* power

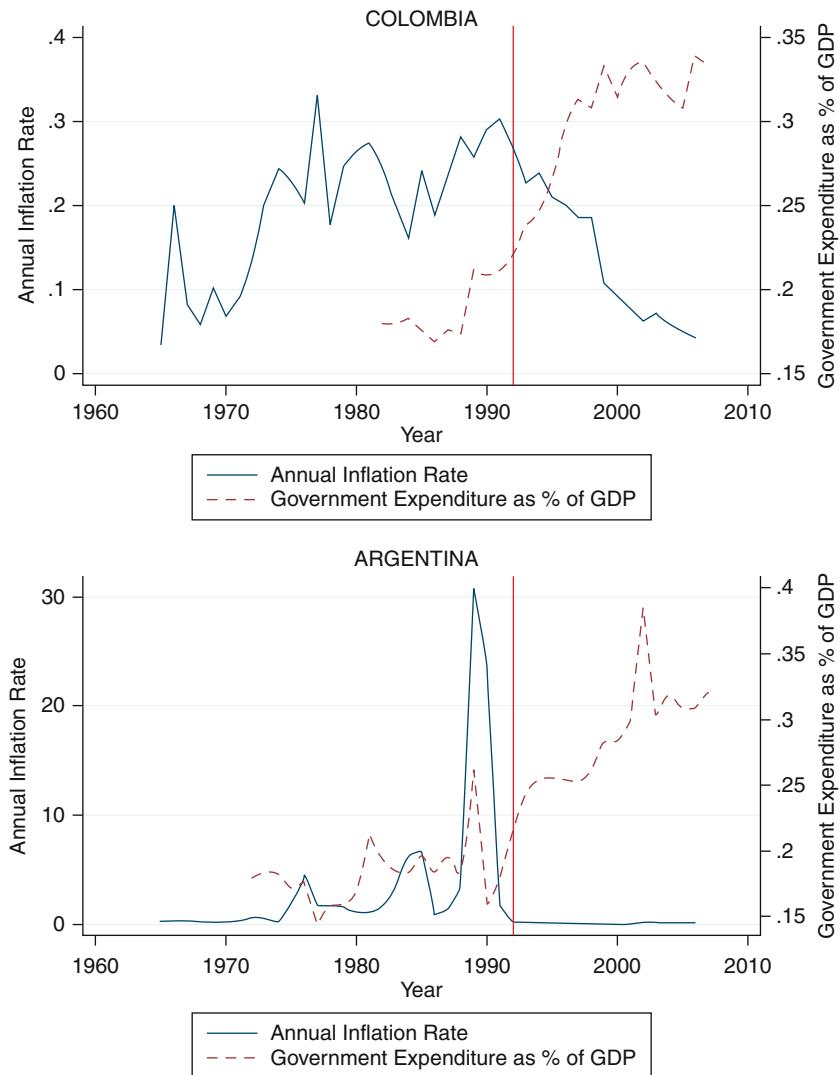


Figure 16 The See-Saw Effect in Colombia and Argentina. Note: Annual inflation rate (scale on the left vertical axis) corresponds to the annual variation in the consumer price index reported in the International Financial Statistics (IMF). Government expenditure data (scale on the right vertical axis) are from World Economic Outlook IMF. Vertical line shows the year of major Central Bank reform taken from Jácome and Vasquez (2005).

(for instance, enfranchising former slaves) or introducing democracy may not be sufficient to induce broader institutional change. The reason why changes in *de jure* power may not be sufficient to trigger a change in the political equilibrium is that the political and economic system is kept in place by a combination of *de jure* and *de facto* political power. An external or internal reform of *de jure* institutions may still leave the

sources of *de facto* power intact, and groups that have lost their *de jure* power may use their *de facto* power in order to recreate a system similar to the one that has departed (Acemoglu & Robinson, 2008a). The new system may be as inefficient as the old one.

This is not to argue that reform of *de jure* institutions is not possible or that it is irrelevant. For example, democratization in many European societies in the nineteenth century appears to have significantly changed economic institutions, for example, leading to sustained expansions of educational systems (Acemoglu & Robinson, 2000; Lindert, 2004). It is to argue, however, that reform comes with pitfalls and as yet we have an incomplete understanding of the circumstances under which such reforms will succeed.

An excellent illustration of these ideas is the evolution of institutions in the US South after the Civil War. Our discussion follows Acemoglu and Robinson (2008b).

6.2.1 Case study: The US South before and after the Civil War

An important example which illustrates our perspective is the continuation of the economic system based on labor repression, plantation, and low-wage uneducated labor in the US South before and after the significant changes in institutions brought about by the Civil War. Most obviously, these changes in *de jure* power included the enfranchisement of the freed slaves.

Before the Civil War, the South was significantly poorer than the US average income at about 70% of GDP *per capita*. The South lacked industry (Wright, 1986, table 2.4, p. 27) and in 1860 the total manufacturing output of the South was less than that of either Pennsylvania, New York, or Massachusetts (Cobb, 1984, p. 6). The South had very low rates of urbanization (around 9% as opposed to 35% in the Northeast) and relatively little investment in infrastructure. For example, the density of railroads (miles of track divided by land area) was three times higher in the North than in Southern states. The situation with respect to canal mileage was similar (Wright, table 2.1, p. 21). Perhaps more importantly, especially in the context of the potential for future economic growth and industrialization, the South was not even innovative in the sectors in which it specialized. The relative backwardness of the South was due to the plantation economy and slavery.

In the aftermath of the Civil War, the income *per capita* of the South fell to about 50% of the US average. If the organization of the slave economy had been the reason why the South had been relatively backward in 1865, one might have imagined that the abolition of slavery in 1865 would have removed this blockage to Southern prosperity. The evidence and historical interpretations show that the abolition of slavery had a surprisingly small effect on the Southern economy. Out of the ashes of the Civil War emerged a low-wage labor intensive economy based on labor repression. Cut off from the rest of the United States, income *per capita* remained at about half the average

until the 1940s when it finally began slowly to converge. Just as before the Civil War, there was systematic underinvestment in education (Margo, 1990).

So why did the economic system of the South change so little following the Civil War, especially given the significant changes in political institutions? At first, this persistence appears at odds with the significant changes in the distribution of *de jure* power that took place after the Civil War, for example, with the enfranchisement of the freed slaves, and the repeal of the Missouri compromise, which had previously cemented the political power of the South in the federal government.

We argue that this persistence is due to the exercise of *de facto* political power by the Southern landed elites to compensate for the loss of their *de jure* political power. There was considerable persistence in the identity and power of these elites. For example, Wiener (1978) studied the persistence of the planter elite in five counties of the black belt of western Alabama. Tracking families from the US census and considering those with at least \$10,000 of real estate, he found that (p. 9) “of the 236 members of the planter elite in 1850, 101 remained in the elite in 1870.” Interestingly, this rate of persistence was very similar to that experienced in the antebellum period; “of the 236 wealthiest planters families of 1850, only 110 remained in the elite a decade later” (p. 9). Nevertheless, “of the 25 planters with the largest landholdings in 1870, 18 (72%) had been in the elite families in 1860; 16 had been in the 1850 elite group.”

After the end of the Civil War, more or less the same group of planter elites controlled the land and used various instruments to reexert their control over the labor force. Though the specific economic institution of slavery did not persist, the evidence shows a clear line of persistence in the economic system of the South based on plantation-type agriculture with cheap labor. This economic system was maintained through a variety of channels, including both control of local politics and exercise of potentially violent *de facto* power. As a consequence, in the words of Du Bois (1903, p. 88), the South became “simply an armed camp for intimidating black folk.”

A key to the persistence of the antebellum system after the Civil War was the continued control over land. For example, in the debate over the redistribution of 40 acres of land to the freedmen (vetoed by President Andrew Johnson in 1865), Congressman George Washington Julian argued (quoted in Wiener, 1978, p. 6):

Of what avail would be an act of congress totally abolishing slavery . . . if the old agricultural basis of aristocratic power shall remain?

Southern elites were able to disenfranchise blacks after 1877 and a whole gamut of segregationist legislation—the so-called Jim Crow laws—was enacted (Woodward, 1955, for the classic analysis). These laws turned the postbellum South into an effective “apartheid” society where blacks and whites lived different lives. As in South Africa, these laws were aimed at controlling the black population and its labor supply. Consequently, the South entered the twentieth century as a primarily rural society.

"It remained an agrarian society with a backward technology that still employed hand labor and mule power virtually unassisted by mechanical implements" (Ransom & Sutch, 2001, pp. 175–176). In 1900, the South's urbanization rate was 13.5%, as compared to 60% in the Northeast (Cobb, 1984, p. 25).

Ransom and Sutch's (2001, p. 186) assessment of the implications of this economic and political system in the South for economic progress is representative of the consensus view: "Southerners erected an economic system that failed to reward individual initiative on the part of blacks and was therefore ill-suited to their economic advancement. As a result, the inequities originally inherited from slavery persisted. But there was a by-product of this effort at racial repression, the system tended to cripple all economic growth":

All in all, the Southern equilibrium, based on the exercise of *de facto* power by the landed elite, plantation agriculture and low wages, and uneducated labor, persisted well into the twentieth century, and only started to crumble after World War II. Interestingly, it was only after the demise of this Southern equilibrium, that the South started its process of rapid convergence to the North.

6.2.2 General lessons

Just as reforming governance or economic institutions without changing the political equilibrium may not improve economic performance, so changing *de jure* power, while leaving the sources of *de facto* power intact, may have little impact. In the US South, the same economic system based on the repression of labor got reinstated after reconstruction. Even though the enfranchisement of the freed slaves meant that there had been a change in *de jure* power, and after the Civil War blacks exercised this power and voted in large numbers, southern elites were able to use their *de facto* power to reassert control over labor and eventually by the 1890s disenfranchise the blacks. The persistence of *de facto* power was facilitated by the fact that white elites had kept hold of the land after the Civil War, and because these elites had avoided being killed during the Civil War and still had a huge comparative advantage over blacks in the ability to engage in collective action. Control was exercised via coercion, lynching, and the Ku Klux Klan and other extralegal methods and eventually institutionalized via control of state legislatures.

The general lesson seems to be that change in institutions which affects the distribution of *de jure* political power, needs to be complemented by changes in the sources of *de facto* political power of the elite and reductions in the benefits that political incumbents have in intensifying their use of *de facto* political power.

6.3 The Iron Law of Oligarchy

The conclusion from the last section seems to be that to change the political equilibrium there needs to be changes in both *de jure* and *de facto* power. For instance, if there is an elite which is structuring institutions to its benefit with adverse aggregate effects,

then to engineer a transition to a better equilibrium both their *de jure* and *de facto* power must be simultaneously reformed.

Unfortunately, things are not quite so simple as this. This is because even if *de jure* and *de facto* power changes, those who acquire the power in the new political equilibrium may not have the correct incentives either. More importantly, their incentives to use their power and the institutions they find it optimal to create may be fundamentally shaped by the status quo they replace—they may be path dependent. If an elite with power is initially structuring economic institutions or the nature of governance to extract rents from society, then the very fact that it is doing this may induce a new elite to do likewise. The replacement of one elite by another may, therefore, do little to improve economic performance. This pitfall is reminiscent of the classic idea in sociology of an *Iron Law of Oligarchy* going back to the work of Michels (1962), Mosca (1939), and Pareto (1968). This hypothesis states that it is never possible to have real change in society because when new groups mobilize or are created in the process of socioeconomic change they simply replace preexisting elites and groups and behave in qualitatively similar ways. There seem to be many circumstances in which “Iron Law” type behavior may occur and there are quite possibly many mechanisms that can generate behavior like this. We now present a case study of this phenomenon in action drawn from Acemoglu and Robinson (2007).

6.3.1 Case study: The Bolivian revolution

Bolivia features centrally in accounts of comparative development in the Americas. It was at the heart of the Inca Empire with a high density of indigenous peoples and during the colonial period economic institutions designed to extract rents—the encomienda, repartimiento, the Potosí mita (forced labor draft for the silver mines)—were all central. Although the mita was abolished at independence, a highly inegalitarian and authoritarian society persisted. In 1950, for example, 6% of landowners owned 92% of all lands and the smallest 60% of landowners owned 0.2% and the tin mines which formed the basis of the export economy were owned by three families. A mere 31% of the adult population was literate and only 4% of labor force was employed in industry. Indians still subject to unpaid *pongueaje* (personal services) for the landowners whose lands they worked (Klein, 1992, for an overview of this evidence).

The remains of this system were swept away by the Bolivian revolution of 1952 which was masterminded by the MNR (Movimiento Nacionalista Revolucionario), a political party which had formed in urban areas in the 1940s to contest the power of the traditional elite. Following the Revolution, the MNR formed a government which implemented land reform, expropriated large estates, and redistributed them to the labor force and Indian communities. It also introduced universal suffrage by abolishing literacy requirement on voting and nationalized the mines of the tin barons.

These appear to be huge, radical institutional changes. In particular, there was a shift in the distribution of both *de jure* and *de facto* power. Surely, Bolivia was launched on a new path of institutional and economic development. At the very least one would have anticipated a sustained fall in inequality. Unfortunately, none of these good outcomes occurred. Following the Revolution, the 1950s saw a failed attempt by MNR to create one-party state and in the process they rebuilt the military which had been disarmed in 1952. They were also able to use clientelism to gain the support of the indigenous majority. Indeed, there are striking comparisons between the traditional clientelism which had existed before 1952 and that which emerged during the regime of the MNR afterward. In a seminal study, Heath (1972) showed that although the identity of the patrons were different and the instruments of clientelism had changed following the institutional changes brought by 1952, there were very strong similarities in the basic structure of the political equilibrium. Kelley and Klein (1981) estimated that 10 years after the Revolution, inequality had returned to 1952 levels.

How can we understand an outcome like this? We believe that there are mechanisms that can generate persistence in the political equilibrium even when *de jure* and *de facto* power changes and can produce an Iron Law of Oligarchy. The idea is quite simple. Initially in Bolivia institutions were structured to the benefit of traditional elites. A new elite emerged, spear-headed by the MNR. The MNR needed to win support of the campesinos and other urban groups. To do this, they had to develop a political strategy, but the form that strategy took was highly influenced by the strategies being used by the traditional elite. The traditional elite were clientelistic, so it was optimal to use clientelism to compete with them. Similarly, the traditional elite ran a political system with few checks and balances. Would the MNR find it optimal to create a political system with checks and balances? Not necessarily. After all, though this might have appealed to citizens and garnered more support, it would also have been disadvantageous to them once they were in power. Hence, there is a well-defined trade-off. Indeed, the MNR were able to attain power and create highly imperfect political institutions which they were then able to undermine.

6.3.2 General lessons

One might conclude from our discussion of the US South that the real problem was the persistence of the elite and their resources. If only the North had implemented land reform and given the freed slaves their 40 acres and a mule, as they had been promised, everything would have been different. The example of the Bolivian Revolution shows that the situation is more complex than this. In Bolivia, the previous elite were expropriated and their power taken away, yet the new elite that emerged (the MNR) used strategies that were very similar to the old elite and which had the same impact on economic institutions. Thus, there can be huge path dependence in political equilibria, even when *de jure* and *de facto* power changes hands from one group to another. This

implies that, for reformers, a policy of changing political institutions and trying to simultaneously undermine the *de facto* power of incumbents may not work. Instead, reformers must change the incentives of new elites and decouple their choices from those of the previous elites.

6.4 Successful reform

Though, so far, we have emphasized the problems of reform which result for the political nature of poor governance, nevertheless there are many cases of improvements in governance. To give a tangible example of the forces leading to institutional change let us return to the US South. Starting in the 1940s, the income *per capita* of the US South began to converge very rapidly to the US average. This period saw the end of the isolation of the labor market. It saw the abolition of institutionalized racial discrimination in labor markets and social life and the re-enfranchisement of blacks culminating in the Voting Rights Act of 1965 (see Wright, 1999, for an overview).

These changes were driven by a number of interacting forces which both changed the ability of Southern elites to maintain the previous system and their incentives to do so. The ability of whites to continue with the institutions which had been in effect since the late nineteenth century was severely undermined by the fact that blacks in South finally solved the collective action problem (McAdam, 1983). The civil rights movement made much of the previous system unenforceable. Black collective action was facilitated by the intervention of the federal government. At the same time, there was a much smaller ability and incentive for the whites to continue to repress labor. Starting in the 1940s, there had been a huge out-migration of blacks from the South which could no longer be controlled. At the same time, technological change, particularly the mechanization of cotton picking (Heinicke, 1994) made it far less important to repress labor.

There was real change, therefore, in the political equilibrium in the South which led to much better governance and economic institutions. The old economy based on extracting rents from the blacks crumbled, along with all the negative externalities that it had for other parts of the economy. No new elite arose to carry on repressing black labor using different instruments and the situation did not turn on its head with blacks extracting rents from whites. This may have been because of the technological changes that took place at the same time, but more likely it was because the South is embedded in a larger economy with well functioning institutions. This latter feature of the South may have considerably raised the opportunity cost of having bad economic institutions and is something which obviously differentiates it from Bolivia or the Congo.

Improved governance can, therefore, come about when changes in state variables or structural parameters influence the costs and benefits of different sets of institutions, and thus change interests, or there is a change in the distribution of political power. These

changes are often endogenous to the system though often in response to shocks. Nevertheless, policy interventions promoted by economists and by external entities such as the World Bank can lead to sustained reform by working on any of these margins. An obvious example of this are the institutional and governance reforms precipitated in Eastern Europe by the carrot of accession to the European Union. This created enormous financial incentives to reform.

6.5 Governance and equity—The Scandinavian model

Though better governance may have arisen in Britain in the seventeenth century as the consequence of a political revolution (Pincus, 2009), in the US South governance improved as the joint outcome both of economic and political changes. Though people certainly refer to the “Civil Rights Revolution” in the United States, one does not need a revolution to change the political equilibrium and indeed, as the Iron Law of Oligarchy emphasizes, revolutions may simply replace one dysfunctional equilibria with another. It is possible that just changes in the economic environment, in technology or in the organization of markets can permanently change the political equilibrium and this is no doubt what Karl Marx had in mind when he observed that “the windmill gives you society with the feudal lord; the steam mill, society with the industrial capitalist” (see also White, 1962). We illustrate this possibility with an in-depth case study of the emergence and persistence of the Scandinavian model of capitalism.

6.5.1 The rise of the Scandinavian model

Social Democratic governments came to power in Sweden and Norway in the midst of the Great Depression committed to reducing unemployment and alleviating poverty. The main slogan of the social democrats in the 1930s was employment for everybody. Both governments increased spending on policies such as unemployment benefits, public housing, and agricultural price supports. In retrospect, the key innovation was not the crisis policies that were adopted in the 1930s, but the institutional response to the problem that threatened the recovery program (Moene & Wallerstein, 2006): What would keep the increased government spending from raising the wages of insiders in the labor market, rather than increasing employment?

The problem came to a head in both countries in the construction industry. Construction workers in Sweden and Norway were highly paid, militant and sheltered from foreign competition. When foreign demand collapsed in the 1930s, workers in the export sectors such as metalworkers accepted large wage reductions in order to stem the decline of employment. Construction workers came under no such pressure, in large part because of increased government spending on housing. Since construction workers were employed in the export sector as well as in home construction, higher construction wages raised labor costs in the export sector, which threatened the jobs of metalworkers. When construction unions called a strike in support of higher wages,

the national confederation of unions intervened to force the strike to an early and from the construction workers' point of view, unsuccessful conclusion.

The intervention of the national union confederation to end the strikes in construction was the initial step in a process of centralization of authority within the union movement in both Norway and Sweden, a process that was encouraged and supported by employers. "Basic agreements" between the national associations of unions and employers establishing rules for collective bargaining at the industry level were reached in 1935 in Norway and 1938 in Sweden. In the 1950s (1956 in Sweden, 1958 in Norway), bargaining at the industry level was replaced by direct negotiations over pay by the national associations of unions and employers. As white-collar and professional union confederations joined the centralized negotiations, the coverage of the central agreements expanded to include most of the working population in the private sector.

The central agreements were necessarily general. The details of how the agreement was to be implemented were decided by subsequent bargaining at the industry and local level. Once the central agreement was signed, however, work stoppages were illegal. Wage increases at the local level were limited to what could be obtained without the threat of a strike. The centralized system of wage setting, which reached its zenith in the 1970s, had three important consequences. The first was the virtual elimination of industrial conflict. From the countries with the highest levels of strikes and lockouts in the world in the interwar year, Norway and Sweden became countries with some of the lowest levels of industrial conflict in the postwar period.

The second consequence was to allow conditions in the export industries to determine the growth of wages throughout the economy. This implied wage moderation. In practice, the centralized system of wage bargaining tied wage growth throughout the economy to the growth of wages in the export sector, since the unions in the export sector, the metalworkers in particular, were the largest and most influential unions within the national confederations.

The third consequence of centralized wage setting was a gradual process of wage compression that, over time, generated the most egalitarian distribution of wages and salaries in the capitalist world. In the 1950s, wage compression was adopted as an explicit goal of the unions in both Norway and Sweden under the title of "solidaristic bargaining."

Solidaristic bargaining was defended more in terms of efficiency than in terms of equity. In the 1950s, two Swedish trade union economists, Gösta Rehn and Rudolf Meidner (Rehn, 1952), argued that equalizing wages across Swedish firms and industries would promote economic development by forcing wages up in low-productivity firms or industries and keeping wages down in high-productivity firms or industries. By reducing profits in low-productivity firms and increasing profits in high-productivity firms, labor and capital would be induced (or coerced) to move from low-productive to high-productive activities, increasing aggregate efficiency as well as improving equality (Agell & Lommerud, 1993; Moene & Wallerstein, 1997).

Whatever the benefits of solidaristic bargaining in terms of efficiency, the cumulative impact on the distribution of wages and salaries was large. In Sweden between 1970, when comprehensive wage data on individuals began to be collected, and 1983, when the system of centralized bargaining temporarily collapsed, the variance of the log of hourly wages among private sector blue-collar workers declined by over 50% (Hibbs & Locking, 2000). That dramatic decrease does not include the equally prominent reduction of the wage differential between blue-collar and white-collar workers. Hibbs and Locking (2000) estimate that a similar decline occurred during the 1960s as well, implying that the variance of log hourly wages in 1983 was only one-quarter of what it was in 1960. In 2003, the ratio of the wage for a worker at the 90th percentile of the wage distribution to the wage for a worker at the 10th percentile was about 2 to 1 in Sweden, Norway and Denmark, the lowest ratios of any country in the OECD. In contrast, the 90 to 10 ratio was above 5 to 1 in the United States in 2003.

To keep highly productive employers from undermining the policy of wage restraint by offering workers generous benefits (which were harder than wages to monitor at the central level), the Swedish employers' confederation lobbied the government to nationalize the provision of health care and pensions (Swenson, 2002). Moene and Wallerstein (2001) show that expenditures on social insurance against the loss of income due to unemployment, disability, sickness, and occupational injury rises as wage inequality declines. If insurance is a normal good, a policy that raises the income of the majority of workers with below average incomes increases the political demand for social insurance policies. The compression of wage differentials, in sum, had far-reaching economic and political consequences, one of which, was to increase the pace of economic development. It introduced a system of governance and incentives that led private businesses to act in socially desirable ways without altering property rights.

6.5.2 How equality multiplies

Today, the societal model of the Scandinavian countries is still distinguished by a large welfare state, encompassing unions and employers associations, and a governance system of routine consultation among government and representatives for interest organizations. In terms of policy, it is characterized by the provision of basic goods for everybody as a right of citizenship; a government committed to full employment; and wage leveling through "solidaristic bargaining" (Moene & Wallerstein, 1993).

From a governance point of view, it is important to notice that these institutional arrangements are complementary. The impact of one of them tends to strengthen the impact of others. Governance is easier and more efficient whenever institutions and policies fit together in this manner. A policy for full employment, for instance, requires wage moderation by the unions. For union members to accept that their union bosses negotiate wage moderation, they must face credible promises of a full

employment policy. Thus, the unions rely on a government commitment to full employment in their wage moderation policies as much as the government relies on union wage moderation in its policy for full employment.

The complementarity between social spending and wage setting is a related aspect of the mutual dependence between politics and markets. Barth and Moene (2008) demonstrate how economic and social equality multiplies due to the complementarity between wage determination and welfare spending. Returning once more to the mechanism that Tocqueville hinted at, we have that (A) a more equal wage distribution fuels welfare generosity via political competition, while (B) a more generous welfare state fuels wage equality further via its support to weak groups in the labor market.

Together the equality magnifying effect (A) and the wage equalizing effect (B) generate a cumulative process that adds up to a sizable social multiplier. Using data on 18 OECD countries over the period 1976–2002, Barth and Moene (2008) are able to identify an equality multiplier of more than 50%. Any exogenous change in either welfare spending or wage setting is thus magnified by 50% by endogenous forces caused by social complementarity. This equality multiplier helps explain why almost equally rich countries differ so much in the economic and social equality that they offer their citizens: With only one-third of the pretax inequality of the United States, the Scandinavian countries of Denmark, Norway, and Sweden have twice as generous welfare spending as the United States.

These calculations are based on a political economic equilibrium approach that incorporates the mutual dependence between the governance of labor markets and the governance of social policies. While social welfare spending depends on the wage dispersion in the labor market, it also feeds back to the determination of this wage dispersion. The political economic equilibrium outcome is wage dispersion and a level of welfare spending that are consistent taking the mutual feedbacks into account. Hence, the new political economic equilibrium gives “new maxims to those who govern and particular habits to the governed,” as Tocqueville said.

These are examples of how certain policies, institutions, and governance systems fit together and strengthen each other. In the long run, the outcomes may look as if societal arrangements come in packages with different social and economic organization.

6.5.3 How the political equilibrium changed

The Scandinavian model was not the result of intelligent design, but rather of social, economic, and political evolution. One of the central groups who supported solidaristic bargaining was the employers (Swenson, 1989, 1991). While the Nordic countries are well known for the strength of unions, employers also achieved an extraordinary level of organization. Employers much preferred to bargain with the “sensible” leadership of the union confederations, rather than with the militant leadership of the shop floor union bodies.

The other important group that supported the policy of wage compression was the leadership of unions of low-wage workers. Since the union movement was encompassing, both low- and high-wage earners had influence in union policy. While the policy of wage compression was controversial in unions of high-wage workers, it was enthusiastically supported by unions of low-wage workers. Thus, the political coalition that prevailed in the 1950s and established the pattern of centralized and solidaristic bargaining that was to last until the 1980s was comprised of the low-wage unions and employers.

High-wage unions were prevented from leaving the centralized negotiations by the threat of lockouts. It is unlikely that the low-wage unions and the leadership of the union confederation would have been able to force the high-wage unions to accept an egalitarian wage policy without the backing of employers and the threat of lockouts against recalcitrant unions.

Many critics would claim that Scandinavian model is only possible in consensual, homogeneous, and affluent societies with an extraordinary commitment to equality. The most common explanations for the Scandinavian experience thus circle around variations of Scandinavian exceptionalism, emphasizing the importance of social homogeneity (Alesina, Glaeser, & Sacerdote, 2002), a Nordic commitment to equality (Therborn, 1986), a consensual model of decision-making (Wilensky, 2002) and affluence. In third world countries that are conflict-ridden, heterogeneous and poor, the model is deemed infeasible. But conditions in Norway and Sweden in the period preceding the social democratic ascent to power were anything but consensual, egalitarian, and affluent.

In the 1920s and early 1930s, Norway and Sweden experienced the highest levels of industrial conflict in the world. In Norway, the number of working days lost in strikes and lockouts in one year—1931—was three times larger than the total number of working days lost in industrial conflict over the 25-year period, 1945–1970. Nowhere else were employers as ready as employers were in Norway and Sweden to fight the unions with lockouts.

The consensus between employers and unions that characterized social democracy after the war was nowhere to be seen when the Social Democrats entered government in the 1930s. While the Nordic countries were relatively homogeneous in terms of religion and language, the working population was far from homogenous in terms of living conditions. In particular, the social and economic cleavages between rural and urban residents were striking. Measured by income *per capita*, the gap between the poorest and richest rural municipalities was 1–14 (Falk & Tovmo, 2000).

Scandinavian social democrats came to power in societies no less economically divided than many poor countries of today. As in developing countries today, there was significant underutilization of labor. Surplus labor in the form of open unemployment was most evident. In addition, there was disguised unemployment in the

countryside that may have been as significant as open unemployment in the cities. Around half of the population lived in sparsely populated areas where most made a living from farming and fishing.

Finally, the economies that the social democrats inherited in the 1930s were far from affluent. The real *per capita* GDP of Sweden and Norway when the social democrats entered government was far below the current real *per capita* GDP of middle-income countries like Brazil or South Africa. The majority of citizens in Scandinavia became rich under the Scandinavian model of governance, not before.

Consensus, homogeneity, and affluence are products of the Scandinavian governance, not prerequisites. Nevertheless, skeptics have from the beginning doubted the long-run feasibility of a governance structure that combines market efficiency with social equality. In 1899, Rosa Luxemburg (1970, p. 43) characterized it as “a sort of a labor of Sisyphus” in which social victories would be continually eroded by market forces. More recently, conservative critics have made the reverse argument that market forces are steadily eroded by social reform with bad consequences for economic performance (Lundberg, 1985). Neither view is proven correct. Social equality and worker security have persisted in the Scandinavian countries and economic growth has been at par with the United States.

6.5.4 General lessons

The Scandinavian experience is interesting for several reasons. First, it is an interesting example of an endogenous change in the political equilibrium. Though the background to this is no doubt the spread of democracy around the time of the First World War, the proximate cause was the shock of the Great Depression. It was certainly not a general phenomenon that this shock led to better political equilibria in the world. In much of Latin America, for example, it led nascent democratic governments to be overthrown and to the emergence of an inward looking authoritarian development strategy that probably significantly retarded economic growth during the second half of the twentieth century. In Europe, it leads to fascist governments in Germany, Italy, and Spain where the left were unable to obtain the support of the landed peasants. While the class struggle in the countryside was contested in Germany, Italy, and Spain, it was settled in the Scandinavian countries of Denmark, Norway, and Sweden. For instance, the campaign for land reform the South of Spain antagonized peasants even in the North. In Scandinavia, landownership was more egalitarian and the rural class struggle was much less intense (This divergent experience is discussed in Luebbert, 1991, see also Roemer 2001, chapter 11).

In Scandinavia, the shock of the Great Depression precipitated a very different reaction from many other countries. While political compromises were sought in the center between social democrats and agrarian parties, the implicit coalitions that emerged in the labor markets were more “ends against the middle” coalitions where the main

beneficiaries were low skilled workers and capital owners. This latter institutional response changed dramatically the way that the labor market worked and fed back into the entire political equilibrium. The *de facto* power changed as a new government was formed backed by the interests that emerged in the reorganized labor market.

While the gap between Scandinavia and the rest of the World in 1930 was no doubt much smaller than that between Sierra Leone and the United States today, the evidence really suggests that these changes did alter the political equilibrium in a way that generated the societies we see today. Scandinavia was not always “different” and if preferences between people in those countries differ today, this is more likely to be an outcome of institutional differences rather than a cause.

Second, this example suggests again the complex way in which good institutions and good governance are created. The Scandinavian model generated rapid economic growth and rising living standards. But it did it with very different labor market institutions and welfare policies that at the same time promoted a radically egalitarian distribution of income. Growth occurred for the standard reasons, there were incentives to accumulate and save and the state provided public goods, but many of these incentives were generated in the context of quite distinct specific economic institutions (a theme of the literature on the “varieties of capitalism,” see Hall & Soskice, 2001; Rodrik, 2008).

6.6 Promoting reform

So far, we have discussed two examples, from the US South and Scandinavia, of endogenous reforms. We conclude this section by returning to Sierra Leone and asking what have we learned that might help us promote reform there or whether reform is likely to come endogenously. We first observe that the *See-Saw Effect* has been powerfully in operation in Sierra Leone (this is one of the sad lessons of the previous experience of structural adjustment in the country which Herbst (1990), Griffiths (2003), Reno (1995, 1998) and made so much of). The experience of Sierra Leone has been precisely that attempted reform in one dimension has led to disreform in another. How then might one push the political equilibrium in Sierra Leone in the direction of better outcomes? One idea comes from our earlier discussion about the construction of patrimonialism in the country. At some deep level, it is difficult to change the attractiveness of patrimonialism as a political strategy. It is obviously an incredibly effective way of binding your supporters to you and disarticulating your opponents. Things of course may have changed. First, Sierra Leone may be more of a nation state now and there may be more of a sense of national identity than there was in the 1960s or 1970s, the lack of which we have argued makes patrimonial rule attractive. Second, people’s preferences (We hesitate to use the word “values”) may have changed in a way which makes the “patrimonial exchange” more difficult to consummate. In essence, citizens may be less happy with patronage now, they want development.

The evidence on this seems to be very mixed, however. So the structural underpinnings of patrimonial rule may have changed, but they probably have not changed much. We do not think the World Bank or international institutions and governments can do much about this in any case. What they can do is to focus more on how institutions were changed to facilitate patrimonialism.

Leaving the social structure aside, it is clear from the facts about Stevens' rule and from the wider African evidence that certain sorts of political institutions facilitate patrimonialism (Robinson and Torvik, 2008). Stevens did not just suspend democracy, he also made himself president and concentrated power in the executive. He also accumulated power in Freetown. The international community clearly recognized this when they pushed for decentralization after 2002 and the reintroduction of district councils and more generally of course the support for multiparty democracy in the country. Trying to decentralize power and resources out of Freetown was surely an excellent idea. But this view which underpinned this policy was not taken to its logical conclusion. In itself, just establishing elected district councils has not done much to undermine executive dominance or strengthen the system of checks and balances which are so crucial to a successful democracy, it is only a small part of reforms in political institutions which might be potentially useful.

The obvious conclusion from the discussion, so far, is that other reforms of political institutions are needed which would further unwind the legacy of patrimonialism. The most important aspect of this is the power of the executive, particularly relative to parliament. This is an obvious conclusion from the experience of Sierra Leone and Africa more generally. We reiterate, however, that we are still far from having a convincing framework which will help us understand how to change the political economy equilibrium of a society and few changes, not even reducing the power of the president or even reversing presidentialism, have such a large impact as the potential ability to join the European Union.

7. CONCLUSIONS

Returning in conclusion to Killick's mango canning factory discussed in the introduction we can ask was this example of bad governance, and many others like it, the reason for such poor economic performance in Ghana? Our argument suggests that such projects were part of the channel via which the political equilibrium created poor performance. Such projects were driven by the patrimonial logic underlying Nkrumah's political strategy and good governance in the sense of choosing rational public sector investments was inconsistent with Nkrumah's goal of consolidating his political power (Robinson and Torvik, 2005). So bad governance in the sense of an ineffective state was economically costly but was nevertheless politically attractive or perhaps expedient. This was because these projects allowed Nkrumah to employ his supporters or generate

rents for contractors in parts of the country where he needed to consolidate his support. Bad governance in this sense was a consequence of bad governance in the other sense, of the nexus of political institutions which determine the political equilibrium. By the time the mango canning factory was built, Nkrumah had suspended democracy and made himself president so there was little accountability and little chance for the immizerized citizens of Ghana to contest this economically costly decision.

It is possible then to attribute important consequences for development to governance. Nevertheless, we have argued in this chapter that the literature on governance is really part of the more general literature on the political economy of institutions and development and so far the proponents of governance have not made a strong case that there is something called “governance” which is conceptually distinct and which is causally related to development. As yet governance has not been unbundled and indeed appears too vague to be unbundled. This being the case, though we have attempted here to enter into the spirit of the literature on governance, we are ultimately quite skeptical about what the term, as currently wielded by its proponents, adds to our understanding of development.

End Notes

- *. Written as a chapter in the Handbook of Development Economics, edited by Dani Rodrik and Mark Rosenzweig to be published by North Holland. We are grateful to Frederico Finan and Robert Putnam for providing us with their data to María Angélica Bautista for her research assistance and Dani Rodrik for his detailed comments on an earlier draft. Karl Moene’s work was conducted under the auspices of the Centre of Equality, Social Organization, and Performance at the University of Oslo.
- 1. See Roe (2003) and Gourevitch and Shinn (2005) on the importance of state politics in explaining corporate governance.
- 2. Though it is somewhat controversial, we include economic policies along with economic institutions. Policies, like institutions, are chosen by those with political power and they have large impacts on the incentive structure of society so for the purposes of the discussion it is not important to distinguish between them.
- 3. <http://info.worldbank.org/governance/beeps2002/>
- 4. <http://www.latinobarometro.org>
- 5. Another source of data on corruption and governance comes from Business International now incorporated into the Economist Intelligence Unit which was first used by Mauro (1995) and also Treisman (2000).
- 6. This discussion presumes that the distribution of political power matters for the efficiency of institutions and policies (see Acemoglu, 2003, 2006).
- 7. On the nature and behavior of the elite in the Sudan, see Seekers of Truth and Justice (2000) and Cobham (2005).
- 8. Interestingly, though it was possible to purchase commissions in the British army until the 1870s this was never possible in the (much more important) navy.
- 9. There is now a small but important theoretical literature on the political economy of state capacity, see Acemoglu (2005b), Acemoglu, Ticchi, and Vindigni (2006), and Besley and Persson (2007).
- 10. Important studies in this enormous literature include Liu (1985), Andvig and Moene (1990), Shleifer and Vishny (1992), Besley and McLaren (1993), Mookherjee and Png (1995), Banerjee (1997),

- Acemoglu and Verdier (1998, 2000), and Chand and Moene (1999); see Bardhan (1997), Rose-Ackerman (1999), Aidt (2003), and Svensson (2005) for useful overviews of the literature on corruption.
11. There is total consensus in the academic literature on this see Davies (2007) for a recent authoritative view, also Luke and Riley (1989) and Kallon (2004).
 12. See Cartwright (1970) and Clapham (1976) for the politics of the early independence period.
 13. For example, the 100-year rule of the True Whig Party in Liberia was overthrown in 1980 by just 18 men led by Master Sergeant Samuel Doe.
 14. Stevens was not of course the first to interfere with chiefship. The British colonial state had done this as did early SLPP governments (see Kilson, 1966).
 15. In Krio “who sai dem tie cow nar day e go eat” quoted in Stevens (1984, p. 35).
 16. The fact that the within variation does not suggest that the change in income *per capita* and control of corruption are correlated may suggest, along the lines of Acemoglu et al. (2008), that the correlation between the levels is due to the effects of an omitted variable.
 17. See also La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1999).
 18. See Haber and Menaldo (2007) who similarly find no effect of oil on the political regime.
 19. As Englebert (2000) has shown the arbitrariness of a country’s boundaries in Africa are correlated with poor development outcomes.
 20. Many other types of arguments appear in the literature, for example, patrimonialism is linked to central traits of African indigenous political systems or maybe African political culture (see Carlson, 1999). Another set of arguments involve the claim that patrimonialism is more attractive when countries are “resource dependent.” Unfortunately, resource dependence is hard to measure in an objective way since economic decline will naturally lead a country to become resource dependent, at least if measured by the share of natural resources in GDP or exports (see Robinson, Torvik & Verdier, 2006, for this point).
 21. Herbst (2000) and Bates (2001) tend to emphasize “fundamentals” such as population density and resource endowments as the main reason for the lack of incentives to build states, rather than the institutional inheritance of colonialism.
 22. See Migdal (1988) for an analysis of indirect rule and its consequences in Sierra Leone.
 23. Though of course there is Somalia and Rwanda as well.
 24. Indeed though Humphreys and Weinstein’s (2008) data show that being Mende does predict membership in the RUF, this effect comes only from those who were coerced into joining the movement. This effect almost certainly stems from the fact that the RUF began in the south of the country, which is more explained by the location of Liberia and Charles Taylor than anything else.
 25. Mentioning Botswana, one might ask why the factors we have discussed here did not lead to patrimonialism in Botswana, and one salient success story in Sub-Saharan Africa. As Parsons and Robinson (2006) discuss, the historical pattern of institutional creation and governance in Botswana was rather unique both before, during and after the colonial experience. They show how the economic success of Botswana can be explained by the absence of the forces here we suggest lead to patrimonialism.

References

- Abharam, A., & Sesay, H. (1993). Regional politics and social service provision since independence. In C. M. Fyle (Ed.), *The state and the provision of social services in Sierra Leone since independence, 1961–1991*. Oxford: Codesaria.
- Acemoglu, D. (2003). Why not a political coase theorem? *Journal of Comparative Economics*, 31, 620–652.
- Acemoglu, D. (2005a). Constitutions, politics and economics: A review essay on Persson and Tabellini’s “The Economic Effects of Constitutions.” *Journal of Economic Literature*, 43, 1025–1048.
- Acemoglu, D. (2005b). Politics and economics in weak and strong states. *Journal of Monetary Economics*, 52, 1199–1226.

- Acemoglu, D. (2006). A simple model of inefficient institutions. *Scandinavian Journal of Economics*, 108, 515–546.
- Acemoglu, D., Cantoni, D., Johnson, S., & Robinson, J. (2009). Consequences of Radical Reform: The French Revolution, NBER Working Paper #14831.
- Acemoglu, D., & Johnson, S. (2005). Unbundling institutions. *Journal of Political Economy*, 113, 949–995.
- Acemoglu, D., Johnson, S., Querubín, P., & Robinson, J. A. (2008, Spring). When does policy reform work? The case of Central Bank Independence. *Brookings Papers of Economic Activity*, 351–417.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001, December). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, 118, 1231–1294.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2005a). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American Economic Review*, 95, 546–579.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2005b). Institutions as fundamental determinants of long-run growth. In P. Aghion, & S. Durlauf (Eds.), *The handbook of economic growth* (Vol. 1A, pp. 385–472). Amsterdam: North-Holland.
- Acemoglu, D., Johnson, S., Robinson, J. A., & Thaicharoen, Y. (2003). Institutional causes, macroeconomic symptoms: Volatility, crises and growth. *Journal of Monetary Economics*, 50, 49–123.
- Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2008). Income and democracy. *American Economic Review*, 98, 808–842.
- Acemoglu, D., & Robinson, J. A. (2000). Why did the west extend the franchise? Growth, inequality and democracy in historical perspective. *Quarterly Journal of Economics*, CXV, 1167–1199.
- Acemoglu, D., & Robinson, J. A. (2006). *Economic origins of dictatorship and democracy*. New York: Cambridge University Press.
- Acemoglu, D., & Robinson, J. A. (2007). A model of the Iron Law of Oligarchy. (unpublished data).
- Acemoglu, D., & Robinson, J. A. (2008a). Persistence of power, elites and institutions. *American Economic Review*, 98, 267–293.
- Acemoglu, D., & Robinson, J. A. (2008b). The persistence and change of institutions in the Americas. *Southern Journal of Economics*, 75, 282–299.
- Acemoglu, D., Ticchi, D., & Vindigni, A. (2006). *Emergence and persistence of inefficient states*. NBER Working Paper No. 12748.
- Acemoglu, D., & Verdier, T. (1998). Property rights, corruption, and the allocation of talent: A general equilibrium approach. *Economic Journal*, 108, 1381–1403.
- Acemoglu, D., & Verdier, T. (2000). The choice between corruption and market failures. *American Economic Review*, 90, 194–211.
- Agell, J., & Lommerud, K. E. (1993). Egalitarianism and growth. *Scandinavian Journal of Economics*, 95, 559–579.
- Aidt, T. S. (2003). Economic analysis of corruption: A survey. *Economic Journal*, 113, F632–F652.
- Alesina, A., Glaeser, E., & Sacerdote, B. (2002). Why doesn't the US have a European-style welfare state? *Brookings Papers on Economic Activity*, 187–277.
- Andvig, J. Chr., & Moene, K. O. (1990). How corruption may corrupt. *Journal of Economic Behavior and Organization*, 13(1990), 63–76.
- Banerjee, A. V. (1997). A theory of missgovernance. *Quarterly Journal of Economics*, 112, 1289–1332.
- Bardhan, P. (1997). Corruption and development: A review of issues. *Journal of Economic Literature*, 35, 1320–1346.
- Barrows, W. (1976). *Grassroots politics in an African state*. New York: Africana Publishing Company.
- Barth, E., & Moene, K. O. (2008). *The equality multiplier*. Mimeo: Department of Economics, University of Oslo.
- Bates, R. H. (1981). *Markets and states in tropical Africa*. Berkeley: University of California Press.
- Bates, R. H. (2001). *Prosperity and violence: The political economy of development*. New York: W.W. Norton & Co.
- Bayart, J. F. (1993). *The state in Africa: The politics of the belly*. New York: Longman.

- Bertrand, M., Djankov, S., Hanna, R., & Mullainathan, S. (2007). Obtaining a driver's license in India: An experimental approach to studying corruption. *Quarterly Journal of Economics*, 122, 1639–1676.
- Besley, T., & McLaren, J. E. (1993). Taxes and bribery: The role of wage incentives. *Economic Journal*, 103, 119–141.
- Besley, T., & Persson, T. (2007). *The origins of state capacity: Property rights, taxation, and politics*. NBER Working Paper No. 13028.
- Besley, T., Persson, T., & Sturm, D. (2005). *Political competition and economic performance: Theory and evidence from the United States*. NBER Working Paper No. 11484.
- Bloom, N., & van Reenan, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*, 122, 1351–1408.
- Brenner, R. (1976). Agrarian class structure and economic development in preindustrial Europe. *Past and Present*, 70, 30–75.
- Brewer, J. (1988). *The sinews of power: War, money and the English State, 1688–1783* (pp. 106–107). Cambridge: Harvard University Press.
- Carlson, R. (1999). *Presidentialism in Africa: Explaining institutional choice*. Unpublished PhD Thesis, Department of Political Science, University of Chicago.
- Cartwright, J. R. (1970). *Politics in Sierra Leone 1947–67*. Toronto: University of Toronto Press.
- Chand, S., & Moene, K. O. (1999). Controlling fiscal corruption. *World Development*, 27, 1129–1140.
- Clapham, C. (1976). *Liberia and Sierra Leone: An essay in comparative politics*. New York: Cambridge University Press.
- Clapham, C. (2003). *Sierra Leone: The political economy of internal conflict*. CRU Working Paper No. 20. The Hague: Clingendael Institute.
- Cobb, J. C. (1984). *Industrialization and southern society, 1877–1984*. Lexington: University Press of Kentucky.
- Cobham, A. (2005). Cause of conflict in Sudan: Testing the black book. *European Journal of Development Research*, 17, 462–480.
- Cooper, F. (2002). *Africa since 1940: The past of the present*. New York: Cambridge University Press.
- Davies, V. A. B. (2007). Sierra Leone's economic growth performance, 1861–2000. In B. J. Ndulu, S. O'Connell, R. H. Bates, P. Collier, C. C. Soludo, J.-P. Azam, A. K. Fosu & J. W. Gunning (Eds.), *The political economy of growth in Africa, 1960–2000* (Vols. 2). New York: Cambridge University Press.
- de Tocqueville, A. (2000). *Democracy in America*. Chicago: University of Chicago Press (New Edition 2000).
- Dixit, A. K. (2004). *Lawlessness and economics*. Princeton: Princeton University Press.
- Du Bois, W. E. B. (1903). *The souls of black folk*. New York: A.C. McClurg & Company.
- Engerman, S. L., & Sokoloff, K. L. (1997). Factor endowments, institutions, and differential growth paths among new world economies. In H. Stephen (Ed.), *How Latin America fell behind*. Stanford: Stanford University Press.
- Englebert, P. (2000). *State legitimacy and development in Africa*. Boulder, CO: Lynne Rienner Publishers.
- Esteban, J., & Ray, D. (1994). On the measurement of polarization. *Econometrica*, 62, 19–851.
- Evans, P. B. (1995). *Embedded autonomy*. Princeton: Princeton University Press.
- Evans, P. B., & Rauch, J. E. (1999). Bureaucracy and growth: A cross-national analysis of the effects of 'Weberian' state structures on economic growth. *American Sociological Review*, 64, 748–765.
- Evans, P. B., & Rauch, J. E. (2000). Bureaucratic structure and bureaucratic performance in less developed countries. *Journal of Public Economics*, 75, 49–71.
- Fafchamps, M. (2004). *Market institutions and Sub-Saharan Africa: Theory and evidence*. Cambridge: MIT Press.
- Falk, T., & Tovmo, P. (2000). *Norwegian local public finance in the 1930s*. Trondheim, Norway: Norwegian University of Science and Technology (unpublished paper).
- Fanthorpe, R. (2005). On the limits of the liberal peace: Chiefs and democratic decentralization in post-war Sierra Leone. *African Affairs*, 105, 1–23.
- Ferraz, C., & Finan, F. (2008). Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123, 703–745.
- Gould, D. J. (1980). Patrons and clients: The role of the military in Zaire politics. In I. J. Mowoe (Ed.), *The performance of soldiers as governors: African politics and the African military*. Washington: University Press of America.

- Gourevitch, P. A., & Shinn, J. J. (2005). *Political power and corporate control: The new global politics of corporate governance*. Princeton: Princeton University Press.
- Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. New York: Cambridge University Press.
- Griffiths, P. (2003). *The economist's tale: A consultant encounters hunger and the World Bank*. London: Zed Books.
- Guiso, L., Paola, S., & Luigi, Z. (2008). *Long-term persistence*. NBER Working Paper No. 14278.
- Haber, S. H., & Menaldo, V. (2007). *Does oil fuel authoritarianism*. Department of Political Science, Stanford University (unpublished).
- Hall, P., & Soskice, D. (Eds.), (2001). *Varieties of capitalism: The institutional foundations of comparative advantage*. New York: Oxford University Press.
- Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics*, 114, 83–116.
- Hausman, R., Rodrik, D., & Velasco, A. (2008). Growth diagnostics. In N. Serra, & J. E. Stiglitz (Eds.), *The Washington Consensus reconsidered: Towards a new global governance*. New York: Oxford University Press.
- Heath, D. (1972). New patrons for old: Changing patron-client relationships in the Bolivian Yungas. In A. Strickton, & S. M. Greenfield (Eds.), *Structure and process in Latin America; patronage, clientage, and power systems*. Albuquerque: University of New Mexico Press.
- Heinicke, C. (1994). African-American migration and mechanized cotton harvesting, 1950–1960. *Explorations in Economic History*, 31, 501–520.
- Henisz, W. J. (2000). The institutional environment for economic growth. *Economics and Politics*, 12, 1–31.
- Herbst, J. I. (1990). The structural adjustment of politics in Africa. *World Development*, 18, 949–958.
- Herbst, J. I. (2000). *States and social power in Africa: Comparative lessons in authority and control*. Princeton: Princeton University Press.
- Hibbs, D. A., Jr., & Locking, H. (2000). Wage dispersion and productive efficiency: Evidence for Sweden. *Journal of Labor Economics*, 18, 755–782.
- Hintze, O. (1975). *The Historical Essays of Otto Hintze*. New York: Oxford University Press.
- Horowitz, D. L. (1985). *Ethnic groups in conflict*. Berkeley: University of California Press.
- Humphreys, M., & Weinstein, J. M. (2008). Who fights? The determinants of participation in Civil War. *American Journal of Political Science*, 52, 436–455.
- Jackson, M. (2004). *In Sierra Leone*. Durham: Duke University Press.
- Jacome, L. I., & Vazquez, F. (2005). Any link between legal central bank independence and inflation? Evidence from Latin America and the Caribbean, IMF Working Paper WP/05/75. Washington: International Monetary Fund.
- Kallon, K. M. (2004). *The political economy of corruption in Sierra Leone*. Lewiston, NY: E. Mellen.
- Kang, D. C. (2002). *Crony capitalism: Corruption and development in South Korea and the Philippines*. New York: Cambridge University Press.
- Kaufman, D., & Kraay, A. (2002). Growth without governance. *Economía*, 3, 169–229.
- Kaufman, D., Kraay, A., & Zoido-Lobatón, P. (1999). *Governance matters*. World Bank Policy Research Working Paper No. 2196.
- Keen, D. (2005). *Conflict & collusion in Sierra Leone*. New York: Palgrave.
- Kelley, J., & Klein, H. S. (1981). *Revolution and the rebirth of inequality*. Berkeley: University of California Press.
- Killick, T. (1978). *Development economics in action; a study of economic policies in Ghana*. London: Heinemann.
- Kilson, M. (1966). *Political change in a West African state; a study of the modernization process in Sierra Leone*. Cambridge: Harvard University Press.
- Klein, H. S. (1992). *Bolivia: The evolution of a multi-ethnic society*. New York: Oxford University Press.
- Kpundeh, S. J. (1995). *Politics and corruption in Africa: a case study of Sierra Leone*. Lanham: University Press of America.
- Kreps, D. M. (1986). Corporate culture and economic theory. In J. Alt, & K. Shepsle (Eds.), *Rational perspectives on political science*. New York: Cambridge University Press.

- Lamoreaux, N. R., & Rosenthal, J. L. (2005). Legal regime and business's organizational choice: A comparison of France and the United States during the mid-nineteenth century. *American Law and Economic Review*, 7, 28–61.
- La Porta, R., Florencio Lopez-de-Silanes, R., Shleifer, A., & Vishny, R. (1998). Law and Finance. *Journal of Political Economy*, 106, 1113–1155.
- La Porta, R., Florencio Lopez-de-Silanes, R., Shleifer, A., & Vishny, R. W. (1999). The quality of government. *Journal of Law Economics and Organization*, 15, 222–279.
- Lindert, P. H. (2004). *Growing public: Social spending and economic growth since the eighteenth century* (2 Vols.). New York: Cambridge University Press.
- Lipton, M. (1977). *Why Poor People Stay Poor*. Cambridge: Harvard University Press.
- Liu, F. T. (1985). An equilibrium queuing model of bribery. *Journal of Political Economy*, 93, 760–781.
- Luebbert, G. (1991). *Liberalism, fascism, or social democracy*. New York: Oxford University Press.
- Luke, D. F., & Riley, S. P. (1989). The politics of economic decline in Sierra Leone. *Journal of Modern African Studies*, 27, 133–141.
- Lundberg, E. (1985). The rise and fall of the Swedish model. *Journal of Economic Literature*, 23, 1–36.
- Luxemburg, R. (1970). *Reform or revolution*. New York: Pathfinder Press.
- Mann, M. (1986). *The sources of social power* (Vol. I). New York: Cambridge University Press.
- Margo, R. A. (1990). *Race and schooling in the South, 1880-1950: An economic history*. Chicago: University of Chicago Press.
- Mauro, P. (1995). Corruption and growth. *Quarterly Journal of Economics*, CX, 681–712.
- McAdam, D. (1983). Tactical innovation and the pace of insurgency. *American Sociological Review*, 48, 735–754.
- Mehlum, H., Moene, K. O., & Torvik, R. (2006). Institutions and the resource curse. *Economic Journal*, 116, 1–20.
- Michels, R. (1962). *Political parties*. New York: The Free Press.
- Migdal, J. S. (1988). *Strong societies and weak states*. New York: Cambridge University Press.
- Moene, K. O., & Wallerstein, M. (1993). What's wrong with social democracy. In P. Bardhan, & J. E. Roemer (Eds.), *Market socialism*. New York: Oxford University Press.
- Moene, K. O., & Wallerstein, M. (1997). Pay inequality. *Journal of Labor Economics*, 15, 403–430.
- Moene, K. O., & Wallerstein, M. (2001). Inequality, social insurance and redistribution. *American Political Science Review*, 95, 859–874.
- Moene, K. O., & Wallerstein, M. (2006). Social democracy as a development strategy. In P. Bardhan, S. Bowles, & M. Wallerstein (Eds.), *Globalization and egalitarian redistribution*. Princeton: Princeton University Press.
- Mookherjee, D., & Png, I. (1995). Corruptable law enforcers: How should they be compensated? *Economic Journal*, 105, 145–159.
- Mosca, G. (1939). *The ruling class*. New York: McGraw-Hill.
- North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge, UK: Cambridge University Press.
- North, D. C., & Weingast, B. R. (1989). Constitutions and commitment: Evolution of institutions governing public choice in 17th century England. *Journal of Economic History*, 49, 803–832.
- Olken, B. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115, 200–249.
- Padró-i-Miquel, G. (2007). The control of politicians in divided societies: The politics of fear. *Review of Economic Studies*, 74, 1259–1274.
- Pareto, V. (1968). *The rise and fall of the elites*. New York: Arno Press.
- Parsons, Q. N., & Robinson, J. A. (2006). State formation and governance in Botswana. *Journal of African Economies*, 15, 100–140.
- Persson, T., & Tabellini, G. (2003). *The economic effects of constitutions*. Cambridge: MIT Press.
- Persson, T., Tabellini, G., & Trebbi, F. (2003). Electoral rules and corruption. *Journal of the European Economic Association*, 1, 958–989.
- Pincus, S. C. A. (2009). *The first modern revolution*. New Haven: Yale University Press.

- Pirenne, H. (1937). *Economic and social history of medieval Europe*. New York: Harcourt, Brace and Company.
- Pirenne, H. (1939). *Mohammed and Charlemagne*. New York: Barnes & Noble.
- Platteau, J. P. (2000). *Institutions, social norms, and economic development*. Amsterdam: Harwood Academic Publishers & Routledge.
- Polillo, S., & Guillen, M. (2005). Globalization pressures and the state: The worldwide spread of central bank independence. *American Journal of Sociology*, 110, 1764–1803.
- Pomeranz, K. (2000). *The great divergence: China, Europe and the making of the modern world*. Princeton: Princeton University Press.
- Postan, M. M. (1944). The rise of the money economy. *Economic History Review*, 14, 123–134.
- Putnam, R. (with Roberto Leonardi & Raffaella Y. Nanetti). (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.
- Ransom, R. L., & Sutch, R. (2001). *One kind of freedom: The economic consequences of emancipation* (2nd ed.). New York: Cambridge University Press.
- Rehn, G. (1952). The problem of stability: An analysis of some policy proposals. In R. Turvey (Ed.), *Wages policy under full employment*. London: W. Hodge.
- Reno, W. (1995). *Corruption and state politics in Sierra Leone*. New York: Cambridge University Press.
- Reno, W. (1998). *Warlords politics and African states*. Boulder: Lynne Rienner.
- Reno, W. (2003). Political networks in a failing state: The roots and future of violent conflict in Sierra Leone. *IPG*, 2, 44–66.
- Richards, P. (1996). *Fighting for the rainforest: War, youth and resources in Sierra Leone*. Oxford: James Currey.
- Robinson, J. A., & Torvik, R. (2005). White elephants. *Journal of Public Economics*, 89, 197–210.
- Robinson, J. A., & Torvik, R. (2008). *Endogenous presidentialism*. NBER Working Paper No. 14603.
- Robinson, J. A., Torvik, R., & Verdier, T. (2006). The political foundations of the resource curse. *Journal of Development Economics*, 79, 447–468.
- Rodrik, D. (2008). *One economics, many recipes*. Princeton: Princeton University Press.
- Roe, M. J. (2003). *Political determinants of corporate governance*. New York: Oxford University Press.
- Roemer, J. (2001). *Political competition. Theory and applications*. Cambridge: Harvard University Press.
- Rose-Ackerman, S. (1999). *Corruption and government: Causes, consequences and reform*. New York: Cambridge University Press.
- Sachs, J. D., & Warner, A. M. (1995). *Natural resource abundance and economic growth*. NBER Working Paper No. 5398.
- Sachs, J. D., McArthur, J. W., Guido, S.-T., Kruk, M., Bahadur, C., Faye, M., & McCord, G. (2004). Ending Africa's poverty trap. *Brookings Papers on Economic Activity*, 2004(1), 117–240.
- Sawyer, A. (2004). Violent conflicts and governance challenges in West Africa: The case of the Mano River basin area. *Journal of Modern African Studies*, 42, 437–463.
- Seekers of Truth and Justice. (2000). *The black book: Imbalance of wealth and power in Sudan*. (unpublished).
- Shleifer, A., & Vishny, R. W. (1992). Corruption. *Quarterly Journal of Economics*, 108, 599–618.
- Stevens, S. P. (1984). *What life has taught me*. Bourne End, Bucks., England: Kensal Press.
- Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives*, 19, 19–42.
- Swenson, P. (1989). *Fair shares: Unions, pay and politics in Sweden and West Germany*. Ithaca: Cornell University Press.
- Swenson, P. (1991). Bringing capital back in, or social democracy. *World Politics*, 43, 513–544.
- Swenson, P. (2002). *Capitalists against Markets*. New York: Oxford University Press.
- Tangri, R. (1978). Local government institutions in Sierra Leone. *Journal of Administration Overseas*, 17, 17–27, 118–128.
- Therborn, G. (1986). *Why some peoples are more unemployed than others*. London: Verso.
- Tilly, C. (Ed.), (1975). *The formation of national States in Western Europe*. Princeton: Princeton University Press.
- Tilly, C. (1990). *Coercion, capital and European States AD 900-1990*. Cambridge: Basil Blackwell.
- Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics*, 76, 399–457.

- Turits, R. (2003). *Foundations of despotism: Peasants, the Trujillo regime, and modernity in dominican history*. Stanford: Stanford University Press.
- Turner, T., & Young, C. (1985). *The Rise and Decline of the Zairian State*. Madison: University of Wisconsin Press.
- van de Walle, N. (2001). *African economies and the politics of permanent crisis, 1979-1999*. New York: Cambridge University Press.
- White, L. T. (1962). *Medieval technology and social change*. London: Oxford University Press.
- Wiener, J. M. (1978). *Social origins of the new South: Alabama, 1860-1885*. Baton Rouge: Louisiana State University Press.
- Wilensky, H. (2002). *Rich democracies*. Berkeley: University of California Press.
- Williams, E. E. (1944). *Capitalism and slavery*. Chapel Hill: University of North Carolina Press.
- Wolfowitz, P. (2006, April). Good governance and development—A time for action. Remarks at Jakarta, Indonesia.
- Woodward, C. V. (1955). *The strange career of Jim Crow*. New York: Oxford University Press.
- World Bank. (2006). A decade of measuring the quality of governance. World Bank's web page Governance and Anti-corruption. www.worldbank.org/wbi/governance.
- Wright, G. (1986). *Old south, new south*. New York: Basic Books.
- Wright, G. (1999). The civil rights revolution as economic history. *Journal of Economic History*, 59, 267–289.
- Young, C. (1994). *The African colonial state in comparative perspective*. New Haven: Yale University Press.
- Zolberg, A. R. (1966). *Creating political order; the party-states of West Africa*. Chicago: Rand McNally.

Labor Regulations, Unions, and Social Protection in Developing Countries: Market Distortions or Efficient Institutions?

Richard B. Freeman

Harvard and NBER, Centre for Economic Performance, LSE

Contents

1. The Debate over Labor Institutions	4661
1.1 Theoretical perspectives	4664
2. Microevidence on Minimum Wages	4667
3. Wage Curve	4670
4. Employment Regulations	4672
5. Mandated Benefits	4674
6. Union Effects	4674
7. Cross-country Evidence	4680
8. Country Cases	4682
8.1 China	4683
8.2 Argentina	4684
9. The Informal Sector	4684
10. Conclusions	4687
Appendix A	4688
Appendix B: Regression	4693
End Notes	4693
References	4695

Abstract

This chapter reviews what economists have learned about the impact of labor market institutions, defined broadly as government regulations and union activity on labor outcomes in developing countries. It finds that:

- (1) Labor institutions vary greatly among developing countries but less than they vary among advanced countries. Unions and collective bargaining are less important in developing than in advanced countries while government regulations are nominally as important.
- (2) Many developing countries' compliance with minimum wage regulations produces spikes in wage distributions around the minimum in covered sectors. Most studies find modest adverse

effects of the minimum on employment so that the minimum raises the total income of low-paid labor.

- (3) In many countries minimum wages “spillover” to the unregulated sector, producing spikes in the wage distributions there as well.
- (4) Employment protection regulations and related laws shift output and employment to informal sectors and reduce gross labor mobility.
- (5) Mandated benefits increase labor costs and reduce employment modestly while the costs of others is shifted largely to labor, with some variation among countries.
- (6) Contrary to the Harris-Todaro two-sector model in which rural-urban migration adjusts to produce a positive relation between unemployment and wages across regions and sectors, wages and unemployment are inversely related by the “wage curve.”
- (7) Unions affect nonwage outcomes as well as wage outcomes.
- (8) Cross-country regressions yield inconclusive results on the impact of labor regulations on growth while studies of country adjustments to economic shocks, such as balance of payments problems, find no difference in the responses of countries by the strength of labor institutions.
- (9) Labor institution can be critical when countries experience great change, as in China's growth spurt and Argentina's preservation of social stability and democracy after its 2001–2002 economic collapse. Cooperative labor relations tend to produce better economic outcomes.
- (10) The informal sector increased its share of the work force in the developing world in the past two decades. The persistence of large informal sectors throughout the developing world, including countries with high rates of growth, puts a premium on increasing our knowledge of how informal sector labor markets work and finding institutions and policies to deliver social benefits to workers in that sector.

JEL classifications: O150, O170, J010, J080, J300, J500, J510, J580, J800

Keywords

informal labor market
labor regulations
minimum wage
productivity
distribution of income

Once about a time—not so long ago—the international financial institutions and many in the economics and policy establishment believed that they knew how to create sustainable growth in developing economies. They had a tool kit of policy prescriptions that they could take from country to country to cure economic ills. For the labor market, the package called for reduced regulations and lower social protection, cuts in public sector pay and employment, weaker unions, and greater reliance on market wage setting compared to collective bargaining or administrative rules. The enemy of growth was “urban bias” (Lipton, 1977)—government or union setting of pay and work

conditions that benefit modern sector workers but that reduce the flow of workers from low productivity informal and rural sectors to the modern sector. The World Bank's 1990 Development Report presented the prevailing wisdom: "Labor market policies—minimum wages, job security regulations, and social security—are usually intended to raise welfare or reduce exploitation. But they actually work to raise the cost of labor in the formal sector and reduce labor demand . . . increase the supply of labor to the rural and urban informal sectors, and thus depress labor incomes where most of the poor are found." (World Bank, 1990, p. 63).

Underlying this perspective was the Harris-Todaro (1970) two-sector model that attributed joblessness in developing countries to institutionally imposed high urban wages. The model posited that the high wages induced rural workers to migrate to urban areas, where they became unemployed while waiting for good jobs. Migration continued until the rate of unemployment equated the expected urban sector earnings (the wage times the probability of employment) to rural earnings. In this situation, an increase in modern sector employment at the institutionally determined wage does not raise GDP. This is because the addition of a high productivity job induces enough rural workers to migrate into urban unemployment to reduce rural output by an amount equal to the increased output due to the new urban job.¹

World Bank and International Monetary Fund economists also worried that labor institutions would undermine structural adjustment programs designed to cure balance of payments deficits or other economic ills. Viewing the archetypical problem as one in which the developing country ran into a balance of payments deficit, they stressed the need to shift resources from labor-intensive nontraded goods and services to capital-intensive traded goods sectors. The least costly way to do this was to devalue the currency, which would raise the price of tradeable goods and services relative to nontradeable goods and services and thus attract resources into the traded sectors. As long as tradeable goods were capital intensive, this would also shift the income distribution toward capital. The fear was that unions or other institutions that raised wages to preserve labor incomes would stop relative prices from moving in the desired direction. In the absence of a price-induced shift in resources, the country would have to undergo a recession to reduce imports and raise exports, which would be far more costly than a real devaluation.

At the 1992 World Bank Annual Conference on Development Economics, I reviewed extant evidence that labor institutions harmed economic development and stymied adjustments to macroeconomic problems per this analysis and found it sparse and unconvincing (Freeman, 1993a). The strongest evidence was Fallon and Lucas's (1989) comparison of the response of employment to output and wages in 35 industries in India and 29 industries in Zimbabwe before and after these countries strengthened labor laws. Their analysis showed that industries adjusted employment to changes in output as rapidly after the laws as before the laws but that employment was lower at the same output after

the laws (i.e., that productivity improved, which could be interpreted as a positive outcome). In the absence of evidence that firms complied with the laws and that other factors did not affect outcomes over the same period (they noted that Zimbabwe became independent coterminous with the change in labor regulations) I viewed the results as inconclusive at best. I was more impressed by the large declines in real minimum wages and average earnings in many African and Latin American countries during the 1980s that suggested that labor regulations were more “sawdust” than “hardwood.”

The quantity and quality of research on labor institutions in developing countries has increased greatly since the early 1990s. Some countries changed labor regulations in ways that provide good pseudoexperiments of whether institutions help or hinder the working of labor markets. Many countries now regularly provide researchers with microdata files on individuals and establishments that permit deeper probing of hypotheses than is possible with aggregate data. Research institutions and individual researchers have developed new datasets with country labor codes and institutional practices that illuminate cross-country differences and provide input into cross-country growth and other regressions.

In light of all this, what have we learned about how labor institutions affect outcomes in developing countries?

The recent research has not uncovered a general law for the effects of institutions on outcomes—economic circumstances and institutions probably vary too much among countries to support any single generalization—but it has yielded new and in some cases surprising findings on how institutions affect outcomes. This has led to a more measured view of what institutions do than in the World Bank’s 1990 proclamation. Here are the main findings:

- (1) Labor institutions vary greatly among developing countries but less than they vary among advanced countries. Collective bargaining is weaker in developing countries than in advanced countries while labor regulations are nominally similar.
- (2) Contrary to my initial skepticism, compliance with regulations in the formal sector of many developing countries is sufficient that minimum wages appear to be binding. They produce spikes in the distribution of wages around minimum. Most studies find that minimum wages reduce employment sufficiently modestly so that minimums generally help the low paid.
- (3) Contrary to Harris-Todaro type models, minimum wages induce spikes in the distribution of earnings in the informal sector in several countries, suggesting that minimum wages determine reservation wages of workers in those sectors.
- (4) Wages and unemployment are negatively related across geographic areas, consistent with the wage curve and contrary to the Harris-Todaro model.
- (5) Mandated benefits increase labor costs and reduce employment modestly while the costs of others are shifted largely to labor, with some variation among countries.

- (6) Some mandated benefits increase labor costs and reduce employment modestly, while the costs of others are shifted largely to workers and thus presumably do not impact employment.
- (7) Unions are associated with higher wages and nonwage shares of compensation and with lower turnover and less dispersion of pay. Estimates of the union effects on profits and productivity differ across countries.
- (8) Cross-country regressions yield inconclusive results on the impact of labor regulations on growth, while studies of country adjustments to economic shocks, such as balance of payments problems, find no difference in the responses of countries by the strength of labor institutions.

The research has led analysts at the World Bank and other international institutions to moderate their initially negative assessments of labor institutions.² Readers familiar with the retreat of the Bank and the IMF from dogmatic edicts on free trade, unrestricted capital flows, and laissez faire policies will note that this fits with the new modesty of these institutions about what economists can scientifically assert about growth-inducing policies.³

- (9) Labor institution can be critical when countries experience great change, as in China's growth spurt and Argentina's preservation of social stability and democracy after its 2001–2002 economic collapse. Cooperative labor relations tend to produce better economic outcomes.
- (10) In the 1990s–2000s, the informal sector's share of employment increased or held steady in virtually all developing countries, including those with healthy growth and limited regulations (Section 8). Even without deregulating the formal sector, an increasing proportion of workers in developing countries are working in largely unregulated markets.

1. THE DEBATE OVER LABOR INSTITUTIONS

Developing countries, like advanced countries, evince substantial differences in labor institutions that *could* impact economic outcomes and growth. To quantify this variation, I summarize in Table 1 the mean and standard deviation of five measures of the institutional orientation of formal sector labor markets. The five measures are: the labor component from the Fraser Institute's (2006) index of economic freedom; the Botero, Djankov, La Porta, López de Silanes, and Shleifer. (2004) indices of the strength of employment laws and laws regarding collective rights; the power of firms to set wages and hire and fire reported in the World Economic Forum's Global Competitiveness Report (2006), and rates of unionization from the ILO (1997). I have scaled the indices so that high values mean that a country relies more on market forces than on institutions in determining outcomes. I differentiate developing countries by

Table 1 Average values (standard deviations in parentheses) of measures of labor institutions in developing countries, level of income compared to developed countries, circa 2005 (e.g., see the correlations among the six indicators in the second part of the table)

Income group	Fraser Institute	Employment law	Collective relations law	Wage setting	Hiring and firing	Union density
Low income	5.410 (1.796)	6.08 (1.831)	6.260 (1.233)	5.224 (1.219)	3.906 (0.755)	13.208 (7.207)
Lower middle income	5.400 (1.202)	5.011 (1.802)	5.428 (1.255)	4.997 (0.664)	3.793 (0.716)	12.705 (8.481)
Upper middle income	5.723 (1.017)	4.937 (1.705)	5.174 (1.049)	4.945 (0.833)	3.532 (0.748)	26.389 (14.354)
All developing	5.501 (1.335)	5.292 (1.813)	5.575 (1.240)	5.036 (0.874)	3.738 (0.740)	17.855 (12.452)
Traditional developed	5.696 (1.374)	4.795 (2.224)	5.433 (1.562)	4.248 (1.174)	3.561 (1.034)	33.552 (19.723)
Recently developed	6.364 (1.696)	5.814 (1.852)	6.000 (0.940)	5.600 (0.711)	4.179 (0.965)	29.589 (16.914)

The correlations among the six indicators					
	Fraser institute	Employment law	Collective relations law	Union density	Wage setting
Fraser Institute	1.000				
Employment law	0.435	1.000			
Collective relations law	0.390	0.487	1.000		
Union density	0.092	0.317	0.113	1.000	
Wage setting	0.346	0.211	0.139	0.171	1.000
Hiring and firing	0.462	0.260	0.424	-0.094	0.507

Source: Tabulated from country data in Fraser (2007), Botero et al. World Economic Forum as given in appendix table, with countries classified by World Bank income levels. Recently developed countries include high-income countries outside the major Western countries.

level of income and distinguish the traditional advanced countries (the West and Japan) from the recently developed Asian Tiger economies. Appendix A gives the measures for each country.

The average values of the measures in Table 1 show that developing countries have lower union density and are more likely to rely on firms/markets than collective bargaining/regulations to set wages compared to advanced countries. The Botero et al. measure of employment law shows stronger laws in developing countries than in advanced countries, but the measures of hiring and firing practices, collective relations law, and the overall Fraser Index show little difference between advanced and developing countries. The newly industrialized Asian countries are usually the most market oriented. The correlation matrix at the bottom of the table shows that the four indices are positively correlated, which indicates that they are measuring broadly similar phenomenon, but the correlations are lower than what one might expect if they were good measures of the same underlying institutional structures. They could reflect genuine differences in country policies or practices along the dimensions that each index covers or it could reflect substantial measurement error in the indicators. Union density has a lower correlation to the four measures than they have among themselves.

The standard deviations below the means show considerable variation in institutions within groups. In five of the six measures, the variation is smaller among the developing countries than it is among the advanced countries. This is due to the divergence between the market-oriented United States and the other English-speaking countries and the more institution-oriented countries in the European Union (Freeman, Boxall, & Haynes, 2007). Finally, to the extent that institutions distort the operation of labor markets per the 1990 World Bank statement, the measures suggest that labor markets work better in some developing countries than in some advanced countries. By the Fraser Institute's index for labor, for example, the Ugandan labor market should work better than the German or Swedish labor markets.

Since institutions usually reduce dispersion of earnings across and among groups, another way to assess the importance of institutions in advanced and developing countries is to compare the dispersion of earnings among nominally similar workers. Figure 1 graphs the standard deviation of ln earnings among occupations and industries in countries by GDP per capita. It shows greater dispersion in lower-income countries than in high-income countries. To the extent that high dispersion reflects informational or other market failures, institutions have greater scope to improve outcomes in developing countries than in advanced countries. To the extent that the lower dispersion reflects institutionally enforced compression of wages from ideal market levels, on the other hand, the higher dispersion in low-income countries might reflect better performing labor markets, per the Fraser Institute rating of the Ugandan market compared to those of EU countries.

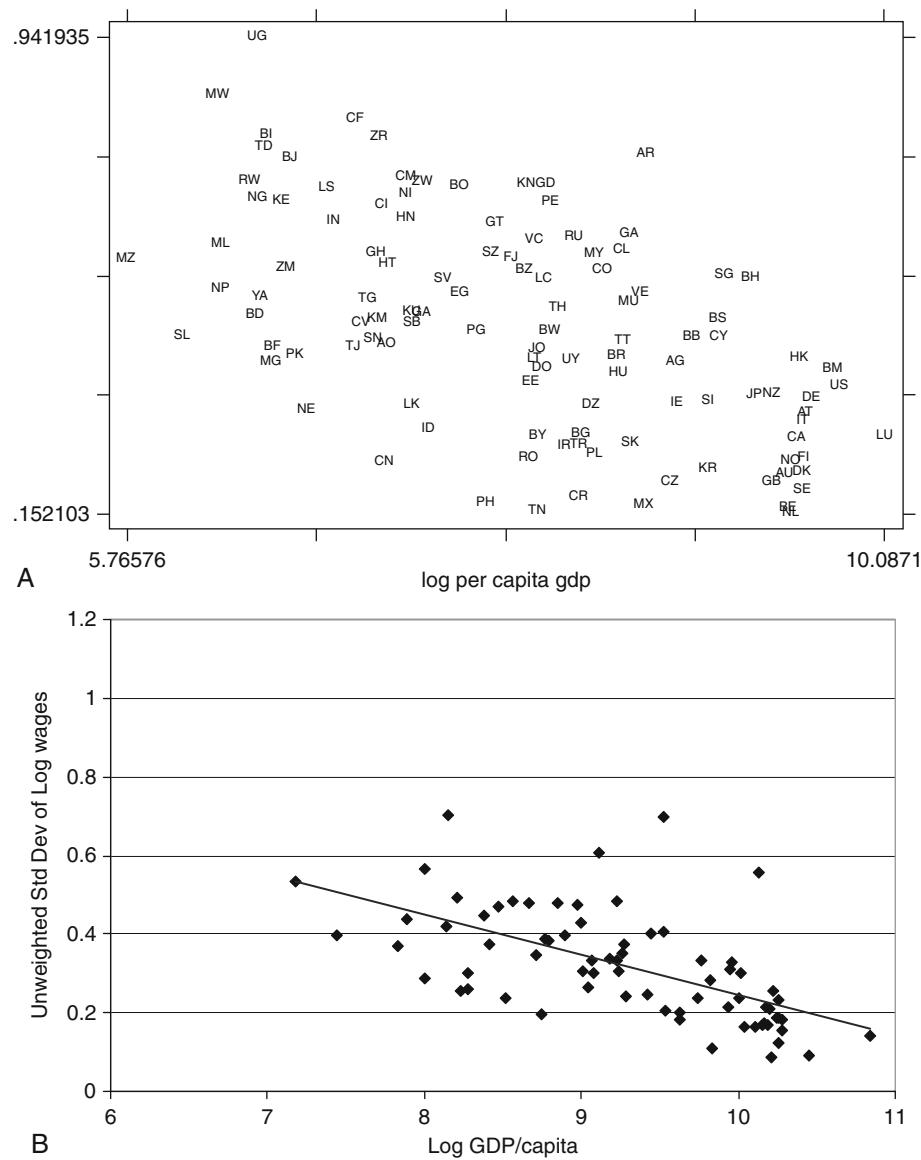


Figure 1 Dispersion of earnings by occupation and industry, by income level of countries (A) standard deviation of ln earnings by Occupation (Freeman & Oostendorp, 2000). (B) Standard deviation of ln earnings by industry (ILO. LABORSTA Internet. Data for 2006, Table 5B).

1.1 Theoretical perspectives

Economists use three types of theories to analyze labor institutions.

The first, which I have labeled *distortionism*, views institutions as distorting otherwise ideal competitive market equilibrium. Consider analyses of unionism in a market where wages are at the competitive level and the union bargains for higher wages. The

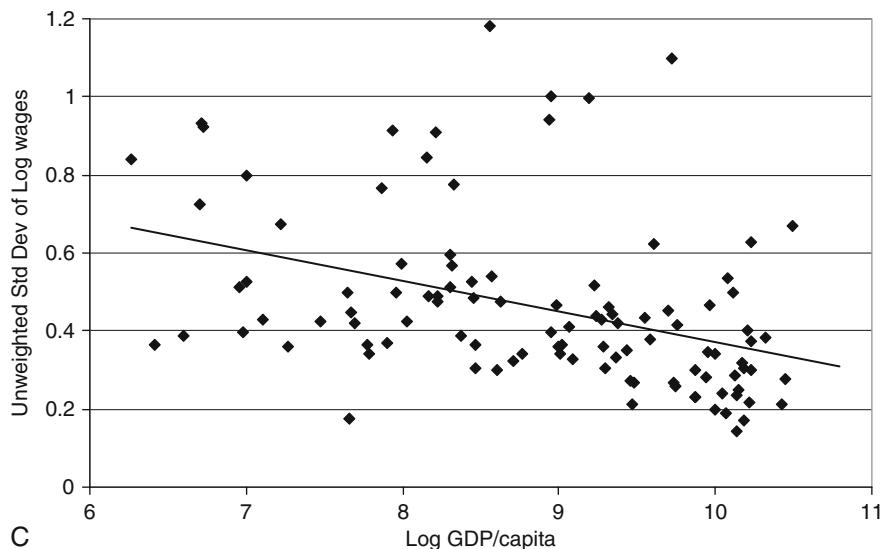


Figure 1, cont'd (C) standard deviation of ln earnings by industry (UNIDO data). Source: A, Freeman and Oostendorp (2002). B, Penn World Table (v6.2); ILO (2006). C, Penn World Table (v6.2); UNIDO Industrial Statistics (2006).

higher cost of labor leads unionized firms to reduce employment, which forces some workers to move to lower-paid, less productive nonunion work, lowering economic efficiency. The higher the elasticity of demand for labor, the greater is the distortion in resource allocation.

The second type of theory treats institutions as mechanisms for efficient bargaining. Models of efficient bargaining predict that when firms/workers bargain they “leave no money on the table” and thus allocate resources optimally. This is the Coase Theorem at work in the world of labor institutions (Freeman, 1993b). This analysis suggests that institutionally determined rules, such as employment protection legislation, affect distribution but not production. More modestly, it suggests that through legal arrangements or shadow economy side payments, there are “natural limits to the efficiency losses engendered by such regulations” (Squire & Suthiwart-Narueput, 1997).

The third type of theory focuses on ways institutions facilitate the flow of information and foster cooperative behavior that *could* raise productivity. In this vein Freeman and Lazear (1995) modeled works councils as institutions that increase communication inside firms and allow management and labor to make more informed and presumably better decisions. In addition, when dispersion of pay is high for noncompetitive reasons such as monopsony (Manning, 2003), informational failures, or other factors, collective bargaining or government regulations can bring wages closer to the market-clearing level.

In sum, there are arguments that institutions reduce efficiency, do not affect efficiency, and raise efficiency. To determine which arguments are valid requires evidence

on the links between institutions and outcomes that isolate the effect of institutions compared to other economic forces. On the basis of extant microdata and statistical tools, I use a five-fold sieve for evaluating research on the impact of institutions:

- (1) *The institution should affect the targeted outcome.* For instance, if the policy is a minimum wage, and the minimum is enforced, it should change the distribution of wages, producing a spike in frequency around the minimum. If an institution does not affect the price or quantity on which it is targeted it is likely to be a pro forma symbolic policy that is not implemented. For example, a government may have strong labor codes consistent with the Conventions of the International Labor Organization but allot no resources to enforce those codes and not impact outcomes.
- (2) *Evidence that the institution alters quantities.* Assuming that an institution affects the targeted outcome, it is critical to determine whether it reallocates resources or simply alters wages with no discernible impact on quantities. If it reallocates resources it is potentially distortionary. If it alters wages but not quantities, it could be efficiently redistributing income while leaving output unchanged, per efficient bargaining models. In the minimum wage case the critical evidence is the effect, if any, on employment.
- (3) *Evidence that the outcome attributed to the institution did not occur in another setting in the absence of the institution.* This applies a difference in difference analysis by treating the setting in which the new institution operates as the treatment and some setting where it does not operate as the control. Since many countries, sectors, or workers without an institution can serve as a counterfactual, there is a danger that the selection of one counterfactual rather than another may determine the conclusion. To minimize such dangers, the analysis should demonstrate that outcomes in the control were highly correlated with outcomes in the affected group *before the institutional change* (Abadie, Diamond, & Hainmueller, 2007). Otherwise, the control is unlikely to be a good guide to how the affected group might have fared in the absence of the policy.
- (4) *Evidence that the estimated effect of the institution is sufficiently large to affect aggregate development.* In advanced countries most workers are in the formal sector so that finding that an institution affects outcomes means that it will have a widespread impact. In developing countries where the share of the work force in the formal sector is small institutions are unlikely to affect aggregate outcomes unless they have *very large* effects on the formal sector, sizeable spillovers to the informal sector, or are located in sectors that may be particularly important for economic development, for instance traded goods.
- (5) *Identification of an institutional innovation from outcome data.* This is a “double blind test” in which the researcher looks at outcome measures as if he/she had no

knowledge of any institutional change and identifies the period/country where the institution was at work. It is based on Andrews' (1993) test for structural changes in time series data with unknown change points. Ideally, the outcome data show a break that lines up with the change of policy that may have caused the break. In cross-country time series growth data, the question would be whether the growth record of countries would identify those that adopted economic reforms that are supposed to help growth. It tests whether any innovation is sufficiently important to affect outcomes when other factors are also at work.

2. MICROEVIDENCE ON MINIMUM WAGES

The response of employment to minimum wages depends on the shape of the labor demand schedule in the range of the minimum. For all its contentiousness, the literature on minimum wages in the United States and other advanced countries generally finds modest responses in employment (Card & Krueger, 1995; Neumark & Wascher, 2006). The evidence that employment responses are often negligible does not mean that demand curves do not slope downward or that a high minimum wage cannot decimate employment. Rather, it suggests that governments set minimum wages with due consideration to the risk that minima can cause more harm than good.

The studies summarized in Table 2 find that in many developing countries, minimum wages raise the pay of low-paid workers in the formal sector adequately to produce spikes in the distribution of earnings and that changes in the minimum change the wages for low-paid covered groups. Studying compliance in Indonesia, Harrison and Scorse (2003) report that during the 1990s' epoch of globalization compliance increased in both multinationals and domestic firms. With respect to employment, most studies find that minimum wages had modest adverse effects on employment, with however considerable variation across countries and studies, even by the same analyst. Alatas and Cameron (2003) and Rama (2001) found that employment effects in Indonesia were limited to small firms. Bell (1997) found larger employment effects in Colombia than in Mexico. Analyzing Brazil, Lemos (2007) found no employment effects in the public or private sectors, while earlier reporting small negative long run employment effects and modest negative employment effects in the formal and the informal sectors (Lemos, 2004a, 2004b). Gindling and Terrell (2007a) estimated an elasticity of employment to the minimum in Costa Rica of -0.10 but estimated an elasticity of employment to the minimum in neighboring Honduras greater than -1.0 (Gindling & Terrell, 2007b). Studies that use panel data to identify the workers directly impacted by the minimum find larger employment effects than studies that estimate the effect from employment statistics for a wider group, but whether the differences among countries reflect differences in labor markets and in enforcement or in research designs is difficult to tell.

Table 2 Studies of minimum wage in developing countries, 1990s-2007s

Area (study, year)	Nature of data	Findings
BRAZIL Lemos (2006, 2004)	Individual data	Compresses distribution in informal and covered, 2006 no effect on employment; 2004 some
Chile , Montenegro and Pages (2004)	Changes in minimum, 1960–1998, individual data	Minimum shifts employment from young and unskilled workers to older and female workers
Costa Rica Gindling & Terrell (2007, 2005, 1995)	Individual data	Raises wages in informal + formal 1995 study: no loss of employment in formal; gain in informal; 2007, minimum lowers employment in public sector
Honduras , Gindling & Terrell (2007)	Individual data to establish spikes, sectors and size of firm over time, 1990–2004	1% increase in the minimum wage increases wage in medium and large firms by 0.29% in the average wage; reduces employment by -0.46%; does not affect small firms where employment grows
Colombia , Arango and Pachon (2004)	Panel date for cities, 1984–2001	Improves earnings of families high in income distribution, reduces it in bottom quintile; reduces employment, especially for women, young, less educated
Colombia, Mexico Bell (1997)	Firm-level data	No effect in Mexico; effect in Colombia;
Mexico (Bosch and Manacorda (2007)	Individual	Minimum shifts distribution for uncovered as well as covered; small employment effect
Mexico (Feliciano, 1998)	Cross-state panel data, 1970–1990	Decline in real value of Mexican minimum wage; increased employment of women with elasticity between -.58 and -1.25.
Trinidad & Tobago Strobl & Walsh (2003),	Individual with longitudinal job loss	Spikes shows that min affects wages, lowers employment of affected group
Puerto Rico Castillo-Freeman & Freeman (1992),	Spike, Uses imposition of US min wage	Modest effect on employment with elasticity of about -0.10, large fall in employment in very small industries

Table 2 Studies of minimum wage in developing countries, 1990s-2007s—Cont'd

Area (study, year)	Nature of data	Findings
LA + Caribbean, Kristensen & Cunningham (2006),	Individual data, focus on spikes	In 10 countries minimum affects informal and covered; 4 it affects only informal
LA Maloney and Nunez Mendez (2004), LA	Individual wage to establish spikes in formal and informal sectors Longitudinal job loss	Affects distribution in 6 of 8 LA countries Stronger effect on informal in Brazil, Mexico, Argentina, Uruguay; employment losses in Columbia; job loss greater for low wage
LA Maloney et al (2001)	Spikes in wage data	Four yes, three no
LA , Lustig & McCleod, 1997, x-country	Aggregate poverty 1990s vs 1980s	Reduces poverty
Indonesia (Alatas & Cameron, 2003)	Individual W spike, firm E geographic time series	Wage effects; no E in large firms; some in small
Indonesia (Harrison & Scorse, 2003)	Plant level data over time	Compliance with minimum wage rises 1990-1999 with both multinationals and exporting firms more likely to comply with labor standards
Indonesia (Rama, 2001)	Time series	Modest wage, little effect on aggregate employment but shift in employment from small firms
Indonesia (SMERU, 2001)	Individual data from National Labor Force Surveys; Firm-based survey; province level panel for employment 1988-1999	Spikes in minimum wages for blue collar workers but not others; compliance rises with firm size and increased over time; estimated elasticity of total employment—0.112, larger for females, youths, less educated;
Ghana Jones (1997)	Individual wage data; time series 21 years	Spike at minimum in distribution; in informal as well as covered. W, E shift to informal elasticity of about -0.10
Many countries (Squire & Suthiwart- Narueput, 1997)	Changes in minimum among countries, 1970-1990	Real minimums fall in 16/23 Min/Average falls in 6/17

That most studies obtain elasticities considerably less than unity imply that increases in the minimum have the potential for being a viable antipoverty tool. Analyzing the link between minimum wages and poverty in 22 developing countries Lustig and McCleod (1997) find that changes in real minimum wages are associated with declines in poverty, which requires that the elasticity of demand be low. The high elasticity found for Honduras is a striking exception, indicating that each country case (and each potential change in the minimum in each country) must be considered carefully.

Studies that have examined the pattern of change in minimum wages as an endogenous variable raise doubts about the standard assumption that changes in minimum wages are exogenous to market conditions. Looking at changes in real minima wages in 23 developing countries in the 1980s–1990s, Squire and Suthiwart-Narueput found declines in the real minimum in 16 countries. Presumably governments allowed real minima to fall because they believed that in times of economic crisis falling real minima saved jobs. Rama (2000) shows that African counties in the CFA zone changed minimum wages in response to changes in the terms of trade, national output, and consumer prices.

The big surprise in studies of minimum wages in developing countries is a substantial body of evidence from Latin America that minimum wages *raise* wages in the informal sector as well as in the covered sector (Gindling & Terrell, 2005, Kristensen & Cunningham, 2006; Lemos, 2004b; Maloney & Nunez Mendez, 2004). In Brazil this is known as the “Efeito Farol” or lighthouse effect. In some cases, increases in the minimum appear to raise wages more in the informal sector than in the formal sector. This effect runs counter to the model that economists often use to analyze minimum wages, in which labor displaced from covered employment moves to the uncovered sector and depresses wages there.

One interpretation of the positive wage spillover is that workers base their reservation wages on the minimum, perhaps because the high dispersion of pay in developing countries provides little guidance as to what wage they might expect. If low-paid workers in the informal sector are paid less than low-paid workers in the formal sector, their wages would increase more if the minimum induced all workers to have the same reservation wage. But perhaps something else is going on that is not captured in the model⁴ or perhaps the measures of wages in the informal sector are poor. In any case, the results devastate the application of the standard model to understanding what minimum wages do in developing countries.

3. WAGE CURVE

The wage curve is an empirical relation between wages (w) and unemployment (u) that is usually written in log-log form: $\ln w = a + b \ln u + X$, where X refers to other factors that affect wages (Blanchflower & Oswald, 1994). Fit with cross-section data across regions,

the wage curve tests the prediction of the Harris-Todaro two-sector model that high wages induce unemployment due to the migration of rural labor to high-wage cities contrary to the finding in advanced countries that wages are lower when unemployment is higher as wages fall to equilibrate the supply-demand imbalance.⁵ If estimated wage curves had the opposite sign in developing countries from those in advanced countries this would signal a major difference between labor markets. Migration would dominate the link between wages and unemployment in developing countries whereas wage adjustments to unemployment would dominate the link in advanced countries.

Between 1993 and mid-2007 researchers estimated wage curves for 13 developing non-transition economies: Argentina (Galiana, 1999), Brazil (Amadeo & Camargo, 1997), Chile (Berg & Contreras, 2004), Mexico (Castro Lugo, 2006), Uruguay (Bucheli & Gonzalez, 2007) in Latin America; Burkina Faso (Lachaud, 1998), Cote D'Ivoire (Hoddinott, 1996), South Africa (Kingdon & Knight, 1998, 2006), in Africa; and Turkey (Ilkkaracan & Raziye, 2003), Korea (Blanchflower & Oswald, 1994), Taiwan (Rodgers & Nataraj, 1999) China (Sabin, 1999; Wu, 2004), and India (Bhalotra, 1993). These studies obtain negative coefficients on ln unemployment of the same order of magnitude as the coefficients in advanced countries—an elasticity of about -0.10. The negative estimated relation between unemployment and wages implies that the Harris-Todaro model does not represent labor markets in the developing countries in the period covered.⁶

This conclusion is consistent with analyses of labor market adjustments in the 1990s in Sub-Saharan Africa, the experience of which motivated the model. For South Africa Kingdon and Knight (2004) reject the two-sector story of urban joblessness on the grounds that unemployed workers do not forego informal sector jobs to search for a high-wage urban job. Similarly, Rama's (2000) analysis of the CFA African countries rejected "the hypothesis that labour market policies and institutions were the obstacles preventing wages in the formal sector from adjusting to a more unfavorable international context" (p. 495). Summarizing research on African labor markets, Kingdon, Sandefur, and Teal (2006) conclude that real wages were "more downwardly flexible than previously thought and...surprisingly responsive to unemployment rates," though large wage differentials remain between the formal and the informal sectors. Sizable flows of labor between the informal and the formal sectors in some countries also argue against the Harris-Todaro model.

In short, the evidence rejects the two-sector model that makes institutions the prime cause of divergences between earnings and productivity among sectors and of the large pay differences among industries and occupations shown in Figure 1. Given that the two-sector model does not fit reality, we must look elsewhere for the causes of the great dispersion of pay in developing countries: Efficiency wages or gift exchanges? Implicit profit sharing between firms and workers? Unmeasured differences in human capital? Transportation or communication problems that make the sectors separate islands in the economy and raise the costs of search? The door is open for innovative analyses.

4. EMPLOYMENT REGULATIONS

Employment protection legislation (EPL) seeks to protect the jobs of current employees by requiring that firms pay severance if they lay them off. Many laws also require the firm to gain approval for dismissals from a labor court or works council. Closely related regulations forbid firms from hiring replacements during a labor dispute and guarantee workers the right to return to work after the dispute. The regulations essentially deal with property rights at work—whether the worker or the firm “owns” the job. Regulations that give greater ownership to workers should reduce job turnover and tilt market outcomes toward experienced workers against new entrants, whereas those that give greater ownership to firms tilt outcomes toward capital and more mobile younger workers. Still, by making layoffs more expensive, EPL increases the cost of hiring workers (whom the firm may lay off in the future), which risks reducing total employment.

For advanced countries, two decades of studies spurred by the OECD have quelled the fears that EPL greatly affects employment. Most studies find that EPL laws redistribute employment to incumbent workers with no clear impact on aggregate employment or unemployment (OECD, 2004, 2006).⁷ For developing countries, research has just begun. In Latin America, job security provisions and costs of dismissals are extensive and thus more costly than in advanced OECD countries (Heckman & Pages, 2002). Even so, studies of EPL legislation that compare employment before and after changes in the law and/or between covered and uncovered groups in particular countries find effects on total employment in some countries but not in others. The strongest country evidence for effects of laws is for Colombia. In 1990 Colombia lowered severance pay and gave firms greater latitude to lay off workers. Using the Colombian National Household Survey, Kugler (2004) found that the weakened EPL was associated with growth of employment and a decline in job tenure in the formal sector relative to the informal sector and with increased job separations and hires in the formal sector. This churning induced a modest fall in unemployment duration and aggregate unemployment.

By contrast, studies of EPL legislation for Chile find no such effects in time series data (Edwards & Edwards, 2000) and household data (Montenegro & Pages, 2004). Montenegro and Pages find that EPL regulations shift employment from young unskilled workers to senior skilled workers as in advanced countries. Petrin and Sivasadan's (2006) analysis of establishment data shows no impact of Chile's 1991 and 1994 changes in employment regulations on employment, the number of hires and fires, or the speed of employment adjustment, though they estimate that wages and marginal productivity changed in ways that they attribute to changing EPL laws.⁸

Using firm-level datasets for 16 countries, including five Latin American countries, Haltwanger, Scarpetto, and Schweiger (2008) show that gross job flows (the sum of

jobs created and jobs destroyed) range from 25% to 30% and vary similarly by size of firm and industry across countries. Using the United States as the metric for what a relatively unregulated competitive labor market would produce, they find that flows in the other countries are lower than in the United States and relate the differences to measures of labor regulations. Latin American regulations are comparable to European Union regulations and both the Latin American and the EU countries have lower gross flows than the United States. However, it is the entry and exit of firms rather than the employment behavior of existing firms that underlies this difference, which they term “somewhat surprising from a theoretical perspective (p. 26).” Given the high level of gross job flows in all the countries, it is unclear whether any regulation-induced difference in job flows adversely impacts net job reallocation or efficiency.

Several studies have looked at how India’s labor laws affect outcomes in the formal sector, exploiting the fact that labor regulations vary across states and over time. Besley and Burgess (2004) created an index of changes in the laws governing employment protection and industrial disputes for Indian states from 1947 to 1997 and made before-after contrasts of employment and output in states that did/did not change the laws and between the formal and the informal sectors within states. Their main specification showed that proworker regulations induced firms to shift employment and output from the formal to the informal sector,⁹ although they report that results “are not robust to including state-specific time trends.” (Besley & Burgess, 2004, fn 21). Bhattacharjea (2006) criticizes the index as being based on: (1) misreading legal changes in three states (Andhra Pradesh, Maharashtra, Rajasthan); (2) coding laws so that passing three laws in 1 year in one direction counts less than passing the same laws in 3 separate years; (3) failing to take account of the way state labor laws interact with other labor laws. The result is that some of the patterns “stand out for anyone who is reasonably familiar with India”: Kerala, known as prolabor, is coded as proemployer while Gujarat and Maharashtra, known as proemployer, are coded as proworker.

Ahsan and Pages (2008) report, however, that the results are robust to Bhattacharjea’s amended measures. They find that the laws that regulate the procedures for the resolution of industrial disputes have more robust and greater effects on formal sector output and employment than those relating to employment protection legislation. Neither Burgess and Besley nor Ahsan and Pages find any gains to workers from the legislation, which raises questions about the purpose of these regulations: why have the states passed these laws if they benefit no one?. Finally, looking at the effect of regulations in the context of trade liberalization, Hasan, Mitra, and Ramaswamy (2007) find that states with greater regulations have smaller elasticities of labor demand, which suggests that the laws may protect jobs by reducing flexibility. Overall, Indian labor institutions come closer to fitting the 1990 World Bank view that regulations do more harm than good than the institutions in other countries, though their impact is limited on the aggregate economy due to the huge size of the Indian informal sector.

5. MANDATED BENEFITS

Mandated benefits are nonwage compensation (social insurance, compensation for injuries at the job, maternity leave, vacation or holiday pay, and so on) that governments require firms to provide workers or to pay taxes for government provision. Employers often complain about the cost of the benefits as if they are add-ons on wages. But if workers desire the benefit and if wages are flexible, workers will pay part and possibly the full cost through lower wages. Summers (1989) argued that mandated benefits can be a more cost-effective way to deliver public goods than government provision through general taxes. Studies of the incidence of mandated benefits for advanced countries show that workers bear the full incidence in some cases (Gruber, 1997a) and a sizable share of the cost in others (Gruber & Krueger, 1991; Ooghe, Schokkaert, & Flechet, 2003). Workers are more likely to bear a large proportion of the cost when they value the benefit highly.

Many developing countries mandate benefits for workers. Most Latin American and Caribbean countries use payroll taxes to fund retirement, work injury benefits, and health care in national social insurance systems. Most have paid vacations, and some have maternity leave as well as other benefits that could increase the cost of labor. Analyses of the incidence of mandated benefits in Latin America give different results by benefit and country. Gruber (1997b) estimated that when Chile switched funding social security from payroll taxes to general revenues, the reduction in payroll taxes was passed entirely onto workers in the form of higher wages. This implies that the payroll taxes had no effect on labor demand. But using a similar methodology in Colombia, Kugler and Kugler (2003) find modest shifting of the payroll tax, which implies that the tax raised the cost of labor in the formal sector and affected demand. When Colombia changed its mode of funding severance pay from a pay-as-you-go system to private accounts in which the employer deposited a proportion of the worker's wage, Kugler (2002) estimated that 60–80% of the cost shifted to workers in the form of lower wages. This suggests that under the pay-as-you-go system workers were dubious that firms would pay severance if financial difficulties forced layoffs, whereas the private accounts guaranteed that they would get severance pay.

6. UNION EFFECTS

Developing country unions range from China's All Chinese Federation of Trade Unions, a government-controlled agency that seeks to advance worker interests within bounds set by the Communist party, to Korea's aggressive independent enterprise unions, to Peronist unions in Argentina, to South Africa's COSATU, which played a major role in the anti-apartheid struggle. Most developing country unions are weak. Many are more involved in political activity than in collective bargaining. But sufficient number of unions in

developing countries engage in collective bargaining to enable researchers to examine what unions do to wages in the tradition of Lewis (1963, 1986) and what they do to non-wage outcomes in the tradition of Freeman and Medoff (1984).

Because data are limited, however, many studies estimate union effects from small nonrepresentative samples and most compare outcomes between union and nonunion workers/firms with similar characteristics rather than undertaking more methodologically preferable longitudinal analyses of workers or of workplaces that change union status. The one advantage of studying union effects in developing countries is that in some countries governments have changed drastically their treatment of unions in short periods of time—for instance from outlawing collective bargaining and restricting unions to allowing them to operate freely, which offers the rare opportunity to compare the functioning of a union-free economy to an economy with free unions.

Table 3 summarizes the findings on the link between unionism and economic outcomes in the countries for which I found quantitative studies.

The most extensive work is on Mexico. Estimated union/nonunion wage differentials range from 5% or less to 10–15%, with unions having their biggest effect on the wages of lower-paid workers. This implies that they reduce wage dispersion and inequality. Popli (2007) confirms this result but reports that the lower dispersion of pay between union and nonunion workers narrowed from 1984 to 2000. Comparing unionized and nonunion establishments, Fairris (2005, 2006) finds that unionism is associated with a higher share of compensation to nonwage payments and more training of workers; with lower quits in foreign-owned firms (though not in Mexican-owned firms). Union firms have higher productivity and comparable profits than in nonunion firms. In contrast, Maloney and Ribeiro (1999) estimate that unions increase employment in Mexico, which suggests that they lower productivity. In the absence of longitudinal changes in union status, however, these patterns could reflect selectivity of unionized firms rather than any causal impact of unions on firm performance.

For Brazil, Arbache estimates a union wage premium in manufacturing on the order of 5–7%, but also finds, contrary to most studies, higher wage dispersion among union workers than among nonunion workers. Menezes-Filho, Zylberstajn, Chahad, and Pazello (2005) estimate that union density within Brazilian manufacturing firms is associated with a union wage effect of 12%, and that unionized firms have lower productivity and profitability but that union firms that introduced profit-sharing schemes had substantial increases in productivity and profits. The studies for other Latin American countries, based on small surveys, find that unions are associated with lower productivity or profits. The study for Peru reports lower profits but comparable productivity in unionized firms as in nonunion firms, which implies that the unions must have gained higher wages. The study for Guatemala finds 10% lower productivity among unionized coffee plantations in a sample of 37 large plantations but finds no union/nonunion differential in a fixed effects analysis that examines changes in productivity in the five plantations that changed union status over the period.

Table 3 Studies of union effects in developing countries

Country	Wage	Dispersion	Turnover	Benefits	Productivity	Profits	Other
Mexico	10-15%	Lower	Lower	Higher	Higher	None	Increased training
Uruguay	5%		Lower		Higher	Higher	Lowers elasticity of demand
Guatemala					Lower in cross-section, not in fixed effects		
Peru					Modest negative but not significant effect	Lower in all cases	Higher capital labor ratios
Brazil	5-12%	Higher			Lower, unless with profit sharing	Lower unless with profit sharing	More likely to go to profit sharing
South Africa	10-20%, bigger for blacks	Lower		Higher			Lowers coefficients on other variables in earnings
Ghana	6-16%	Lower					
Cameroon a	14%	Lower					
Cameroon b	-8% to -11%						

Senegal	-13%				
Zimbabwe	-17% 14%				
Malaysia	15–20%	Lower	Higher	Higher	Increases training
Korea	3–7% bigger for women	Lower, but nonstandard	Lower	Higher	Works council raises Unions lower, WC raises job sat; substitute nonstandard

Source: Mexico: Fairris (2003, 2005), Popli (2007), Malone and Ribiero (1999), Panagides and Patrinos (1994); Uruguay: Cassoni, Labadie, and Fachola (2005), Cassoni, Allen, and Labidie (2004); Peru: Saavedra and Torero (2005); Brazil: Menezes-Filho et al. (2005), Arbaché, 1998; Guatemala: Uriar and Lee (2005). South Africa: Kingdon, Sandefur, and Teal (2004), Butcher and Rouse (2001), Hofmeyr and Lucas (2001), Schultz and Mwabu (1998), Moll (1993); P-C: Michaud and Vencatachellum (2001); Ghana: Blunch and Verner (2004), Verner (1999); Cameroon a: Tsafack-Nanfoso (2007), 1999 survey of 1074 wage earners; Cameroon b: Miller and Vallee (1995), Vallee and Thomas (1994), 1993 and 1994 survey; Senegal: Terrell and Svejnar (1989); Zimbabwe: (Verner, 1999); Malaysia: Standing (1992); Korea: Lee and Na (2004), Fields and Yoo (2000), Park (1991), Cheon (2006)

Quantitative studies of unions in Africa have concentrated on wage effects. In South Africa, unions appear to raise wages by 10–20%. Since South Africa has high unemployment, one might be tempted to blame the unemployment on high union wages per the Harris-Todaro model, but Butcher and Rouse note that the union share of the South African work force is too small for the estimated wage effects to explain the country's unemployment. The wage curve for South Africa shows an inverse relation between wages and unemployment. In other African countries, Blunch and Verner (2004) estimate that unions raise wages by about 6% in Ghana, with the effect coming largely from higher wages for lower-paid workers, while Verner reports a larger union–nonunion differential in Ghana of 16%. Tsafack-Nanfoso (2007) estimates a union/nonunion wage differential of about 14% in Cameroon¹⁰ and finds that the standard deviation of log earnings among unionized workers is considerably lower than among nonunion workers.

Three studies find lower wages among unionized workers in Africa. Estimates for Zimbabwe show higher productivity as well as lower pay in union workplaces, which is not easily reconcilable with other research (Verner, 1999). Estimates for Cameroon give lower union wages in 1993 and 1994 (Thomas & Vallee, 1996). Estimates for Senegal find lower union wages in 1980–1985 (Terrell & Svejnar, 1989). Since it makes little sense for independent unions to negotiate lower wages for members, in all these cases, the unions are presumably not “normal unions” doing collective bargaining. At various times, governments in these countries suppressed independent unions or ran unions as appendages of the governing party.¹¹ What is needed is a study of the circumstances and behavior of African trade unions to inform the quantitative analyses.

Turning to Asia, Malaysian unions operate primarily at the enterprise level because the government restricts industrial unions. Using establishment data, Standing (1992) finds that unions reduce skill differentials and lower quits, and are associated with higher productivity, more product and process innovations, more firm-sponsored training, and greater enrollment in a pension plan. The differences between the outcomes for firms with and without unions far exceed the differences in outcomes between firms with plant level unions and those that had industrial level unions. Bhandari's (2008) data on the wage differences between union and nonunion Indian workers based on a small (551 persons) sample reports differences of wages 6% for contractual workers and 25% for permanent workers.¹²

Like Malaysia, Korean unions are largely enterprise based, with membership restricted to regular employees. Studies compare wage and other outcomes between regular employees in firms with and without collective bargaining agreements, and between nonregular workers between those firms as well. The estimated impact of Korean unions on wages is in the low single digits. Fields and Yoo (2000) found that the union differential doubled during the period of rapid union growth, but this was from a modest 3% to a still modest 6%. Park (2006) reports a union differential of 5% for all workers, larger for women (12%) than for men (2%), but also finds that

nonstandard workers, who are excluded from the enterprise union, are paid less in union establishments. Unionism is also associated with greater employment of nonstandard workers, suggesting that firms substitute them for more expensive union labor. Using a 1991 survey of establishments, Kleiner and Yee (1997) estimated the impact of unions and works councils deemed "effective" by management and labor leaders on nonwage outcomes. They found lower turnover, similar productivity, and lower job satisfaction among workers in the unionized settings. By contrast, they found that works councils were positively associated with productivity and satisfaction. Both unions and works councils appear to have affected decisions about terminations and downsizing.

The country which experienced the most striking change in union status is Uruguay. From 1973 to 1984 Uruguay's military dictatorship outlawed collective bargaining. The return of democracy gave unions the right to bargain with employers, first under a tripartite system in which unions, management, and government bargained together (1985–1991, covering some 60% of production workers), and later by management and unions without government involvement. Labor market outcomes differed with these regime shifts. During the dictatorship real wages fell 49% while the rate of unemployment varied cyclically, reaching 16% in 1983. Thereafter unions raised wages, which induced firms to invest in capital. Unionized firms had sufficiently higher productivity that they suffered no profits loss (Cassoni, Labadie, & Fachola, 2005). Using establishment data, Cassoni, Allen, and Labadie (2004) found that elasticities of employment to output and wages were lower with collective bargaining than in the union-free environment, suggesting that unions reduced employment flexibility.

Korea, like Uruguay, has had distinct regimes governing unionization. Before Korea achieved democracy in 1987, the government allowed only an official union movement to operate and used state police powers to assist employers in labor disputes. Democracy brought with it a near doubling of union membership, but union density fell in the 1990s and 2000s (Jeong, 2007, table 2.1). Some Korean unions formed a second more militant federation, and the older federation became more aggressive as well. Governments have tried unsuccessfully to ameliorate the intensity of disputes by establishing Tripartite Committees for social dialogue and requiring that firms introduce works councils. Wages increased greatly during democraticization and income inequality fell, but the 1997 Asian financial crisis weakened union strength. Real wages fell by 9.3% in 1998, as unions agreed to low nominal wage changes despite high inflation. Jeong's (2007, chap. 4) analysis of collective bargaining contracts shows that the unions shifted from seeking wage gains to protecting jobs. Lee and Na's (2004) 1999 survey of 300 firms found only a weak union impact on firm responses to the crisis: unionized firms were more likely to downsize through retirements than layoffs.

Finally, taking a more aggregate approach, Feldmann (2009) related the reports of executives in the World Economic Forum's Global Competitiveness Report on the

quality of labor-management relations to aggregate unemployment. With measures of labor relations for 1995–2003, he conducted a longitudinal regression analysis with country-specific random effects that shows that cooperative relations reduce unemployment, as does flexible labor regulations as reported by executives.

In sum, research finds that unions are associated with higher wages (save for the African studies noted), higher nonwage compensation, lower dispersion of earnings (save for the Brazilian study noted), reduced quits, and greater training. But the studies show varying relations between unions and productivity or profitability across studies and countries. While this variation is consistent with the notion that what unions do differs across countries, it also reflects differences between the often small surveys and groups covered and limitations of the study designs.

7. CROSS-COUNTRY EVIDENCE

Many economists are dubious about what we can learn from cross-country growth regressions.¹³ The datasets are small, the measures of institutions are often weak, and our knowledge of economics is too limited to identify the “right” model for analyzing how an institutional intervention affects an entire economic system. Analysts have trouble pinning down the effects of education or trade on growth in cross-country regressions (Krueger & Lindahl, 2001; Rodriguez & Rodrik, 1999), so what hope is there of uncovering the more subtle effects of labor institutions? These are valid concerns but since the bottom line in development is growth, there is something to be learned from regressions of growth rates on measures of labor institutions.

One strand of cross-country analyses focused on labor institutions uses data from the World Bank’s Database of Labor Market Indicators Across Countries (Rama & Artecona, 2002). This database contains measures of indicators of institutions such as the number of ILO conventions a country has ratified, the level of minimum wages, mandated days of maternity leave, union and government shares of the work force. Because the data are sparse, Forteza and Rama (2001) average the four measures to form an aggregate indicator of institutional “rigidity” for 1970–1999 and warn analysts that “*time-variant indicators of labor market rigidity cannot be used in the empirical analysis.*”

By the indicators measure, the ten most flexible labor markets are: South Africa, Uganda, Zimbabwe, Kenya, Tanzania; and China, Hong Kong, Jordan, Indonesia; and Chile. The ten most rigid labor markets include France, Sweden, Belgium, Denmark; and four former Soviet bloc countries, Hungary, Belarus, Krygyz Republic, Bulgaria; and Algeria and Uruguay. I list the countries to highlight the problems with categorizations of this type. There is remarkable institutional variation among countries at both ends of the flexibility spectrum. In the flexible group, South Africa has strong trade unions and an extensive labor code based on those in Europe, whereas unions could barely operate in Idi Amin’s Uganda or Robert Mugabe’s Zimbabwe; China did not have a working labor

market until the 1990s while Hong Kong had a free market under British law. In the inflexible group, Belarus is a residual Soviet-style dictatorship, Hungary is part of the EU, Denmark is famed for high labor mobility and “flexicurity,” while France relies extensively on minimum wages. Changes in labor regulations in Uruguay, the ex-Soviet countries, the African countries, and China suggest that no single index can capture their institutions over the entire period. Given these measurement issues, institutions would have to be incredibly powerful to show up as determinants of growth in cross-country regressions.

Forteza and Rama (2001) test whether their measures of rigidity were associated with the success of World Bank adjustment programs. They do this by interacting the indices with dummy variables for the timing of World Bank credits and loans and entering the interaction terms into regressions of the growth of GDP per capita. The idea is that countries that obtaining Bank loans/credits should increase growth more if they have less rigid labor markets.¹⁴ Growth regressions with fixed country effects yield, however, similar coefficients on the labor rigidity index before and after receipt of Bank aid.¹⁵ Decomposing the index into its separate parts, Forteza and Rama find that different indicators have different relations with growth: “relatively high minimum wages are associated with better economic performance before adjustment” (p. 24), while union and government shares of the work force were associated with slower growth before and after the receipt of loans.¹⁶

Ignoring the warning that the Labor Market indicators data should not be used as time-variant measures, Calderon and Chong (2005) create rigidity indices over time to conduct a panel analysis of growth rates. For developing countries their panel regressions with country fixed effects show that the number of ILO conventions has a large negative effect on growth, while the average of indicators has little effect. Models estimated with a generalized method of moments with instrumental variables show stronger negative effects on growth,¹⁷ but regressions that relate growth over the entire period to the average indicators find that one of the four measures is *positively* related to growth while the others have no significant relation.

Botero et al. (2004) and Calderon and Chong (2005) use the Botero et al. measures of labor laws introduced in Table 1 to analyze aggregate outcome. Botero et al. report that across all countries in the dataset the employment law and collective action indices are associated with lower utilization of labor overall and with higher unemployment for young workers. Calderon and Chong (2005) regress growth rates on the Botero measures of labor codes and obtain an insignificant coefficient on the index of employment laws but a positive coefficient on an index for social security laws, that at face value implies that countries with social security laws grow more rapidly than others (Table 6).

The labor subindex of the Fraser Institute, which the Institute reports at 5-year intervals from 1970 to 2000 and yearly thereafter, provides another source of data on labor institutions across countries. To see whether the labor subindex is associated with economic growth I estimated the following equation:

$$\Delta \ln \text{GDP} = a + b \text{INDEX} + c \text{LABOR} + d \ln \text{GDP}(-1) + Dc + Dt + u_{ij}$$

where $\Delta \ln \text{GDP}$ is the annual growth rate in $\ln \text{GDP}$ per capita over each period, 1970-1975, 1975-1980, and so on through 2000-2004; INDEX is the Fraser index of economic freedom excluding the labor subindex; LABOR is the subindex of measures of minimum wages, hiring and firing practices, collective bargaining, unemployment benefits, and use of the draft; Dd are country dummy variables; and Dt are time dummy variables.

If market-driven economies improve growth, the b coefficient in this regression should be positive. Similarly, if labor institutions harm growth the c coefficient will be positive. The regression, which is summarized in Appendix B, yields a significant positive coefficient on the INDEX variable for developing and advanced countries, which supports the notion that market-oriented economies enjoy higher growth. But for both sets of countries the coefficient on the LABOR measure is negative and insignificant, which rejects the hypothesis that labor institutions reduce the rate of growth.

In sum, there is no strong support for the proposition that labor institutions affect economic growth positively or negatively.

8. COUNTRY CASES

Case studies of country experiences are the methodological polar opposite of cross-country regressions. Country studies allow analysts to use country-specific data rather than generic cross-country indicators and to situate their interpretations of the data in hopefully deep knowledge of country institutions and practices. But country cases are not a random sample from the countries in the world, and they invite attention to country specifics—great leaders, particular events—that do not readily generalize. Still, there is much to be learned from them.

In 1988 the World Bank initiated studies of labor market adjustments in twelve developing countries (Chile, Argentina, Bolivia, Costa Rica, Brazil; Ghana, Kenya, Egypt, Cote D'Ivoire; and Korea, Malaysia, Thailand). The analysis (Horton, Kanbur, and Mazumdar, 1994) reached three conclusions: (1) “real wages were more flexible than generally supposed, which would support adjustment”; (2) “labor reallocation across sectors has been more or less in the desired direction” (i.e., toward tradable goods); and (3) “labor market institutions such as unions and minimum wages, often argued to be an impediment to adjustment, have more subtle effects on the workings of the labor market” (Preface, p. x). These conclusions fit with the econometric findings reviewed earlier and with the more measured view of labor institutions and economic outcomes that has emerged in recent years.

That labor institutions have modest impacts on economic performance on average does not mean that in particular situations they may not affect growth or help or hinder

economic adjustments. Some countries in the Bank study underwent major changes after the researchers completed their empirical investigations—the Chilean economy improved greatly; Argentina went from boom to bust; Cote D'Ivoire suffered civil war; East Asia suffered financial crisis. Consideration of how labor institutions performed under these circumstances would enrich the conclusions of the Bank studies. Much might also be learned by analyses of cases where institutions appear particularly harmful or beneficial to the economy or of institutions in countries with great growth failures or successes. From this perspective I examine next institutions in China's move to a market economy and in Argentina's recovery from economic meltdown.

8.1 China

China's transformation from communist planned economy to a market economy and entry to the global trading system is one of the most important developments in recent economic history. As part of this shift, the country radically changed its labor market institutions. From the 1960s through the late 1980s, China did not have a working labor market. The state/party controlled the demand and supply of labor (Walder, 1986). The state was the primary employer. It set pay through a national pay grid and mandated benefits such as health care, retirement, and housing. It controlled hiring and firing and assigned workers to jobs. It limited migration through Hukou residency restrictions. Fearing that unemployed urban youth might create a social problem in the 1980s, the state encouraged older workers to retire by promising to assign their jobs to their offspring.

Shifting from communist planning to markets in the mid-1980s through the 1990s, China freed firms and workers to make demand and supply decisions. It gave management of state-owned enterprises authority to hire and fire and introduce performance-linked wages. It allowed greater private employment. In the 1990s it privatized many small and medium state-owned enterprises and laid off several million workers from state enterprises. It weakened the Hukou system. With opportunity to find work, buy food, and obtain housing in the market, upward of 100–150 million workers moved from rural to urban areas—the largest internal migration in history. Workers and firms matched themselves in the job market rather than relying on state assignment.

Could China have successfully transitioned to a market economy with its prelabor market institutions? Studies of wages, employment, and productivity during China's growth spurt suggest that the labor market reforms were an essential part of China's success (Knight & Song, 2005; Meng, 2000). Wages for skilled work increased, raising the return to education. Tens of millions found work in the private sector. Productivity rose in state-owned enterprises. It is hard to imagine these changes occurring with the state setting pay and assigning workers to jobs, though it is hard to imagine a counterfactual test for this assertion.

The new Chinese labor market produced a huge income gap between rich and poor, evinced in a Gini coefficient that reached US levels by 2005, disturbing Party leaders, who worried about “mounting public anger over inequality and corruption”

(Eckholm, 2001). To deal with these concerns, China enacted a new labor code in 2007 that required employers give written contracts to workers, restricted use of temporary laborers, made layoffs more difficult, and strengthened the power of the All China Federation of Trade Unions to organize and to bargain for wages and benefits. With World Bank (2006b) advice, the government sought to establish a national pension system. China's effort to develop labor institutions to reduce income inequality and insure against social disorder reflects a very different perspective on the role that institutions play in development than the fear that institutions are inimical to growth that motivates many economic studies.

8.2 Argentina

In the 1990s Argentina was the poster child of globalization and Washington Consensus policies. With the advice and assistance of the IMF, the country pegged the peso to the dollar, privatized many public enterprises, loosened controls on banking and foreign currency, and greatly increased the market orientation of its economy. Between 1990 and 2000, Argentina rose in the economic freedom index of the Fraser Institute (which measures reliance on market forces rather than institutions) from 84th in the world to 27th. Seemingly reflecting these changes, the Argentine economy grew rapidly in the 1990s, although with double-digit unemployment and increased inequality. At its fall 1999 meeting, the IMF lionized Argentine President Carlos Menem for his economic stewardship, and he responded by thanking the IMF for its guidance. Washington Consensus? Think Argentina.

Two years later, the Argentine economy collapsed. The value of the Peso dropped to one-third of the dollar, interest rates zoomed, real GDP fell by 18%, and unemployment rose to 21.5%. Poverty increased greatly. The government froze bank accounts. In winter 2001 Argentina went through five Presidents or acting Presidents in less than 2 months. Angry protestors filled downtown Buenos Aires. The government gave up dollarization. It refused to meet the demands by the IMF and creditors to repay loans quickly. It funded an emergency unemployment benefits program and worked with Peronist unions to lessen social disorder. These institutions helped Argentina maintain its democracy and social stability. The country went on to a strong recovery based on an export-led boom that lowered unemployment and poverty. As with China's development of a labor market, there is no good counterfactual to assess how Argentina might have fared if it had weaker labor institutions or had chosen to suppress protesters and unions to repay its international debts. But during the crisis labor institutions helped maintain political and economic order.

9. THE INFORMAL SECTOR

The majority of workers in developing countries work in the informal sector.¹⁸ The traditional view has been that economic growth shrinks the informal sector and that as it does, more workers will gain the higher pay and economic security of the formal sector.

This underlies the unease that the 1990 World Bank Development Report expressed toward labor institutions: if only unions and governments leave well enough alone, the natural process will move workers out of the low productivity informal sector into good formal sector jobs.

The employment data summarized in Table 4 contravene the picture of a naturally declining informal sector for the 1980s–2000s. In this period of time the informal sector expanded its share of global employment, even in developing countries with rapid economic growth. Line 1 indicates that the proportion of workers in self-employment increased in every region in the 1980s and 1990s.¹⁹ Line 2 shows that the nonpaid share of employment (the self-employed, employers, members of cooperatives, contributing family members, and nonclassified workers) increased. Line 3, which gives estimates of informal sector employment in eight of the ten most populous developing countries from a diverse set of country sources, tells a similar story. In seven of the countries the informal sector share of employment rose, while it held stable in Indonesia. Line 4 shows that even in Korea, arguably the world's greatest growth success, the informal sector remains a large and growing share of the work force.

The shocker in Table 4 is that informal sector employment rose in rapidly growing countries such as China, India, the Philippines, and Korea as well as in countries with anemic growth records. Why did rapid growth fail to shrink the informal sector? One likely reason is that productivity growth in the modern sector was so great that even massive expansion of output did not increase employment. China is instructive in this respect. Despite becoming the world's manufacturing center in the 2000s, China had fewer workers in manufacturing in 2002 than in 1987 (Bannister, 2005b, table 1). Presumably this is because it adopted modern labor-saving technology, though it could also reflect Chinese firms outsourcing manufacturing jobs to the informal sector.²⁰

Traditional models of developing country labor markets treat the informal sector as a last choice safety net where individuals seek employment when nothing is available in the formal sector. But this is not the full story. Some workers and firms may prefer the informal sector to avoid taxes and regulations and may make a relatively good living in those sectors. Studies of sectoral mobility for Mexico (Bosch & Maloney, 2007), Brazil (Bosch, Goni, & Maloney, 2007), and Chile (Packard, 2007) show that formal/informal boundaries are porous, with many workers shifting from one sector to the other in response to economic changes. The dynamics of mobility in the sectors may, however, differ. Hoek (2007) finds that in the Brazilian formal sector, reductions in employment take the form of a reduced job finding rate while in the informal sector, reductions in employment take the form of higher separation rates. Ulyssea and Szerman (2007) find that more educated and older workers had longer job duration in the formal sector and shorter duration in the informal sector. They also show a rapid decline in the hazard rate for exiting the informal sector that implies long spells if workers do not leave within 6 months.

Table 4 Indicators of the trend in the informal sector share of employment

(1) Self-employment share of nonagricultural employment (ILO, 2002) 1980s-1990s		
World		26–32%
Africa		44–48%
Latin America		29–44%
Asia		26–32%
2. Unpaid workers as share of total employment (ILO, LaborSta, 2007), 1993–2002		
World		55–60%
Developing countries		
Advanced countries		
3. Selected country measures, by population of country ^a		
China	1990–2005	51–52.8%
India	1993/4–2004/5	92.7–94.1%
Indonesia	1990–2003	28.2–28.2%
Brazil, urban	1990–2003	40.6–44.6%
Pakistan, urban	1997/8–2001/2	64.6–66.5%
Bangladesh	2000s	65%
Nigeria, urban	1960s–2005	25–45%–60%
	1970s–1990s	50–65%
Mexico	1991–1998	61.2–63.6%
Philippines Vietnam	2000s	70%
4. Informal sector employment in Korea, by definition		
Self-employed	1990–2004	39.5–34.0%
Nonregular	2001–2005	26.8–36.6%
Temporary	2001–2005	16.6–29.4%
Nonstandard	1990–2004	45.8–48.8%

Source: Line 1: ILO, Women and Men in the Informal Economy: a statistical picture (ILO, 2002), table 2.5.

Line 2: Estimated from ILO, Laborstats (2007).

Line 3: China, India, Brazil from OECD, March 2007, table 1.1; Indonesia, Felipe and Hasan (2006), Pakistan, Gennari (2004); Nigeria a, Nwaka (2005); Nigeria b, Mexico, Martin (2000), table 4; Philippines, Venida; UN, ILO; <http://www.ilo.org/public/english/employment/skills/informal/who.htm> for Bangladesh.

Line 4: Korea, 1990–2004 OECD, Factbook 2006, Economic, Environmental, and Social Statistics, Self employment rates as percentage of total civilian employment 2001–2005, defined as nonregular; temporary, OECD, Korea, table 5.13; 1990–2004, Ki Seong, Park (2006), nonstandard employment, Figure 1.

^aChina, India, Brazil from OECD, 2007, table 1.A.3. For Indonesia, Wiebe reports 68% for 1993; <http://www.socialalert.org/pdf/IndonesieBROCH%5B1%5D.pdf> reports over half informal in 2000s. <http://indonesiaurbanstudies.blogspot.com/2007/05/urban-planning-and-informal-sector-in.html> reports 64%.

Comparisons of wages across sectors and activities show that in Mexico and Brazil employees in the informal sector earn less than employees in the formal sector but that the self-employed earn as much or more than formal sector workers (Bosch, Goni, & Maloney, 2007; Maloney, 1999). Still, workers laid off from formal sector manufacturing suffer earnings decline of nearly 50% 1 year later and about 33% 2 years later, with middle-aged workers hit hardest by a layoff (Hoek, 2006). For El Salvador and Peru, Marcouiller, de Castilla, and Woodruff (1997) find that informal sector workers earned significantly less than formal sector workers. In Argentina poverty among older persons is greater when a greater proportion of their working lives is spent in the informal sector.

Given the size and persistence of the informal sector in developing economies, we need more studies of how workers fare in this sector, of the links between formal and informal job markets, and of possible innovative social insurance schemes to improve the working lives of those in the informal sector, be it through social security pensions (Auerbach, Genoni, & Pagés, 2007), occupational health and safety laws, increased enforcement of regulations (Almeida & Carneiro, 2007), or perhaps through associations of the self-employed like SEWA (<http://www.sewa.org/>).

10. CONCLUSIONS

This review has found that regulations and unions are not the bugbear to development that many believed them to be years ago. Some labor policies have adverse effects on employment in some countries, but the magnitudes are generally modest. Evidence on the wage curve, the spillover of minimum wages to the informal sector, and the mobility of workers between the informal and the formal sectors shows that the two-sector model that has guided much thinking about labor markets in developing countries does not capture the way those markets operate. The evidence suggests instead that labor markets adjust to conditions in developing countries much as they do in advanced countries and do not impede macroeconomic adjustments. The Chinese example suggests that a functioning labor market can help development, while the Argentine example suggests that labor institutions can help preserve social stability during turbulent times.

The 800-pound gorilla in this review is the increased share of the informal sector in successful developing countries. Because research has focused largely on formal sector labor markets, we know far too little about the informal market setting in which most workers make their living, and about policies and institutions that can help raise productivity in the informal sector, improve occupational health and safety, and deliver social services and protections to them. The informal sector is going to be the locus of work for the majority of workers for the foreseeable future and should be the focus of labor market analyses as well.

APPENDIX A**Table A.1** Developing country labor laws and practices

Country	Fraser Institute (2004)	Employment law	Collective relations law	Union density	Co. Flexibility of Wage determination 7—employer vs 1 cb	Co. Control of Hiring and Firing Practices 7—employer 1—regulated
Uganda	8.6	0.35	0.38	3.9	6.3	5.0
Zambia	8.6	0.15	0.29	12.5	—	—
Pakistan	7.8	0.34	0.31	5.5	5.7	4.5
Haiti	7.3	na	na	—	—	—
Bangladesh	6.7	na	na	4.3	5.3	4.0
Ghana	6.2	0.29	0.48	25.9	3.7	4.4
Kenya	5.8	0.37	0.23	16.9	5	4.9
India	5.7	0.44	0.38	5.4	5.4	2.6
Malawi	5.7	0.18	0.25	—	5	3.5
Nigeria	5.5	0.19	0.21	17.2	4.6	4.5
Mongolia	5.4	0.33	0.23	—	5.9	4.6
Zimbabwe	5.2	0.25	0.44	13.9	2.8	2.2
Tanzania	5.1	0.68	0.32	17.4	4.5	3.7
Vietnam	4.8	0.54	0.48	—	5.4	4.2
Benin	4.1	na	na	—	5.3	4.1
Chad	3.7	na	na	—	5.4	3.6
Mali	3.6	na	na	13.7	4.5	3.4
Madagascar	3.1	0.47	0.46	—	5.3	3.6
Senegal	2.8	0.51	0.57	21.9	—	—
Mozambique	2.5	0.79	0.58	—	8.7	3.6
Guyana	8.2	na	na	25.2	5.1	4.5
Namibia	7.2	na	na	22.0	4.3	3.0
Thailand	7.0	0.41	0.36	3.1	4.7	4.2

Table A.1 Developing country labor laws and practices—Cont'd

Country	Fraser Institute (2004)	Employment law	Collective relations law	Union density	Co. Flexibility of Wage determination 7—employer vs 1 cb	Co. Control of Hiring and Firing Practices 7—employer 1—regulated
Nicaragua	6.8	na	na	23.4	5.5	3.7
Cameroon	6.6	na	na	14.7	4.8	3.6
Jamaica	6.6	0.16	0.23	—	5.5	4.0
Jordan	6.6	0.7	0.38	—	5.9	3.4
Georgia	6.5	0.77	0.57	—	6.0	4.9
Sri Lanka	6.4	0.47	0.51	—	4.5	2.7
Dominican Republic	6.0	0.6	0.27	17.3	5.1	3.9
Macedonia	5.8	na	na	—	5.5	3.3
Albania	5.7	na	na	—	5.7	4.8
El Salvador	5.5	na	na	7.2	5.5	4.9
Azerbaijan	5.4	na	na	—	5.2	4.7
Egypt	5.3	0.37	0.41	29.6	5.9	3.9
Armenia	5.2	0.6	0.52	—	5.6	4.9
Indonesia	5.0	0.68	0.39	2.6	4.2	3.6
Philippines	5.0	0.48	0.51	22.8	4.4	2.6
Honduras	4.9	na	na	4.5	4.3	3.1
Ukraine	4.8	0.66	0.58	—	5.0	4.8
Bolivia	4.8	0.37	0.46	16.4	5.4	3.1
China	4.7	0.43	0.33	54.7*	5.2	4.5
Colombia	4.3	0.34	0.49	7	5.5	3.8
Tunisia	4.3	0.82	0.38	9.8	3.7	4.2
Peru	4.3	0.46	0.71	7.5	5.7	3.3

Continued

Table A.1 Developing country labor laws and practices—Cont'd

Country	Fraser Institute (2004)	Employment law	Collective relations law	Union density	Co. Flexibility of Wage determination 7—employer vs 1 cb	Co. Control of Hiring and Firing Practices 7—employer 1—regulated
Guatemala	4.2	na	ns	4.4	4.4	3.6
Ecuador	4.2	0.4	0.64	9.8	4.5	3.0
Morocco	4.1	0.26	0.49	4.8	5.1	3.7
Algeria	3.7	na	na		3.9	3.4
Paraguay	2.9	na	na	9.3	3.8	2.7
Malaysia	7.9	0.19	0.19	11.7	5.6	4.0
Uruguay	7.3	0.28	0.35	11.6	5.0	3.7
Hungary	7.1	0.38	0.61	52.5	5.7	4.0
Mauritius	6.7	na	na	25.9	3.1	3.6
Slovak Republic	6.6	0.66	0.45	52.3	5.6	4.8
Botswana	6.5	na	na	11.5	5.1	3.7
Costa Rica	6.1	na	na	13.1	4.4	4.1
Bulgaria	6.0	0.52	0.44	51.4	5.6	4.5
Russia	5.8	0.83	0.58	74.8*	5.6	5.0
Latvia	5.8	0.72	0.53	—	5.7	3.9
Panama	5.6	0.62	0.46	14.2	4.0	3.0
Mexico	5.5	0.59	0.58	31	5.0	3.0
Croatia	5.5	0.49	0.45	—	5.1	3.8
Lithuania	5.4	0.62	0.5	—	5.8	3.1
Chile	5.3	0.47	0.38	15.9	6.1	3.4
South Africa	5.1	0.32	0.54	21.8	3.4	2.3
Argentina	5.1	0.34	0.58	25.4	4.0	2.7
Brazil	5.0	0.57	0.38	32.1	4.1	3.1

Table A.1 Developing country labor laws and practices—Cont'd

Country	Fraser Institute (2004)	Employment law	Collective relations law	Union density	Co. Flexibility of Wage determination 7—employer vs 1 cb	Co. Control of Hiring and Firing Practices 7—employer 1—regulated
Romania	5.0	0.33	0.57	40.7	5.6	3.1
Poland	4.8	0.64	0.57	27.0	5.0	3.5
Turkey	4.3	0.4	0.47	22.0	5.1	3.5
Venezuela	3.5	0.65	0.54	14.9	4.2	1.9
United Arab Emirates	8.9	na	na	—	6.3	5.0
Kuwait	8.6	na	na	—	6.1	4.5
Hong Kong	8.6	0.17	0.46	18.5	6.4	5.8
Bahrain	7.4	na	na	—	6.2	3.1
Trinidad & Tobago	7.1	na	na	—	4.9	4.0
Singapore	6.9	0.31	0.34	13.5	6.0	5.9
Malta	6.8	na	na	57.9	5.1	3.0
Estonia	6.1	na	na	26.4	6.2	4.4
Slovenia	5.6	0.74	0.49	—	4.6	2.9
Taiwan	5.3	0.45	0.32	27.9	6.2	4.6
Israel	4.8	0.29	0.31	23.1	5.1	4.4
Cyprus	3.1	na	na	53.7	4.2	3.3
Iceland	8.0	na	na	70.7	4.5	5.4
United States	7.9	0.22	0.26	12.7	6.1	5.4
Switzerland	7.4	0.45	0.42	20.0	5.8	5.4
Canada	7.1	0.26	0.2	31.0	5.6	4.3
Japan	7.0	0.16	0.63	18.6	5.9	3.6

Continued

Table A.1 Developing country labor laws and practices—Cont'd

Country	Fraser Institute (2004)	Employment law	Collective relations law	Union density	Co. Flexibility of Wage determination 7—employer vs 1 cb	Co. Control of Hiring and Firing Practices 7—employer 1—regulated
United Kingdom	6.9	0.28	0.19	26.2	5.9	4.4
Australia	6.6	0.35	0.37	28.6	4.6	3.5
New Zealand	6.5	0.16	0.25	23.2	5.8	3.8
Spain	6.2	0.74	0.59	11.4	4.8	2.9
Portugal	5.9	0.88	0.65	18.8	4.4	2.8
Ireland	5.8	0.34	0.46	36.0	3.2	3.4
France	5.7	0.74	0.67	6.1	4.6	2.4
Luxembourg	5.7	na	na	39.5	4.3	3.2
Italy	5.5	0.65	0.63	30.6	3.3	2.6
Netherlands	5.4	0.73	0.46	21.8	3.1	2.8
Denmark	5.1	0.57	0.42	68.2	4.2	5.4
Belgium	5.0	0.51	0.42	38.1	3.0	2.6
Czech Republic	5.0	0.52	0.34	36.3	5.6	3.5
South Korea	4.9	0.45	0.54	9.0	5.5	4.1
Austria	4.4	0.5	0.36	36.6	2.7	3.8
Norway	4.2	0.69	0.65	51.7	3.9	3.0
Greece	4.0	0.52	0.49	15.4	3.4	2.9
Finland	3.9	0.74	0.32	59.7	2.5	3.6
Sweden	3.5	0.74	0.54	77.2	3.3	2.4
Germany	3.3	0.7	0.61	29.6	2.8	2.3

Note: *are for government official unions, not for worker-formed independent unions.

na means data are "not available".

APPENDIX B: REGRESSION

Table B.1 Coefficients (standard errors) on regression of annual growth rates of GDP per capita on the labor subindex and the economic freedom index of the Fraser Institute (with the labor subindex removed), 1970–2004.

Lagged GDP per capita	-0.066 (0.007)
Labor subindex for developing countries	-0.002 (0.002)
Economic freedom index* for developing countries	0.010 (0.002)
Labor subindex for OECD countries	-0.001 (0.002)
Economic freedom index* for OECD countries	0.009 (0.003)
Year dummies	Yes
Country dummies	Yes
Number of observations	314
Number of countries	85

Sources for data used in the calculations: Indices of Economic Freedom, Fraser Institute (2006) Gwartney, James, and Robert Lawson (2006). *Economic Freedom of the World: 2006 Annual Report*. Vancouver: The Fraser Institute. Data retrieved from www.freetheworld.com.

<http://www.freetheworld.com/2006/2006Dataset.xls>. Labor subindex. Listed as 5B on the dataset.

Economic freedom index*—simple average of subindices for separate areas, excluding the labor subindex. Analysis using the full economic freedom index gives comparable results, but then has the labor index entered in two places.

GDP growth rates—change in natural log of real GDP per capita in const. prices (Laspeyres) from Penn World Tables divided by number of years (5 years except 4 years for 2000–2004).

Lagged GDP—log GDP per capita with 5-year lag (4 year in the case of 2000–2004).

End Notes

1. Let W be the wage in the urban sector and W_r be the wage in the rural sector. Then the two sectors have equal expected earnings when $eW = W_r$, where e is the ratio of employment (E) to labor force (L) in the urban sector. Since this means that $EW_r = LR$, $dL = W/W_r dE$. An increase in E increases the urban work force, which in turn reduces the rural output.
2. The 1995 World Development Review was the first major Bank statement in this regard: “Free trade unions are the cornerstone of any effective system of industrial relations. Unions act as agents for labor . . . monitor employers’ compliance with government regulations . . . can help raise workplace productivity and reduce workplace discrimination . . . (contribute to) . . . political and social development” (World Bank, 1995, p. 79). In 2003 the Inter-American Development Bank declared: “Labor regulations are not cost-free, but deregulation is not the answer . . . Unions are neither the sand in the wheels of the labor market nor the solution to low wages . . . better labor market performance is compatible with lower earnings inequality . . . The new agenda requires a strengthened labor

- authority and a complex network of public and private institutions" (Inter-American Development Bank, 2004, pp. 7–8).
3. Indicative of this thinking: "Rising trade volumes are unambiguously related to growth, but the direction of causation is unclear." Zagha, Nankini, and Gill (IMF, 2006); "some of the more extreme polemic claims made about the effects of financial globalization on developing countries, both pro and con, are far less easy to substantiate than either side generally cares to admit." Kose, Prasad, Rogoff, and Wei (IMF, 2007); "greater caution toward certain forms of foreign capital inflows might be warranted." Prasad, Rajan, and Subramian (IMF, 2007); "expectations about the impact of reforms on growth were unrealistic . . . our knowledge of economic growth is extremely incomplete. . . . an economic system may not always respond as predicted . . ." (Zagha et al., 2006). "The Washington Consensus has been dead for years" Wolfensohn (2004). On the role of government see World Bank (1993).
 4. In a general equilibrium closed economy the minimum would shift capital as well as labor to the informal sector, which could raise wages there if the shift raised its capital/labor ratio.
 5. Note that the inverse relation indicates that the adjustment process is insufficiently powerful to clear the labor market and eliminate the correlation in the time frame covered. There is debate about the applicability of the wage curve analysis to the United States, but US studies also find unemployment and wages positively related.
 6. See Babecky, Ramos, and Sanromá (2008) for a meta-analysis of wage curves that cover all countries that is consistent with this conclusion for developing countries.
 7. In 2004 the OECD summarized the literature as saying that "the evidence of the role played by EPL on aggregate employment and unemployment rates remains mixed" (p. 81).
 8. I find their direct analyses of employment and hiring more convincing than their interpretation of their estimated gap between wages and marginal productivity, which show a rise in the gap after the 1984 change in regulations but not after the 1991 change and a rise in the gap for blue collar workers in 1995 when the law did not change.
 9. Indicative of the political sensitivity of the analyses, the Indian Ministry of Finance misread the study as showing that "states which have more pro-worker regulations lost out on industrial production" when the finding is that more prolabor laws shifted production to the informal sector, conditional on omitting state time trends, as cited by Bhattacharjea (2006).
 10. He reports a 51% union wage effect using a selectivity term, but has no way to identify the likelihood of being union beyond functional form. Models of this type often yield extremely high or low estimates that reflect the absence of any genuine instrumental variable (Freeman & Medoff, 1982).
 11. Unions in Senegal and Cameroon have been in conflict with governments over time, and the unions in Zimbabwe have long opposed the Mobutu dictatorship.
 12. Using a regression model decomposition, he reports a 57% union wage effect for permanent workers. This appears to reflect the way the decomposition treats huge differences in regional and migration characteristics between union and nonunion workers.
 13. T.N. Srinivasan (National Academy of Sciences, 2004) worries about "data of varying quality from disparate economies . . . (subject to) . . . measurement errors and biases . . . (and) a two-way relationship between growth and some of the explanatory variables" (p. 14).
 14. Someone skeptical of the value of Bank assistance and advice might argue the opposite but the regressions show that growth rates tend to improve after receipt of Bank assistance.
 15. For instance, the estimated coefficient in Table A.1 on the interaction terms for the aggregate rigidity indicator on the years right before a program is -0.094 while the coefficient on the years right after the program is -0.097—a negligible difference of 0.003.
 16. These measures are negatively associated with growth rates only when industrial countries are included in the dataset (Forteza & Rama, p. 23).

17. By contrast, for advanced countries, the indices have negative impacts on growth regardless of the statistical methodology, which reflects the less rapid growth of institution-driven EU countries than of the more market driven United States and other English-speaking countries.
18. The ILO defines informal sector employment as self-employed; wage workers in insecure and unprotected jobs (unregistered, casual, temporary); household workers.
19. Data for particular countries not covered in the table confirm the trend. ILO (2001) estimates that the informal share of employment rose in 12 Latin American countries for which it obtained data. Scattered estimates show informal employment up in Ghana, Ethiopia, Kenya, and Nigeria.
20. Loayza, Oviedo, and Serven (2005) show that rapid growth reduces the informal sector share of output, while labor and product market regulations raise the informal sector share of GDP. There is no inconsistency between their regressions of the informal share of *output* on growth and the failure of observed growth to reduce informal sector shares of *employment*.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2007, January). *Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program*. <http://www.people.fas.harvard.edu/~jha1nm/Paper/ADH2007.pdf>
- Ahsan, A., & Pages, C. (2008, March). *Are all labor regulations equal? Evidence from Indian manufacturing*. IZA Discussion Paper No. 3394.
- Aidt, T., & Zafirov, T. (2003). *Unions and collective bargaining: Economic effects in a global environment*. World Bank.
- Alatas, V., & Cameron, L. (2003). *The impact of minimum wages on employment in a low income country: An evaluation using the difference-in-differences approach*. Working Paper No. 2985. World Bank Policy Research.
- Almeida, R., & Carneiro, P. (2007). *The efficiency-equity trade-off in labor market regulations: Evidence from Brazilian cities*. In 2nd IZA/World Bank Conference, Employment and Development June 8–9.
- Amadeo, E. J., & Camargo, J. M. (1997). Brazil: Regulation and flexibility in the labor market. In S. Edwards & N. C. Lustig (Eds.), *Labor markets in Latin America. Combining social protection with market flexibility* (pp. 201–234)). Washington, DC: Brookings Institution Press.
- Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61, 821–856.
- Arango, C., & Pachón, A. (2004, February). *Minimum wages in Colombia: Holding the middle with a bite on the poor*. Borradores de Economía Series No. 280. Banco de la República de Colombia. <http://www.banrep.gov.co/docum/ftp/borra280.pdf>
- Arbache, J. (1998, January). *The impact of unions on wages in Brazilian manufacturing*. <http://www.kent.ac.uk/economics/papers/papers-pdf/1998/9805.pdf>
- Auerbach, P., Genoni, M. E., & Pagés, C. (2007, August). *Social security coverage and the labor market in developing countries*. IZA Discussion Paper No. 2979.
- Babecsky, J. R., Ramos, R., & Sanromá, E. (2008). Meta-analysis on microeconomic wage flexibility (wage curve). *Sozialer Fortschritt*, 57(10–11), 273–279.
- Banister, J. (2005a, August). Manufacturing earnings and compensation in China. *Monthly Labor Review*, 128(8), 22–40.
- Bannister, J. (2005b, July). Manufacturing employment in China. *Monthly Labor Review*, 128(7), 11–29.
- Bell, L. (1997). The impact of minimum wages in Mexico and Columbia. *Journal of Labor Economics*, 15 (3, Pt. 2), S102–S135.
- Berg, J., & Dante, C. (2004, April). Political-economic regime and the wage curve: Evidence from Chile, 1957–1996. *International Review of Applied Economics*, 18(2), 151–165.
- Besley, T., & Burgess, R. (2004). Can labor regulation hinder economic performance? Evidence from India. *Quarterly Journal of Economics*, 19(1), 91–134.

- Bhalotra, S. (1993). *Geographical differences in unemployment and wage rates in India*. Wolfson College, Oxford University. Mimeo.
- Bhandari, A. (2008, October). *Union membership effect on wage premiums: Evidence from organized manufacturing industries in India*. IZA Discussion Paper No. 3747.
- Bhattacharjea, A. (2006, April-June). Labour market regulation and industrial performance in India: A critical review of the empirical evidence. *Indian Journal of Labour Economics*, 39(2).
- Blanchflower, D., & Oswald, A. (1994). *The wage curve*. Cambridge and London: The MIT Press.
- Blunch, N. H., & Verner, D. (2004). Asymmetries in the union wage premium in Ghana. *World Bank Economic Review*, 18(2), 237–252.
- Bosch, M. (2007). *Job creation and job destruction in the presence of informal labour markets*. IZA Working Paper.
- Bosch, M., Goni, E., & Maloney, W. (2007). *The determinants of rising informality in Brazil: Evidence from gross worker flows*. IZA Discussion Paper No. 2970.
- Bosch, M., & Manacorda, M. (2007, August 2). *Urban inequality and minimum wages in Mexico*. LSE. Mimeo.
- Bosch, M. M., & Maloney, W. (2007, June). *Gross worker flows in the presence of informal labor markets: Evidence from Mexico, 1987–2002*. IZA Discussion Paper No. 2864.
- Botero, J., Djankov, S., La Porta, R., López de Silanes, F., & Shleifer, A. (2004). The regulation of labor. *Quarterly Journal of Economics*, 119, 1339–1382.
- Bucheli, M., & Gonzalez, C. (2007). *An estimation of the wage curve for Uruguay UdeLaR June*. <http://www.bcu.gub.uy/autoriza/peiees/jor/2007/iees03j3400807.pdf>.
- Butcher, K., & Rouse, C. (2001). Wage effects of unions and industrial councils in South Africa. *Industrial and Labor Relations Review*, 54(2), 349–374.
- Calderon, C., & Chong, A. (2005). Are labor market regulations an obstacle for long-term growth? In J. Enrique Restrepo & A. Tokman (Eds.), *Labor markets and institutions*. Santiago, Chile: Central Bank of Chile.
- Calderon, C., Chong, A., & Leon, G. (2006, October). *Institutional enforcement, labor market rigidities, and economic performance*. IADB Working Paper 589.
- Calderon, C., Chong, A., Leon, G., & Valdes, R. (2005). Labor market regulations and income inequality: Evidence for a panel of countries. In J. E. Restrepo, & Tokman A. (Eds.), *Labor markets and institutions*. Santiago, Chile: Central Bank of Chile.
- Card, D., & Kreuger, A. (1995). *Myth and measurement: The new economics of the minimum wage*. Princeton, NJ: Princeton University Press.
- Cassoni, A., Allen, S., & Labadie, G. (2004). Unions and employment in Uruguay. In J. Heckman & C. Pages (Eds.), *Law and employment: Lessons from Latin America and the Caribbean*. Chicago, IL: University of Chicago Press.
- Cassoni, A., Labadie, G., & Fachola, G. (2005). The economic effects of unions in Latin America: Their impact on wages and economic performance of firms in Uruguay. In P. Kuhn & G. Marquez (Eds.), *What difference do unions make?*. Washington: Inter-American Development Bank.
- Castillo-Freeman, A., & Freeman, R. (1992). When the minimum wage really bites: The effect of the U.S.-level minimum on Puerto Rico. In G. J. Borjas & R. B. Freeman (Eds.), *Immigration and the work force: Economic consequences for the United States and source areas*. Chicago: University of Chicago Press for NBER.
- Castro Lugo, D. (2006). Curva salarial: una aplicación para el caso de México 1993–2002. *Estudios Económicos*, 21, 233–273.
- Cheon, B. Y. (2006, July 14). Suggestions for improving labor market flexicurity and active labor market policy in Korea. In *National employment strategy and its vision for job creation*. Seoul, Korea: KLI.
- Chor, D., & Freeman, R. (2005). *The 2004 global labor survey: Workplace institutions and practices around the world*. NBER Working Paper No. 11598.
- Easterly, W., Kremer, M., Pritchett, L., & Summers, L. H. (1993). Good policy or good luck? Country growth performance and temporary shocks. *Journal of Monetary Economics*, 32(3), 459–483.
- Eckholm, E. (2001, June 3). China's inner circle reveals big unrest. *New York Times*, Section 1, pg 14.
- Edwards, S. E., & Edwards, A. C. (2000). *Economic reforms and labor markets: Policy issues and lessons from Chile*. NBER Working Papers 7646. National Bureau of Economic Research, Inc.

- Fairris, D. (2003). Unions and wage inequality in Mexico. *Industrial and Labor Relations Review*, 56(3), 481–497.
- Fairris, D. (2005). *What do unions do in Mexico?* University of California Riverside. <http://www.economics.ucr.edu/people/fairris/mexunioneffects>.
- Fairris, D. (2006). Union voice effects in Mexico. *British Journal of Industrial Relations*, 44(4), 781–800.
- Fairris, D., A., Popli, G., & Zepeda, E. (2008, May). Minimum wages and wage structure in Mexico. *Review of Social Economy*, 66(2), 181–208.
- Fallon, P. R., & Lucas, R. E. B. (1989). *Job security regulations and the dynamic demand for industrial labor in India and Zimbabwe*. Institute for Economic Development Series, No. 2. Boston University.
- Fallon, P. R., & Lucas, R. E. B. (1991). The impact of job security regulations in India and Zimbabwe. *World Bank Economic Review*, 5(3), 395–423.
- Feldman, H. (2009). The quality of industrial relations and unemployment in developing countries. *Review of Development Economics*, 13(1), 56–69.
- Feliciano, Z. (1998). Does the minimum wage affect employment in Mexico? *Eastern Economic Journal*, 24 (2), 165–180.
- Felipe, J., & Hasan, R. (2006, August). *Labor markets in Asia: Issues and perspectives*. Hampshire, England: Palgrave Macmillan.
- Fields, G., & Yoo, G. (2000, June). Falling labor inequality in Korea's economic growth: Patterns and underlying causes. *Review of Income and Wealth*, 46(2), 139–159.
- Forteza, A., & Rama, M. (2001). Labour market rigidity and the success of economic reforms across more than one hundred countries. *Journal of Policy Reform*, 9(1), 75–105.
- Fraser Institute. (2001). *Economic freedom of the world: 2001 annual report*.
- Fraser Institute. (2003). *Economic freedom of the world: 2003 annual report*.
- Fraser Institute. (2006). *Economic freedom of the world: 2006 annual report*.
- Fraser Institute. (2007). *Economic freedom of the world: 2007 annual report*.
- Freeman, R. B., Boxall, P., & Haynes, P. (Eds.) (2007). *What workers say: Employee voice in the Anglo-American workplace*. Ithaca, NY: ILR Press an imprint of Cornell University Press.
- Freeman, R. B. (1993a). Labor market institutions and policies: Help or hindrance to economic development? *Proceedings of the World Bank Annual Conference on Development Economics 1992*, Washington, DC: The World Bank.
- Freeman, R. B. (1993b, May). Labor markets and institutions in economic development. *American Economic Review*, 83(2), 403–408.
- Freeman, R. B. (2002, January 4). Varieties of labor market institutions and economic performance. *Presented at the IRRA Session on Labor Market Institutions and Economic Outcomes*.
- Freeman, R. B. (2005, June). Labour market institutions without blinders: The debate over flexibility and labour market performance. *International Economic Journal* 19(2), 129–145. NBER Working Paper 11246. Korean International Economic Association.
- Freeman, R. B., & Lazear, E. P. (1995). An economic analysis of works councils. In J. Rogers & W. Streeck (Eds.), *Works councils: Consultation, representation, cooperation* (pp. 27–50). University of Chicago Press for NBER.
- Freeman, R. B., & Medoff, J. (1982). The impact of collective bargaining: Illusion or reality? In J. Stieber, R. B. McKersie & D. Q. Mills (Eds.), *U.S. industrial relations 1950–1980: A critical assessment*. Madison, WI: Industrial Relations Research Association.
- Freeman, R. B., & Medoff, J. (1984). *What do unions do?* New York: Basic Books.
- Freeman, R. B., & Oostendorp, R. (2002). Wages around the world: Pay across occupations and countries, with Remco Oostendorp, Chapter 1. In R. Freeman (Ed.), *Inequality around the world*. London, UK: Palgrave in association with the IEA, Conf Vol. #134.
- Galiani, S. (1999). *Wage determination in Argentina: An econometric analysis with methodology discussion*. Buenos Aires: University Torcuato Di Tella. Mimeo.
- Gennari, P. (2004). The estimation of employment and value added of informal sector in Pakistan. In *Proceedings of 7th Meeting of the Expert Group on Informal Sector Statistics*, New Delhi, February 2–4.
- Ghose, A. K. (2005). *Employment in China: Recent trends and future challenges*. Employment Strategy Papers 2005/14. Geneva: Employment Strategy Department, ILO.

- Gindling, T. H., & Terrell, K. (1995, August). The nature of minimum wages and their effectiveness as a wage floor in Costa Rica, 1976–1991. *World Development*, 23(8), 1439–1458.
- Gindling, T. H., & Terrell, K. (2005). The effect of minimum wages on actual wages in the formal and informal sectors of Costa Rica. *World Development*, 33(11), 1905–1921.
- Gindling, T. H., & Terrell, K. (2007a). The effects of multiple minimum wages throughout the labor market: The case of Costa Rica. *Labour Economics*, 12, 485–511.
- Gindling, T. H., & Terrell, K. (2007b). *Minimum wages and the welfare of workers in Honduras*. IZA Discussion Paper No. 2892.
- Grimshaw, D., & Miozzo, M. (2003). *Minimum wages and pay equity in Latin America*. Geneva: International Labor Organization.
- Gruber, J. (1997a). The incidence of mandated maternity benefits. *American Economic Review*, 84(3), 622–641.
- Gruber, J. (1997b). The incidence of payroll taxation: Evidence from Chile. *Journal of Labor Economics*, 15 (3), S72–S101.
- Gruber, J., & Kreuger, A. (1991). The incidence of mandated employer-provided insurance: Lessons from workers' compensation insurance. In D. Bradford (Ed.), *Tax policy and the economy* (Vol.5, p.111–144). Cambridge MA: MIT Press.
- Gwartney, J., & Lawson, R. (2006). *Economic freedom of the world: 2006 annual report*. Vancouver: The Fraser Institute. Data retrieved from www.freetheworld.com.
- Haltiwanger, J., Scarpetta, S., & Schweiger, H. (2008, April). *Assessing job flows across countries: The role of industry, firm size, and regulations*. NBER Working Paper No. 13920.
- Harris, J., & Todaro, M. (1970). Migration, unemployment and development: A twosector analysis. *American Economic Review*, 60(1), 126–142.
- Harrison, A., & Scorse, J. (2003). Globalization's impact on compliance with labor standards. *Brookings Trade Forum*, 45–95.
- Hasan, R., Mitra, D., & Ramaswamy, K. V. (2007, August). Trade reforms, labor regulations, and labor demand elasticities: Empirical evidence from India. *Review of Economics and Statistics*, 89(3), 466–481.
- Heckman, J., & Pages, C. (2002, June). *The cost of job security regulation: Evidence from Latin American labor markets*. NBER Working Paper No. W7773.
- Hoddinott, J. (1996). Wages and unemployment in urban African labour market. *The Economic Journal*, 106 (439), 1610–1626.
- Hoek, J. (2006, September). *Life cycle effects of job displacement in Brazil*. IZA Discussion Paper No. 2291.
- Hoek, J. (2007). *Labor flows in formal and informal labor markets in Brazil*. In *Proceedings of 2nd IZA/World Bank Conference, Employment and Development*, June 8–9.
- Hofmeyr, J., & Lucas, R. (2001). The rise in union wage premiums in South Africa. *Labour*, 15(4), 685–719(35).
- Horton, S., Kanbur, R., & Mazumdar, D. (1994). *Labor markets in an era of adjustment* (Vols. 1 and 3). Washington, DC: World Bank.
- Hussmans, R. (2004). *Defining and measuring informal employment*. <http://www.ilo.org/public/english/bureau/stat/download/papers/meas.pdf>.
- Ilkkaracan, I., & Raziye, S. (2003, September). The role of unemployment in wage determination: Further evidence on the wage curve from Turkey. *Applied Economics*, 35(14), 1589–1598.
- ILO. (1997). *World Labour Report 1997–1998: Industrial Relations, democracy and social stability*. Geneva: International Labour Office.
- ILO. (2002). *Key indicators of the labour market, 2001–2002*. Geneva: ILO.
- ILO.(2002). *Women and men in the informal sector: A statistical portrait*. <http://www.ilo.org/dyn/infoecon/docs/441/F596332090/women%20and%20men%20stat%20picture.pdf>.
- ILO. (2006). *Key indicators of the labour market, 2005–2006*. Geneva: ILO.
- ILO. (2007). *Laborstats*. Web site:<http://laborsta.ilo.org/>.
- ILO. LABORSTA Internet. <http://laborsta.ilo.org/STP/guest>.
- Inter-American Development Bank. (2004). *Good jobs wanted: Labor markets in Latin America*. Available at: <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=419963>.
- Jeong, J. (2003). Is Korean enterprise bargaining a bona fide tool for employee representation? *Journal of Industrial Relations*, 45(1), 94–100.
- Jeong, J. (2007). *Industrial relations in Korea*. London: Routledge.

- Jeong, P. (1997, December). *The impact of minimum wage legislation in developing countries where coverage is incomplete*. Working Paper Series 98–2. Oxford: Centre for the Study of African Economies.
- Kingdon, G. G., & Knight, J. (2004, March). Unemployment in South Africa: The nature of the beast. *World Development*, 32(3), 391–408.
- Kingdon, G., & Knight, J. (1998). *Wages and unemployment in South Africa: A spatial analysis*. Working Paper. Centre for the Study of African Economies, University of Oxford.
- Kingdon, G., & Knight, J. (2006). How flexible are wages in response to local unemployment in South Africa? *Industrial and Labor Relations Review*, 59, 471–495.
- Kingdon, G., Sandefur, J., & Teal, F. (2006, December). Labor market flexibility, wages and incomes in sub-Saharan Africa in the 1990s. *African Development Review*, 18(3), 392–427 (36).
- Kleiner, M., & Lee, Y. M. (1997, January). Wage councils and unionization: Lessons from South Korea. *Industrial Relations*, 36(1), 1–16.
- Knight, J. (1998). *What underlies the wage curve?* Working Paper. Institute of Economics and Statistics, University of Oxford.
- Knight, J., & Song, L. (2005). *Towards a labour market in China*. Oxford: Oxford University Press.
- Kose, M. A., Prasad, E., Rogoff, K., & Wei, S. (2007, March). Financial globalization: Beyond the blame game. *Finance & Development*, IMF, 44(1).
- Kristensen, N., & Cunningham, W. (2006, March). *Do minimum wages in Latin America and the Caribbean Matter? Evidence from 19 countries*. Working Paper 3870. World Bank Policy Research.
- Krueger, A. B., & Lindahl, M. (2001, December). Education for growth: Why and for whom? *Journal of Economic Literature*, XXXIX, 1101–1136.
- Kugler, A. (1999). The impact of firing cost on turnover and unemployment: Evidence from Columbian labor market reforms. *International Tax and Public Finance*, 6, 389–410.
- Kugler, A. (2002, February). *From severance pay to self-insurance: Effects of severance payments savings accounts in Colombia*. Discussion Paper No. 3197. Centre for Economic Policy Research.
- Kugler, A. (2004). The effect of job security regulations on labor market flexibility: Evidence from the Colombian labor market reform. In J. J. Heckman & C. Pages (Eds.), *Law and employment: Lessons from Latin America and the Caribbean*. Chicago: University of Chicago Press.
- Kugler, A., & Kugler, M. (2003, September). *The labour market effects of payroll taxes in a middle-income country: Evidence from Columbia*. Discussion Paper No. 4046. CEPR.
- Lachaud, J. P. (1998). *Efficiency wage, labour vulnerability and urban unemployment in Burkina Faso*. France: University Montesquieu-Bordeaux IV.
- Lee, Y. M., & Na, I. (2004). How did unions protect workers rights after the 1997 economic crisis in Korea? *Industrial Relations Research Association Proceedings of the 56th Annual Meeting*, 182–194.
- Lemos, S. (2004a, Fall). Minimum wage policy and employment effects: Evidence from Brazil. *Economia*, 219–266.
- Lemos, S. (2004b). *The effects of the minimum wage in the formal and informal sectors in Brazil—Bonn*. IZA Discussion Paper 1089. Institute for the Study of Labor. repec.iza.org.
- Lemos, S. (2007, May). Minimum wages across the public and private sectors in Brazil. *Journal of Development Studies*, 43(4), 700–720.
- Lewis, H. G. (1963). *Unionism and relative wages in the United States*. Chicago, IL: University of Chicago Press.
- Lewis, H. G. (1986). *Union relative wage effects: A survey*. Chicago: University of Chicago Press.
- Lipton, M. (1977). *Why poor people stay poor: Urban bias in world development*. Cambridge, MA: Harvard University Press.
- Loayza, N., Oviedo, A. M., & Serven, L. (2005, June). *The impact of regulation on growth and informality cross-country evidence*. Working Paper No. 3623. World Bank Policy Research.
- Lustig, N., & McCleod, D. (1997). Minimum wages and poverty in developing countries; some empirical evidence. In S. Edwards & N. Lustig (Eds.), *Labor markets in Latin America* (pp. 62–103). Washington, DC: Brookings Institution.
- Maloney, W. F. (1998a). *Are labor markets in developing countries dualistic?* Policy Research Working Paper 1941. Washington, DC: World Bank.
- Maloney, W. F. (1998b). *Self-employment and labor turnover in LDCs: Cross country evidence*. Working Paper 2102. Washington, DC: World Bank.

- Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *The World Bank Economic Review*, 13(2), 275–302.
- Maloney, W. F. (2003). Informal self-employment: Poverty trap or decent alternative? In G. S. Fields & G. Pfeffermann (Eds.), *Pathways out of poverty. Private firms and economic mobility in developing countries* (pp. 65–82). London: Kluwer.
- Maloney, W. F. (2004). Informality revisited. *World Development*, 32(7), 1159.
- Maloney, W. F., Nunez, J., Cunningham, W., Fleiss, N., Montenegro, C., Murrugarra, E., et al. (2001, April). *Measuring the impact of minimum wages: Evidence from Latin America*. Working Paper 2597. World Bank Policy Research.
- Maloney, W. F., & Nunez Mendez, J. (2004). Measuring the impact of minimum wages: Evidence from Latin America. In J. Heckman & C. Pagés (Eds.), *Law and employment: Lessons from Latin America and the Caribbean*. Cambridge, MA: NBER and University of Chicago.
- Maloney, W. F., & Ribeiro, E. P. (1999). *Efficiency wage and union effects in labor demand and wage structure in Mexico: An application of quantile analysis*. Policy Research Working Paper Series 2131. The World Bank.
- Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton, NJ: Princeton University Press.
- Marcouiller, D., de Castilla, V. R., & Woodruff, C. (1997, January). Formal measures of the informal-sector wage gap in Mexico, El Salvador and Peru. *Economic Development and Cultural Change*, 45(2), 367–392.
- Martin, G. (2000, November). Employment and unemployment in Mexico in the 1990s. *Monthly Labor Review*, 123(11), 3–18.
- Menezes-Filho, N., Zylberstajn, H., Chahad, J. P., & Pazello, E. (2005). Unions and the economic performance of Brazilian establishments. In P. Kuhn & G. Marquez (Eds.), *The economic effects of unions in Latin America*. Washington: Inter-American Development Bank.
- Meng, X. (2000). *Labor market reform in China*. Cambridge, England: Cambridge University Press.
- Micco, A., & Pages, C. (2006, November). *The economic effects of employment protection: Evidence from international industry-level data*. IZA Discussion Paper No. 2433.
- Michaud, P. C., & Vencatachellum, D. (2001, October 15). *The union wage premium for Blacks in South Africa*. http://www.commerce.uct.ac.za/research_units/dpru/Conf2001PDF/michaud&vencatachellum.pdf.
- Miller, V., & Vallee, L. (1995). Labour regulation and segmentation in Cameroon. In B. Guathier, (Ed.), *Economic reform and the development of the manufacturing sector in Cameroon, RPED country study series*. Washington, DC: World Bank.
- Moll, P. G. (1993, January). Black South African unions: Relative wage effects in international perspective. *Industrial and Labor Relations Review*, 46(2), 245–261.
- Montenegro, C., & Pages, C. (2004). Who benefits from labor market regulations? Chile, 1968–1988. In J. Heckman & C. Pages (Eds.), *Law and employment: Lessons from Latin America and the Caribbean*. NBER.
- National Research Council. (2003). Monitoring International Labor Standards: Quality of information, Summary of a Workshop. In Margaret Hilton (Ed.), *Division of Behavioral and Social Sciences and Education and Policy and Global Affairs Division*. Washington, DC: The National Academies Press.
- Neumark, D., & Wascher, W. (2006, November). *Wages and employment: A review of evidence from the new minimum wage research*. NBER WP 12663.
- Nwaka, G. (2005, May). The urban informal sector in Nigeria: Towards economic development, environmental health, and social harmony. *Global Urban Development Magazine*, 1(1)<http://www.globalurban.org/Issue1PIMag05/NWAKA%20article.htm>.
- OECD. (2004). *Employment outlook 2004*. Paris: OECD.
- OECD. (2006a). *Employment outlook 2004*. Paris: OECD.
- OECD. (2006b). *OECD factbook 2006*. Paris: OECD.
- OECD. (2007). *OECD employment outlook*. Paris: OECD.
- OECD. (2007). Working party on employment, labour markets in the BRICs. EDLSA/ESLA/WP5(2007).
- Ooghe, E., Schokkaert, E., & Flechet, J. (2003). The incidence of social security contributions: An empirical analysis. *Empirica*, 30, 81–106.

- Packard, T. (2007, May). Do workers in chile choose informal employment? A dynamic analysis of sector choice. Working Paper Series No. 4232.
- Panagides, A., & Patrinos, H. A. (1994). *Union/non-union wage differentials in the developing world: A case study of Mexico*. Policy Research Working Paper 1269. Policy Washington, DC: World Bank.
- Park, K. S. (2006, July 12). *Non-standard employment*. Korea Labor Institute, Beyond Flexibility.
- Park, Y. B. (1991, December). Union/nonunion wage differentials in the Korean manufacturing sector. *International Economic Journal*, 5(4), 79–91 Korean International Economic Association.
- Pencavel, J. (1997). The legal framework for collective bargaining in developing economies. In S. Edwards & N. Lustig (Eds.), *Labor markets in Latin America* (pp. 62–103). Washington, DC: Brookings Institution.
- Petrin, A., & Sivadasan, J. (2006, December). *Job security does affect economic efficiency: Theory, a new statistic, and evidence*. NBER Working Paper No. 12757.
- Popli, G. (2007). Rising wage inequality in Mexico, 1984–2000: A distributional analysis. *Journal of Income Distribution*, 16(2), 49–67.
- Rama, M. (2000). Wage misalignment in the CFA countries: Were labour market policies to blame? *Journal of African Economies*, 9(4), 475–511.
- Rama, M. (2001). The consequences of doubling the minimum wage: The case of Indonesia. *Industrial and Labor Relations Review*, 54(4), 864–881.
- Rama, M., & Artecona, R. (2002). A database of labor market indicators across countries. Washington: World Bank, Development Research Group.
- Rodgers, Y. V., & Nataraj, S. (1999, October). Labor market flexibility in East Asia; lessons from Taiwan. *Economic Development and Cultural Change*, 48(1), 51–69. Working Paper.
- Rodriguez, F., & Rodrik, D. (1999). *Trade policy and economic growth: A Skeptic's guide to cross-national evidence*. NBER Working Papers 7081. National Bureau of Economic Research.
- Saavedra, J., & Torero, M. (2005). Union density changes and union effects on firm performance in Peru. In P. Kuhn & G. Marquez (Eds.), *What Difference Do Unions Make?: Their Impact on Productivity and Wages in Latin America*. Washington: Inter-American Development Bank.
- Sabin, L. (1999, February). The development of urban labour markets: China's urban wage curve, 1980–92. *Journal of Development Studies*, 35(3), 134–152.
- Schultz, T. P., & Mwabu, G. (1998, July). Labor unions and the distribution of wages and employment in South Africa. *Industrial and Labor Relations Review*, 51(4), 680–703.
- SMERU Research Institute. (2001 October). *Wage and employment effects of minimum wage policy in the Indonesian urban labor market*. The SMERU Research Institute. www.smeru.or.id.
- Squire, L., & Suthiwart-Narueput, S. (1997). The impact of labor market regulations. *World Bank Economic Review*, 11, 119–143.
- Srinivasan, T. N. (2004). *Comments in National Research Council, monitoring International Labor Standards, techniques and sources of information*. Washington: National Academies Press.
- Standing, G. (1992). Do unions impede or accelerate structural adjustment? Industrial versus company unions in an industrializing labour market. *Cambridge Journal of Economics*, 16(3), 327–354.
- Strobl, E., & Frank, W. (2003). Minimum wages and compliance: The case of Trinidad and Tobago. *Economic Development and Cultural Change*, 51(2), 427–450.
- Summers, R. (1989, May). What can economics contribute to social policy? Some simple economics of mandated benefits. *The American Economic Review*, 79(2), 177–183. Papers and Proceedings of the Hundred and First Annual Meeting of the American Economic Association.
- Teal, F. (1996). The size and sources of economic rents in a developing country manufacturing labour market. *The Economic Journal*, 106, 963–976.
- Terrell, K., & Svejnar, J. (1989). *The industrial labor market and economic performance in Senegal*. Boulder, CO: Westview Press.
- Thomas, M., & Vallée, L. (1996). Labor market segmentation in Cameroonian manufacturing. *Journal of Development Studies*, 32(6), 876–898.
- Tsafack-Nanfoso, R. (2007). *Union wage differential: Cameroon*. <http://www.csae.ox.ac.uk/conferences/2002-UPaGiSSA/papers/Tsafack-Nanfoso-csae2002.pdf>.
- Ulyssea, G., & Szerman, D. (2007). *Job duration and the informal sector in Brazil*. In *Proceedings of 2nd IZA/World Bank Conference, Employment and Development*, June 8–9.

- UNIDO. (2006). *The international yearbook of industrial statistics 2006*. UK: Edward Elgar.
- United Nations. (2005). *United Nations Human Development Report 2005*.
- Urizar, C., & Lee, S. (2005). The effects of unions on productivity; evidence from large coffee producers in Guatemala. In P. Kuhn & G. Marquez (Eds.), *What difference do unions make?* Washington: Inter-American Development Bank.
- Vallee, L., & Thomas, M. (1994). Labor market segmentation and labor absorption in Cameroon. In B. Guathier (Ed.), *Manufacturing enterprises under adjustment in Cameroon: A survey perspective*. RPED Country Study Series. Washington, DC: World Bank.
- Venida, V. S. (1998). *Employment, productivity and the informal sector in the Philippines, 1974-1988: An input-output analysis*. <http://www.iioa.org/pdf/12th%20conf/nydrfr98.pdf>.
- Verner, D. (1999). *Are wages and productivity in Zimbabwe affected by human capital investment and international trade?* World Bank Policy Research Working Paper No. 2101.
- Walder, A. (1986). *Communist neotraditionalism: Work and authority in Chinese industry*. University of California Press.
- Wolfenson, J. (2004, May 25). *Comments on 'scaling up poverty reduction' in Shanghai*.
- World Bank. (1990). *World development report 1990*. New York: Oxford University Press.
- World Bank. (1993). *The east Asian miracle*. New York: Oxford University Press.
- World Bank. (1995). *World development report*. Washington, DC: The World Bank.
- World Bank. (2006a, January). *Minimum wages in Latin America and the Caribbean: Impact on employment, inequality, and poverty*. Office of the Chief Economist.
- World Bank. (2006b, July 10). *China—Pension Reform Project, Report No.: 36024*.
- World Economic Forum. (2005). *Global Competitiveness Report. 2005–2006*.
- World Economic Forum. (2006). *Global Competitiveness Report. 2006–2007*.
- Wu, Z. (2004, June). Wage curve for urban China: A panel data approach. *Applied Economics Letters*, 11(7), 425–428.
- Yamada, G. (1996). Urban informal employment and self-employment in developing countries: Theory and evidence. *Economic Development and Cultural Change*, 44, 289–314.
- Zagha, R., Nankani, G., & Gill, I. (2006, March). Rethinking growth. *Finance & Development, IMF*, 43(1).

Access to Finance

Dean Karlan

Yale University

Jonathan Morduch*

New York University

Contents

1. Introduction	4704
1.1 Mechanisms matter	4706
1.2 Testing what works	4707
2. Global Financial Access	4710
3. Financial Intermediation and Economic Growth	4712
4. Returns to Capital	4715
4.1 Framing the question	4717
4.2 Evidence from estimating profit functions	4717
4.3 Evidence from field experiments	4719
4.4 What do average returns tell us?	4720
4.5 Sensitivity to interest rates	4721
4.6 A final caveat	4722
5. Credit Market Innovations	4723
5.1 Nature of frictions and policy examples	4724
5.2 Interventions and mechanisms	4729
5.3 Impacts from solving credit market failures	4736
6. The Economics of Saving	4739
6.1 Basic models of saving	4742
6.2 Constraints to saving	4745
6.3 Impacts of saving	4757
7. Risk Management and Insurance	4759
7.1 Why insurance markets fail	4762
7.2 Partnership models and index-based insurance	4766
7.3 Health insurance	4769
8. Governments, Businesses, and Nonprofit Institutions	4770
9. Concluding Comments	4774
End Notes	4775
References	4777

Abstract

Expanding access to financial services holds the promise to help reduce poverty and spur economic development. But, as a practical matter, commercial banks have faced challenges expanding access to poor and low-income households in developing economies, and nonprofits have had limited reach. We review recent innovations that are improving the quantity and quality of financial access. They are taking possibilities well beyond early models centered on providing “microcredit” for small business investment. We focus on new credit mechanisms and devices that help households manage cash flows, save, and cope with risk. Our eye is on contract designs, product innovations, regulatory policy, and ultimately economic and social impacts. We relate the innovations and empirical evidence to theoretical ideas, drawing links in particular to new work in behavioral economics and to randomized evaluation methods.

JEL classifications: O16, O17, G21, G22, D03, D03, I32

Keywords

- microfinance
- microcredit
- microinsurance
- credit savings
- insurance
- behavioral economics
- financial intermediation
- economic growth
- randomized controlled trials

1. INTRODUCTION

Many interventions have been proposed to solve entrenched development problems, or at least to make noticeable dents in poverty levels. The list of accumulated hopes is long, including better nutrition to catapult levels of productivity and wages; control of population growth to free resources for human capital investment; education for girls to fight inequalities and bring empowerment; and stronger property rights to unleash markets. Each hope is grounded in good reason, and each intervention holds a place in the larger scheme of development strategies. But none on its own has proved to be a catalyst on the scale imagined by its chief proponents.

In recent years, much hope has been placed on the transformative power of financial access. It is, in many ways, the boldest claim so far, and the most unlikely. The best-known advocate has been Muhammad Yunus, the cowinner of the 2006 Nobel Peace Prize alongside Grameen Bank, the bank Yunus founded to serve the poor of Bangladesh. Yunus speaks eloquently and forcefully about the power of access to small loans—dubbed microcredit—to transform the businesses of poor households.¹ With

those loans, Yunus argues, incomes will grow and, with rising incomes, children will be given long-denied opportunities. As Yunus (2006) declared in his Nobel lecture in Oslo: “we are creating a completely new generation that will be well equipped to take their children out of the reach of poverty.”

Yunus’s argument has grounding in economic theory. The argument aligns with explorations of credit rationing that show that when lenders lack good information on customers and contracts are costly to enforce, outcomes are not necessarily Pareto efficient (Besley, 1994; Stiglitz & Weiss, 1981). Innovations in credit markets can thus, in principle, bring gains in both efficiency and equity. The common assumption that the marginal return to capital is large when capital is scarce reinforces the claim that the “unbanked” poor have sizeable returns to reap from financial access.

Yet, as a practical matter, commercial banks have had difficulty providing such access profitably. The unbanked (and underbanked) tend to be poor and often lack titled assets to offer as security for loans. Moreover, many of the unbanked want to make transactions at too small a scale to attract much interest from profit-seeking institutions (Cull, Demirgürç-Kunt, & Morduch, 2009b; Johnston & Morduch, 2008). The “microfinance revolution” has thus had to contend with incentive problems alongside more prosaic challenges imposed by transactions costs. More fundamentally, the list of other factors correlated with poverty is long (including low education levels, poor health, discrimination, and weak labor markets), and these challenges risk undermining the effectiveness of financial access in raising incomes. The evidence to date shows that access to capital may be powerful for some, but it does not yield high returns for all.

This way of thinking, centered on the productive potential of capital, requires scrutiny in part because the evidence so far is mixed. But, more importantly, the conceptual frame centered on providing small loans for small businesses is too limiting. While Yunus’s vision of unleashing the productive potential of millions of small-scale entrepreneurs yields a powerful narrative, it risks blinding policymakers and practitioners (and researchers) to the broader financial needs of poor and low-income households. Those broader financial needs are, in many ways, similar to the needs of richer households: mechanisms to manage cash flows, devices for accumulating assets in both the short term and long term, and tools for coping with risk. Access to capital to expand businesses can generate income that facilitates these tasks. But as Collins, Morduch, Rutherford, and Ruthven (2009) show through year-long “financial diaries” that track the financial lives of poor and “near poor” households in Bangladesh, India, and South Africa, financial activities are most often driven by a basic set of needs—for example, to get food on the table every day, deal with illness, pay school fees and other sizeable expenses, and seize investment opportunities as they arise. None of these needs is necessarily tied to running small businesses, and all are as important for employed people in cities as they are for village women running microenterprises.

Work on “access to finance” is shifting to embrace the idea of providing banking services (credit, savings, and insurance) rather than primarily delivering microcredit for small-scale business. This chapter describes ways of thinking about this transformation, with an eye on innovations that help expand and improve financial access in poor communities. We focus sharply on contract designs, product innovations, and regulatory policy, complementing earlier surveys on access to finance, including *Handbook of Development Economics* chapters written by Gersovitz (1988) and Besley (1995), which, to a far greater degree, focus on theories of banking, macroeconomic frameworks, and informal finance.² While most poor and low-income households continue to conduct most financial transactions through informal mechanisms, our focus on informal finance is largely instrumental. This is not to diminish the importance of the informal sector. But given the parameters of the chapter, we look to informal mechanisms mainly as guides for product design, contract possibilities, and context in understanding the measured impact of specific innovations.

1.1 Mechanisms matter

For all the unknowns, we have acquired one central understanding about financial behavior: that mechanisms matter. Old debates—about whether the poor can repay loans reliably, or whether they can pay high interest rates, or save, or insure—need to be recast. Yunus’s fundamental insight was to show that the poor are bankable if *the right lending mechanism is used*. The earliest mechanism to gain attention is the “group lending” contract, in which neighbors meet together to take loans and collectively assume responsibility for their repayment, mitigating problems imposed by information asymmetries and costly external contract enforcement. New evidence shows that this is only one of several key mechanisms, and probably not the most important (Armendariz de Aghion & Morduch, 2000, 2010; Giné & Karlan, 2009). But no matter their individual roles, when taken together the successes of the financial mechanisms have changed the terms of debates. Grameen Bank (like other microfinance institutions) reports loan repayment rates above 98% despite lending to poor households, most of whom lack collateral and experience with banks. Policymakers, “social” investors, and academic researchers have taken note.

Similar understandings of the importance of mechanisms have emerged with regard to saving, and, to an extent, insurance. For example, studies show that poor households often seek specific, structured financial tools to achieve their savings goals—not just generic savings accounts (Ashraf, Karlan, & Yin, 2006b; Collins et al., 2009; Duflo, Kremer, & Robinson, 2006). Old prejudices held that poor households lack the surpluses to save much (Bhaduri, 1973). The idea conformed to a notion of poverty defined as having income that falls below a minimal threshold necessary for basic subsistence. The logic holds that if you are struggling to meet your needs today, saving up will inevitably be a slow process. Yunus’s initial push for microcredit (rather than

“microsaving” or “microfinance”) thus made sense as a way to speed the process of transformation. But by the 1990s, the view that the very poor are unable to save has been turned back, prompted in particular by lessons on wide-scale saving in Indonesia (Patten & Rosengard, 1991). As Banerjee and Duflo (2007) show in their analysis of tens of thousands of households in thirteen developing countries, even the very poor, living on under \$1 a day per person, spend relatively heavily on what appear to be nonpressing expenditures (like social and religious expenditures). Against that background, it is easier to see room for saving within the budgets of the poor, and Grameen Bank itself has now introduced an array of flexible and structured saving products.

Insights follow as well from paying closer attention to the psychology of financial decision making (e.g., Thaler, 1990). Lessons from behavioral economics are naturally relevant for choices by poor and low-income households, and we see applications in research on pricing, saving, insurance, and debt traps. The new work shows the roles of limited self-control, loss aversion, and mental accounting and their implications for product design and marketing (Bertrand, Mullainathan, & Shafir, 2006). While we are far from having a grand, unified theory based on the psychology of financial choices, the evidence so far demonstrates gains from expanding beyond (but not away from) traditional economic intuition—in which the way that products and choices are presented to consumers is essentially immaterial. The empirical evidence is mounting that both product design features and presentation can matter greatly.

1.2 Testing what works

From a macroperspective, expanding financial access holds the promise of increasing economic growth by spurring investment in underfunded enterprise, following the logic of Gurley and Shaw (1955) and McKinnon (1973). On the savings side, expanding access to reliable, low-cost deposit accounts promises to increase the capital stock. Given that the expansion of access favors lower income populations, these steps also promise to reduce poverty and inequality. All this is true in principle, but there is little evidence so far that expanding financial access through microfinance has had notable macroimpacts anywhere. Only in a few countries—Bangladesh and perhaps Indonesia and Bolivia—is the scale of microfinance large enough to even imagine the possibility. We do know from cross-country evidence that financial deepening correlates with inequality reduction (Demirguc-Kunt & Levine, 2008), but the lack of scale means we have no firm results with regard to microfinance specifically, and endogeneity and sample size issues hamper causal inference in cross-country regressions.

Much of the action has thus been at the microlevel, and turning there requires a different set of lenses. The macroperspective puts a natural focus on savings primarily as a way to increase wealth and borrowing as a way to fuel investment. But for households, borrowing is also an important way to cope with emergencies and to pay for household and social expenses. To this extent, borrowing is welfare enhancing if not always

output-increasing. Saving too can be an important way to smooth consumption from month to month and to cope with within-year expenses, and not chiefly as a means to build up long-term balances (Rosenzweig, 2001). Again, saving may be welfare enhancing even if not particularly output-increasing. In their close look at the financial lives of poor households, for example, Collins et al. (2009) find a common pattern of intensive use of saving instruments but relatively small average balances. Turning to the microlevel also gives risk mitigation a prominent place in expanding financial access, and we review the growing movement to provide microinsurance. Without much formal or informal insurance, borrowing (whether at zero interest from neighbors and relatives or at high prices from moneylenders) becomes by default a primary way to cope with emergencies.

The evidence so far suggests that financial access will not, on its own, be enough to take children out of the reach of poverty on a massive scale. Nor does the evidence suggest that finance alone is necessarily as powerful as finance coupled with other interventions—like training and healthcare. But the most striking conclusion from the available evidence is in fact that many of the big questions are left unanswered. There have been few fully convincing studies of impacts, and little rigorous investigation of whether the very poor can benefit from financial access to the same degree as the less poor—or perhaps whether the very poor will benefit *more* than others. Either possibility is consistent with economic theory and is at root an empirical issue. Similarly, the knowledge of saving behaviors and risk management strategies of the poor is only now accumulating, as is our understanding of price sensitivity and the demand for particular qualities of service.

Establishing appropriate counterfactuals is a critical challenge for researchers. Convincingly teasing apart the roles of mechanisms and their impacts on customers has been slow-going, though progress is now being made through the adoption of approaches to evaluation that incorporate experimental elements, most importantly randomized controlled trials of various kinds. The new approaches draw on decades of experience with evaluations of medical treatments, and represent the best ways developed yet to address the selection biases and omitted variable biases that undermine the credibility of evaluations.³

The potential biases are particularly acute when assessing financial interventions. Microfinance customers tend to be especially entrepreneurial and energetic relative to their nonparticipating neighbors. This causes self-selection issues, which make non-experimental evaluations challenging—and which tend to bias estimates of impact to overstate actual benefits. Even if self-selection is not an issue, financial institutions are apt to carefully screen potential customers, filtering the pool to find the most promising customers, and seeking the most promising locations in which to operate. Again the biases tend to lead to overstating actual benefits if banks target as described earlier. Alternative selection processes can, correspondingly, lead to underestimates of impact.

Randomized controlled trials can eliminate the resulting selection biases by building evaluation methods into program design. Recent methodological innovations in experimental design aim to ensure that evaluations are cost-effective, ethically appealing, and useful for the programs and customers. But experimental approaches have limits (many of which are shared with nonexperimental modes of evaluation), and have only picked up steam in the past 5 years in their application to issues around financial access.

To form future policies wisely, randomized controlled trials should be pushed in two directions: first, researchers need to replicate studies in different settings. Learning that a given approach to microfinance worked in one place, with one institution, at one point in time is not sufficient to know what to do in the future. What works in Bangladesh may not work in Argentina. What works in the city may not work in a small town or a village. The problem of “external validity” (i.e., uncertainty around how far it is appropriate to generalize a particular study to other contexts) is an old but often ignored problem in applying empirical research to policy decisions. Replicating studies allows analysts to begin to address the problem of external validity by building a clear understanding of the necessary context for an intervention to work (i.e., robust tests of theories that account for how the context will influence the outcome). It may be that, on a macrolevel, certain interventions work best in boom economies, but not in low-growth scenarios. Or, on an individual level, certain interventions may, say, work best with women who have little-to-no preexisting power in the household, while having little impact on women who already exhibit considerable control within their families. Replication, combined with attention to theoretical relationships, can help us understand the underlying failures and the contexts in which innovations succeed.

The second direction involves learning *why* things work. We ask this on two levels. The first concerns market structure. In order to think about whether an idea has promise for solving a market failure (as is the claim of microfinance), it is necessary to understand why the market failed in the first place, and how this intervention was able to solve the specific failure. The second level is micro, regarding individual decision making. Here, we need to understand better the mechanics of choice, particularly for the choices faced by poor households. The information allows us to predict how the choices made by low-income households (and the outcome that can be achieved) will be affected by changing the available financial options and tools, including pricing. The aim is to learn information that is forward-looking, rather than confining efforts to only looking backward to assess the impacts of existing interventions.

Clearer data on impacts, market structure, and household-level (and often individual-level) decision making are critical for weighing major public policy issues—and are necessary complements to ethnographic, financial, and administrative data. The most volatile debates concern the appropriate use of subsidies and the setting of price regulations for financial institutions serving poor households. But a broader set of concerns

has received less systematic attention: whether investing in the sector as a whole is the most cost-effective way for donors to achieve their missions, relative to alternative interventions that reach poorer households, reach larger businesses, or that focus on interventions like health, education, and infrastructure. The microfinance movement has proved the possibility of creating viable economic institutions on a large scale, and the challenge now is to more carefully assess social and economic impacts.

The next section describes what we know about the gaps and accomplishments in providing financial access globally. Section 3 reviews the links between financial intermediation and economic growth, drawing on both theoretical and empirical work. Much of that hinges on assumptions about the returns on assets in small businesses, the topic of Section 4. Section 5 turns to credit market innovations, and Section 6 to savings. Section 7 then describes emerging work on risk management and insurance. Section 8 focuses on the policy landscape and the roles of governments, businesses, and nonprofit institutions. The final section draws conclusions.

2. GLOBAL FINANCIAL ACCESS

Gaps in financial access remain stark. Using survey data combined with aggregate indicators Demirgüç-Kunt, Beck, and Honohan (2007) report estimates of the share of populations with accounts in formal and semiformal (e.g., microfinance) financial institutions. More than 80% of households in most of Western Europe and North America have an account with a financial institution. In Central Asia and Eastern Europe 60–80% are estimated to have accounts, with Latin America exhibiting variation ranging from less than 20% in Nicaragua to more than 60% in Chile. Estimated usage in Asian countries generally ranges from 40% to 60%. A World Bank study in rural India, for example, finds that about 40% of households have deposit accounts, 20% have outstanding loans, and only 15% report having any insurance (Basu, 2006). In much of Sub-Saharan Africa, fewer than 20% have accounts. Only in Botswana, The Gambia, and South Africa are the estimates above 60% (Demirgüç-Kunt, Beck, & Honohan, 2007).

Taken together, the results suggest that the number of “unbanked” and “underbanked” adults worldwide could be 2–3 billion people: precise figures have not been aggregated. Against that backdrop, the rapid expansion of microfinance has been stunning but still leaves substantial gaps. The Microcredit Summit Campaign Report of 2007 reports a growth of 885% in the number of clients from 1997 to 2006—an average annual growth rate of 29% per year. In 2006, 3316 institutions reported to the organization, and those institutions reached 133 million clients; a year later, the number had swelled to 154 million. In 1997, only 618 institutions were found, cumulatively serving 13.5 million clients; remarkable growth considering that 93 million of the 133 million at the end of 2007 are judged to be among the “poorest,” an income segment that traditional banking institutions have long considered unbankable.⁴

Of the poorest customers microfinance counted in the survey, 90% are in Asia, mostly in Bangladesh and India (Daley-Harris, 2009). Overall, most microfinance customers are found in Bangladesh and India, with the next largest group in East Asia and the Pacific.

Still, even in Bangladesh, there are substantial gaps in financial access. The number of loans per 100,000 people in Bangladesh, for example, is 54.73 and the number of deposits for 1000 people is 228.75 placing it at the 31st and 43rd spot, respectively, in a World Bank survey of 53 developed and developing countries (Beck, Demirguc-Kunt, Peria, & Soledad, 2006). Without the spread of microfinance institutions, Bangladesh would have ranked considerably worse, but the numbers show that there is further to go in spreading access.

It is not just availability that matters. Fees, costs, and documentation requirements also serve to limit financial access. Beck et al. (2006) report on an important survey of the largest commercial banks in a large sample of countries, documenting price and nonprice barriers associated with deposit, credit, and payment services. The survey shows critical variations across countries in the degree of physical access to formal financial institutions, documents required to maintain accounts, and costs (e.g., minimum balance requirements and fees). In one dramatic example, they find that opening a checking account in a commercial bank in Cameroon required a minimum deposit of over \$700 (a figure greater than Cameroon's GDP *per capita*). In Sierra Leone, maintaining a checking account required annual fees exceeding 25% of Sierra Leone's GDP *per capita*. Getting a small business loan processed in Bangladesh, Pakistan, or the Philippines can take over a month. And transferring \$250 dollars abroad cost \$50 in the Dominican Republic. These extreme examples are made more striking by the fact that other banks have managed to drop minimum balance requirements, cut annual account fees, speed up loan processing, and slash costs for sending remittances.

Microfinance has expanded in part due to the rise in foreign capital investment. Between 2004 and 2006, foreign capital investment in microfinance tripled to \$4 billion; by 2007, investment had reached \$5.4 billion. Institutional investors lending to microfinance institutions reached US\$550 million in 2006. The majority of the global capital flows go to about 30 countries in three regions, though: Latin America, Eastern Europe, and Central Asia. Africa and Asia receive only 6–7% of foreign investment (Forster & Reillie, 2008). The mismatch of capital flows and the locations with a greater prevalence of poverty is startling given the emphasis by microfinance leaders on poverty reduction, but investors have been wary about the perceived lack of management capacity and regulation that imposes hurdles.

Despite the capital flows from social and commercial investors, the greatest microfinance outreach at this juncture is not from commercial institutions but from public sector banks, nongovernmental organizations, and self-help groups (an Indian hybrid based on partnerships between NGOs and banks; Cull et al. (2009b). One report reviewed 2600 microfinance institutions to better understand the institutional landscape (Gonzalez &

Rosenberg, 2006). They found that nongovernmental organizations served 25% of the 94 million borrowers found in 2004, with self-help groups serving 29%. Commercially based microfinance banks and licensed “nonbank financial institutions” served only 17%, though the composition is shifting toward commercial players, pushed by the transformation of nongovernmental organizations into nonbank financial institutions.

These shifts are likely to affect the nature of services delivered to customers. Cull et al. (2009b) use data from the Microfinance Information Exchange (MIX) to analyze lending models and outreach of 346 leading microfinance institutions serving 18 million active clients in 2002–2004. They find that two-thirds of commercially oriented microfinance banks lend through individual methods (i.e., standard bilateral loan contracts), while three-quarters of nongovernment organizations used group-lending methods in the original spirit of the Grameen Bank. The latter tend to target poorer households and often use the groups for social support, while the individual lenders tend to target “upmarket” clients looking for larger loans.

This broad picture of financial access is starting to gain detail, but it remains too imprecise to guide local policy. Details that may seem trivial—how a survey question is phrased, for instance—turn out to strongly shape responses (Cull & Scott, 2009). More generally, large, one-time surveys tend to miss important information, partly because respondents hesitate to disclose intimate information about their financial lives to outsiders, especially about informal activities. Such discrepancies are revealed by the collection of “financial diaries.” In an intensive data collection effort, Collins et al. (2009, figure A1.1) collected information on all household financial inflows and outflows for small samples in Bangladesh, India, and South Africa, repeating the interviews every 2 weeks for a year. The initial interviews in South Africa greatly undercounted inflows, a deficit that was narrowed to within 6% only after about six meetings (i.e., 3 months of repeated interviews). Much of what was undercounted was informal. Savings clubs, reciprocal credit arrangements with friends and family, and other informal financial mechanisms turned out to be abundant, but seldom picked up by large one-time surveys of the sort collected by government agencies and research organizations. The 42 households in the Bangladesh sample reported using 33 different devices, with no household using fewer than 10, while two-thirds percent of South African diarists belonged to at least one informal savings club. In all three countries, informal mechanisms were used more frequently by the poor than any other kind to form lump sums of money, even in the South African sample where many respondents held bank accounts.

3. FINANCIAL INTERMEDIATION AND ECONOMIC GROWTH

Economists have long linked the expansion of financial markets to the spread of broader economic activity. By the same token, economists have focused on ways that barriers to financial markets undermine economic efficiency. In the 1970s, economists

turned their focus on regulations in many countries that capped interest rates on loans. Interest rates serve many roles, and one is to screen the quality of investments. When interest rates are set artificially low, borrowers are undeterred in investing in businesses that have relatively low returns. Artificially low interest rates also lead to excess demand for credit—and thus, inevitably, to credit rationing. Goldsmith (1969) stitched together these pieces of analysis to argue that interest rate caps undermine the average quality of investment, yielding “financial repression.” The notion of financial repression was extended by McKinnon (1973) and Shaw (1973) who turned to savings, focusing on the ways that interest rate caps ultimately reduced returns on saving as well, ultimately reducing both the quality and the quantity of investment. The McKinnon-Shaw treatises drove broad arguments for financial liberalization (a push, notably, to allow interest rates to rise to levels determined by markets), and their ideas fueled a specific assault on rural credit “directed lending” programs, led by researchers associated with the Ohio State University Rural Finance Program (e.g., Adams et al., 1984).

The association of financial expansion and economic growth is well established in the empirical literature. The causal link is harder to establish, however, since economic growth spurs financial expansion just as financial expansion can spur growth. Levine (2005) reviews the basic empirical associations, arguing that the link from finance to economic growth cannot be explained merely by reverse causation (drawing on cross-country regression analyses including those by Rajan and Zingales (1998) and Beck, Levine, and Loayza (2000)). These empirical findings are based on data aggregated at a country level. The empirical linkages cannot be tied to the expansion of financial access by households (as opposed to firms), nor to the spread of microfinance. At this date, the penetration of microfinance is too low in most countries to draw reasonable inferences about broad economic impacts (Honohan, 2008). Indeed, the challenge at this point is to establish basic household-level impacts of microfinance.

A related strand of cross-country literature, though, turns to the distributional impacts of financial expansion. It does not ask about the impact of financial access by the poor on macroindicators, but instead asks about the impacts of financial deepening on poverty and inequality. The impacts are set out in the theoretical model of Loury (1981), for example. The focus of his model is the intergenerational transmission of inequality; parents’ inability to borrow to fund investment in their children’s human capital means that inequality of resources in a given generation translates into inequality in the next generation. In Loury’s model, redistribution can thus improve economic efficiency. While it is not a stress of the paper, relaxing borrowing constraints will also improve efficiency, as well as reduce inequality and its persistence over time. The basic result—that borrowing constraints reduce efficiency and exacerbate inequality by diverting capital from low-income households with high-return investments—emerges in a string of more recent theoretical papers, including Galor and Zeira (1993), Aghion and Bolton (1997), and Banerjee and Newman (1993). Greenwood and Jovanovic

(1990) build a model in which financial development can increase inequality as better-off households are, at first, best positioned to take advantage of finance. The logic follows from Townsend (1978, 1983) who builds off the idea that investment in creating financial systems is costly. In Greenwood and Jovanovic (1990), richer segments of the population thus invest in financial infrastructure first; over time a broader swath of the economy benefits, so that inequality widens then narrows with financial development. In Greenwood and Jovanovic (1990), though, financial deepening is poverty reducing at all points.

As with the empirical literature on economic growth, directions of causality are difficult to establish. Demirgüç-Kunt and Levine (2008, p. 1) are left to note that economic researchers have done “an inadequate job of examining how formal financial systems affect the poor. We find this surprising because many of the profession’s most influential theories on intergenerational income dynamics advertise the central role of financial market imperfections in shaping the economic opportunities of the poor.” Clarke, Xu, and Zou (2006) and Beck, Demirgüç-Kunt, and Levine (2007) are among the few papers to investigate the link across countries. Both focus on the role of private credit on measures of inequality. The private credit variable captures the value of credit offered by financial intermediaries (excluding the central bank and state-owned development banks) to the private sector as a fraction of GDP. Clarke et al. find that financial development is associated with inequality reduction in a dataset for 83 countries in 1960–1995, in line with Galor and Zeira (1993) and Banerjee and Newman (1993). The result is robust to instrumental variables estimation using the origin of the country’s legal system as a determinant of the degree of financial development—under the assumption that historical origins play no current role in explaining outcomes once contemporaneous variables are included in specifications.

Beck et al. (2007) provide similar results, extending the analysis by adding countries and years, taking the number of observations from 170 to 245. They rely on the timing of trends to make causal claims, given a dearth of alternative, credible instrumental variables. Their main conclusion is that financial development “disproportionately boosts incomes of the poorest quintile” and thus reduces income inequality. Financial development is associated with a reduction in the population share living on less than \$1 a day as well. Most of the long-run gain made by the poorest fifth (60%) comes from general growth effects and the balance (40%) results from reductions in income inequality. The broad conclusion is that financial development is good for the poor—though, here, the link occurs mainly through trickle-down effects.

One of the reasons that so little work has been done to tackle these kinds of links in cross-country data rests with the lack so far of breakthrough empirical approaches to solve statistical identification problems. In contrast, there has been a great growth of microstudies that attempt to link financial access to household well-being and decision making. The microstudies have the advantage of isolating the impacts of particular

kinds of financial intermediation (rather than focusing on financial development broadly measured in an economy). The application of general equilibrium models calibrated to specific economies also holds promise as a way to integrate micro- and macroanalyses (see, for example, Townsend & Ueda, 2006 calibration of a model with fixed financial costs to Thai data, 1976–1996). The microstudies, in addition, hold the promise of evaluating specific assumptions underlying theoretical models, such as the nonconvex production technologies that undergird models like those of Galor and Zeira (1993) and Banerjee and Newman (1993)—a research program outlined by Banerjee and Duflo (2005).

4. RETURNS TO CAPITAL

If there is one fundamental argument in the global microfinance movement, it boils down to beliefs about patterns of returns to capital. On one hand is the belief that poor households can earn higher returns than richer households. The idea stems from the assumption that poorer households are more likely to face binding financing constraints and thus will get an especially big boost in productivity from access to finance. The other side argues that this logic holds only to a point: the very poorest households likely lack the wherewithal to be reliable bank customers and are better off being served by other economic and social interventions (like education and health services that build human capital). We argue later that the terms of that frame are too stark and that generalizations based on income level alone conceal as much as they reveal. All the poor are not alike. More interesting questions surround (1) How to identify nonincome dimensions along which patterns of returns can be differentiated and (2) how to identify *other* interventions (e.g., financial literacy, skills training, marketing, health) that may raise returns to capital for low-income populations.

Theoretical models that yield credit constraints usually depict efficiency gains from expanded financial access; relaxing constraints means that the productive potential of entrepreneurs is unleashed—farmers who lack the cash to buy enough fertilizer at planting time, weavers who cannot buy sufficient yarn, shopkeepers who cannot adequately build their inventory (e.g., Banerjee & Newman, 1994). Microcredit advocates like Muhammad Yunus similarly focus on the gains from promoting “microenterprise.” Yunus argues that the returns to financial access are bound to be large—large enough in some cases to transform livelihoods and permit sustainable exit from poverty (Yunus, 2006). Even a small bit of extra cash, Yunus argues, can transform money-starved, microscale businesses.

The idea gives a place to start, though it ignores the observation that much credit is used for nonbusiness purposes. Still, the idea gives the simplest defense of the claim that poor households can afford the high interest rates often charged by microfinance

institutions. Rosenberg (2002) has put forward the claim most sharply in a much-cited publication of the Consultative Group to Assist the Poor. The implication is that poor entrepreneurs can afford high-priced credit (perhaps even better than some richer customers), and that poor entrepreneurs can and should pay the fees required to cover costs, be they 20% or 40% per year or possibly higher.

The case is backed with anecdotes. Take, for example, the story of Vidalia Mamami, a 43-year-old vegetable seller in Tacna, Peru. She sells vegetables from a stand in a local market, and her earnings help support her husband and five children. She had been in business for 21 years, but only recently turned to Pro Mujer, a micro-finance NGO:

With my first loan I was able to buy more merchandise for my business and I was able to add vegetables and condiments, which have increased my earnings. Before, I earned 18 to 20 soles per day by selling only fruit, but now that I have added vegetables and condiments, I earn an average of 30 to 35 soles per day. This money has allowed my family to eat better and allows me to do things for my children that I could never do before. I remember how my older children were not able to go to school because we didn't have enough money.⁵

The story puts together a dramatic increase in earnings and ties it to broader social impacts. For economists, it resonates in large part because it aligns with the well-understood model of production under imperfect credit markets. Still, the anecdotes tend to reflect the best cases, and the theoretical analyses assume that constraints bind. The theorist's job is to focus narrowly, abstracting from other variables that are apt to determine profitability in practice.

The belief that that many poor households are in fact relatively *weak* prospects for loans, and that they can take better advantage of other interventions (schools, health clinics, savings accounts, insurance, and the like), also makes sense. For those who argue from this side, it is unclear how many “unbanked” entrepreneurs have the skills, business connections, political access, and other inputs that can help in running a truly thriving enterprise; thus, their returns to capital may remain low. Second, households with more capital may be able to reap returns to scale unavailable to poorer households. Marguerite Robinson (2001) has, for example, drawn this conclusion in her sweeping assessment of the “microfinance revolution,” and it drives Dale Adams’ wariness of microfinance as a poverty-reduction tool (e.g., Adams & Pischke, 1992). So while microfinance advocates like Muhammad Yunus see credit as a human right (Yunus, 2006), others counter that poorer households may have such low returns that expanding credit access to the poorest might only create a heavy debt burden.

Thus, much of where one stands on ongoing microfinance policy controversies—Should credit be targeted to the poorest? Are there better interventions for donor dollars? Should interest rates be subsidized for the poorest? Is there a trade-off between

financial sustainability and depth of outreach?—is bound up with what one believes about patterns of returns to capital.

4.1 Framing the question

Though tempting, these are not questions that can be fully answered by simply looking at whether poor households *do* pay high interest rates. First, this kind of “market test” gives no sense of the *level* of gain that households experience. To see the point, consider the case in which microcredit is priced so that loans are only just worth taking. The interest rate, for example, might be 40% while the expected return to capital is 45%. The 5 percentage point gain is an important incremental gain (and will keep customers coming back for loans), but it is not a transformative change—and not the kind of gains asserted by Yunus. Second, households may be caught in debt traps, paying interest but falling deeper into a hole.

More important, the market test tells us whether *some* people can pay high interest rates, but it does not tell us anything about people who are not borrowing. Are they not borrowing because they cannot afford to? Or because they have no desire to (but could afford to if necessary)? Household surveys that look at a broad population are needed to see the bigger picture.

Johnston and Morduch (2008) show how this matters with evidence from a survey in which loan officers employed by Bank Rakyat Indonesia, a pioneering microfinance bank, were employed to assess the creditworthiness of a nationally representative sample in Indonesia (basing their judgments on expected returns to investments and on the stability and predictability of household cash flows). Households with incomes above the poverty line were deemed far more likely to be creditworthy than poor households. Still, the loan officers identified 38% of poor households as being ready and able to borrow from Bank Rakyat Indonesia with existing financial products. Johnston and Morduch (2008) conclude that the right question is not the one that has generated debate: Are the poor and very poor *as a group* creditworthy? Rather, the key question is: How many? And, most importantly, can the creditworthy portion be cost-effectively identified and served?

4.2 Evidence from estimating profit functions

Researchers measuring returns to capital run into the same difficulties that make impact evaluations so challenging. The biggest hurdle is to disentangle the pure return to capital (i.e., the improvement in profit that occurs relative to a situation where all else is the same, but the business owner has less capital) from the effect of qualities and conditions correlated with having capital. People with better access to capital tend also to have better access to other resources like labor and markets. They may also be more entrepreneurial, less risk averse, and higher skilled. So when we see that people with more capital have higher profits, it does not necessarily mean that having more capital caused the higher profits. The gains may be due to the other attributes.

Two approaches are taken to measuring returns to capital.⁶ The first approach uses econometric methods to estimate profit or output functions, and identifies returns to capital parametrically. Identification then turns on the extent of control variables for typical confounding variables like basic ability and entrepreneurial skill. Feder, Lau, Lin, and Luo (1990) provide an example. Their model of farm production in Jilin province in northeast China uses a switching regression (following Maddala, 1983) in which farm households are assumed to face a binding liquidity constraint (case 1) or to be unconstrained (case 2). The two cases are determined endogenously, modeled as a probit in which the dependent variable is an indicator of credit constraints. The statistically significant variables in the probit are last season's income and current savings levels, and, identification rests on the authors' assertion that neither directly affects output once capital is accounted for in the output equations. Feder et al. (1990) find reason to think that liquidity constraints bind: surveys yield that 41% of farmers with access to formal finance indicated that they would like to borrow even more, and 28% of non-borrowers wished to borrow but were denied access. But the estimates yield that one additional yuan of liquidity would yield only about one-quarter of one yuan of additional output. Feder et al. (1990) are left to conclude that constraints may not in fact bind so strongly in practice and that a fair amount of "production credit" is likely getting diverted to consumption purposes (about a third, they estimate). Their conclusion is thus relatively pessimistic about the general proposition that financial access will raise incomes in dramatic ways (although the welfare gain from consumption smoothing should not be ignored).

Newer work is more optimistic; indeed, some of the estimated returns to capital are puzzlingly high. In a study that closely follows from Feder et al. (1990), Guirkinger and Boucher (2007) use a switching regression to identify constrained and unconstrained farmers in Peru, yielding an estimate that implies that relaxing credit constraints would raise the value of output per hectare by 26%. The result is, of course, hypothetical, but suggests the possible gains in efficiency from expanding access to finance. In keeping with this result, Udry and Anagol (2006) also find high returns to capital in a sample of small-scale farmers in Ghana. Farmers growing nontraditional crops generated returns to capital of 250% per year on the median-sized plot. Farmers growing traditional crops generated returns of 50% per year on the median-sized plot.

In turning to small enterprise, rather than farm finance, McKenzie and Woodruff (2006), use data from the Mexican National Survey of Microenterprises (ENAMIN) and find marginal returns to capital in the range of 10–15% per month for the smallest firms—that is, those with capital stocks of less than US\$200. Each \$100 of extra investment raises earnings by \$10–15 per month, a handsome profit. Firms with capital stocks above \$500 have a more modest average marginal return to capital of 35%. These results are robust to a wide range of controls for ability and emerge using a semiparametric estimator that allows substantial freedom in the estimated pattern of returns.

The pattern leads McKenzie and Woodruff (2006) to reject the notion that production is characterized by important nonconvexities in production here, and thus they rule out technology-based poverty traps. Instead, like the case of Mrs. Vidalia Mamami, the microentrepreneur described, production can be expanded incrementally, as with her move to sell vegetables and condiments as part of her fruit-selling business. McKenzie and Woodruff also find little to suggest that the smallest businesses are particularly risky or newly established. The high returns thus seem to be bound up with capital constraints. The puzzle is that if returns are indeed so high for the poorest entrepreneurs, then why have they not saved their way out of those constraints (a point developed by Armendáriz & Morduch, 2010, Section 6.4)?

One possibility is that the McKenzie and Woodruff (2006) measures are overstated. Unmeasured ability might partly drive the results, a problem that panel data alone cannot fully remove (since changes in capital stocks over time would likely be affected by unmeasured factors like demand shocks; McKenzie and Woodruff, 2008).

This last problem is part of a larger challenge in understanding connections between informal labor and capital markets. At a basic level, the studies here do not account for the time that small-scale entrepreneurs put into their businesses: enterprise profits are generally measured without accommodation for the value of unpaid labor, though it is often the most important input into production. Without more complete data, we cannot determine the degree to which high returns to physical capital in fact reflect returns to both physical capital and unmeasured human capital. Accounting for unpaid labor is challenging given difficulties measuring the quality of labor inputs, and a first useful step would be to put a bound on the effects by reestimating enterprise profits under alternative assumptions about the value of own labor. Samphantharak and Townsend (2008) offer a well-structured framework for measuring enterprise profits that draws on accounting principles used by corporations; it yields clarity, for example, on how to treat income and expenses made in different periods. Ideally the framework would be extended to fully address the cost of labor.

4.3 Evidence from field experiments

A second approach uses experimental methods to generate exogenous variation in capital usage. The new work attempts to address econometric problems by creating interventions that distribute capital in poor communities based in part on a randomized process. In these interventions, some people get larger transfers, some smaller, depending on a decision formula that leaves an important part of the allocation to chance. de Mel, McKenzie, and Woodruff (2008b), for example, study 408 small firms in Sri Lanka and offer them a range of cash or in-kind prizes (the in-kind grants are either equipment or inventories, selected by the business owners). The prizes (worth either roughly \$100 or \$200) were large enough to make a difference to the businesses, all of which functioned with capital investments under about \$1000. The researchers

picked winners and losers using random numbers. The random element (which means that people get access to capital independent of whether they are more talented, more connected, etc.) provides a key to estimating the pure return to capital. The real returns to the capital infusions in Sri Lanka turned out to be about 60% per year—an impressive return, especially given that nominal interest rates on loans are 12–18% per year.

Experimental evidence from Mexico also shows high returns to capital, especially for smaller businesses. While the earlier ENAMIN-based study had showed returns to capital of about 15% per month, a follow-up experimental study in Leon, in the state of Guanajuato, yields average returns of 20–33% per month for small, male-owned retail businesses with no employees other than the owner (McKenzie and Woodruff, 2008). Businesses that are identified by their owners as being financially constrained, moreover, have estimated returns to capital of 70–79% per month—and these businesses are most likely to be run by poorer households. Assuming that financially constrained households do indeed tend to be poorer, the result suggests that poorer households have a greater ability to pay for capital than better-off households, and it makes interest rates of even 10% per month seem reasonable in this context (though the result says nothing about female-owned businesses, nor small-scale enterprises engaged in services or manufacturing).

4.4 What do average returns tell us?

The averages, though, cannot speak to what is perhaps the biggest policy debate within the microfinance community—is microcredit an effective tool for the very poor (or should the focus be on households with incomes only slightly below poverty lines and above them)? The de Mel, McKenzie, and Woodruff (2008a) and de Mel et al. (2008b) result yields that returns are higher for a group identified as being constrained than one that is not, but the mapping into levels of poverty has not been done. We still have little sharp evidence to adjudicate whether it is true, as Vijay Mahajan, the founder of BASIX in Hyderabad, India, has said in summing up the early academic literature, that most microfinance borrowers starting below the poverty line “end up with less incremental income after getting a microloan,” and that borrowing “seems to do more harm than good to the poorest” (cited by Tripathi, 2006). Or can microcredit be a powerful tool to help the very poor, as long argued by Muhammad Yunus and others?

One important—and surprising—result of disaggregating the evidence in Sri Lanka yields that the average impact when female-owned businesses got more capital was “not different from zero” (de Mel et al., 2008a, 2008b). The heterogeneity is remarkable: 59% have returns less than zero, 14% in the 0–5% per month range, and only 27% of women had measured returns to capital over 5% per month. (The higher returns were for poor women with high cognitive ability, as measured by a test of number recall.) Men do better on average, but about a fifth of male owners generate returns

below market interest rates. The finding poses a puzzle in the context of the strong focus on serving women by many microfinance institutions (Armendáriz & Morduch, 2010, chap. 7), and it calls for further investigation in other settings.

A piece of the puzzle is found in Emran, Morshed, and Stiglitz (2007). They point to the role of a missing (or imperfect) labor market, especially for women, as one explanation for the success of the microfinance movement. Where the labor market opportunities for women are undervalued, it is argued, possibilities for self-employment are particularly appealing. Under these assumptions, women are willing to pay high rates of interest to microfinance institutions given limited outside options, even when their returns to capital are relatively low. Their model also shows that if using extra capital requires expanding the scale of operations, the expansion could be far more difficult for women than for men as it would require a discontinuous jump in wages if expansion requires the business to move to hiring outside workers at market wages, becoming a net demander of labor.

4.5 Sensitivity to interest rates

Another vantage on the ability to pay for loans comes from a parallel set of studies on the sensitivity of loan demand to interest rates. Researchers investigate loan demand directly, rather than focusing on the nature of underlying profit functions. The existence of high returns to capital in poor communities, of course, undergirds arguments that poor households can pay high interest rates—rates that are high enough to allow microlenders to sustain themselves without donor help. A recent survey of about 350 leading microfinance institutions finds most institutions charge interest rates and fees clustered between roughly between 20% and 40% per year, after taking inflation into account (Cull et al., 2009b). To some observers, these rates are very high and deserve justification. One defense has been that given high returns to capital, households are happy to pay seemingly high rates of capital as long as it is reliably delivered and in sizeable volume (Rosenberg, 2002).

Dehejia, Montgomery, and Morduch (2009) investigate changes in loan demand when a microfinance institution in Dhaka raised interest rates on loans from 2% per month to 3% per month in one branch but not in a similar location. Using the comparison across branches before and after the price change, they find a substantial short-term reduction in demand (roughly a unitary elasticity), although responsiveness is moderated over the longer term. In Dhaka, prices clearly matter to customers. From an institutional viewpoint, the price hike raised revenues that allowed it to cover costs and stay afloat—and was thus seen as a clear benefit to the lender, even if some customers shifted their behaviors. The Dehejia et al. (2009) study opens an important set of questions, but relies on an assumption of comparability across branches to identify the elasticity.

Karlan and Zinman (2008a) take an experimental tack to answer the same basic question. In one case, they work with a consumer lender in South Africa who charges

very high interest rates for installment credit (charging nearly 12% per month as interest). The researchers measured clients' sensitivity to interest rates by mailing out over 50,000 credit offers to customers, with the letters offering interest rates that were selected at random. Borrowers turned out to be less sensitive to changes in price than expected (the elasticity is modestly negative), a finding consistent with the notion that the lenders' customers have limited outside options for access to finance—a natural finding given the high prices routinely demanded for this lenders' loans. While the lender is atypical (well off the charts in terms of interest rates when compared to typical microfinance interest rates), the methods and questions are essential.

In a second case, Karlan, Mullainathan, and Zinman (2009) work with a large for-profit bank in Mexico, Compartamos, and randomize interest rates at the community level in 80 geographic clusters (containing 138 branches) across the country. This allows for longer term effects, both in terms of the decision-making process for individuals as well as competitive responses. Half of the geographic clusters were randomly assigned to receive a 0.50 percentage point reduction in their monthly interest rate (which translates to about a 10 percentage point reduction in the annual percentage rate). Customers turned out to be strongly sensitive to interest rates: the price reductions led to more clients, both new and retained, as well as larger loan sizes. The demand response was high enough to generate higher gross revenue: increased number of loans and size of loans outweighed lost revenue from lowering the interest rate on those who would have borrowed anyhow. Costs increased with the higher volume of lending, but not as much as revenues, thus profits increased.

The question begs for replication, as the elasticity should depend on underlying fundamentals of the economy, competitive landscape, financial literacy, disclosure policies, and population. Thus as with many empirical questions, more work needs to be done in a broader range of settings.

4.6 A final caveat

There is a final caveat to bear in mind. As noted in the context of Johnston and Morduch (2008), Bank Rakyat Indonesia loan officers focus in part on whether households have the cash flow (taking into account all of a household's income, from small business and from employment) to repay loans reliably. The loan officers are not focused exclusively on the ability to repay microcredit loans only from small business profits. Poor households have a wide variety of financial needs that go beyond enterprise—for example, financing healthcare, paying for school fees, and facilitating purchases of consumer goods. Not all poor households even run small businesses, especially in urban areas, and the poorest members of villages are often landless agricultural laborers. Thus, getting business loans is not always a priority for poor households, and the focus on business loans risks blinding policymakers and practitioners to a wider range of opportunities (Collins et al., 2009; Johnston & Morduch, 2008).

Answering narrow questions around the sizes of returns to capital in enterprise is necessary for assessing the degree to which the available evidence supports basic claims—notably that when borrowers do fund small businesses, the profits are sufficient to justify the interest rates charged by microlenders. However, given the plethora of nonenterprise needs for access to credit, it is worth noting that generating returns to capital in enterprise is only one part of what determines the demand for loans.

5. CREDIT MARKET INNOVATIONS

Theorists and practitioners often assume that problems, such as information asymmetries and difficulties enforcing legal contracts, undermine credit markets in developing countries. Yet in practice, deep methodological challenges have frustrated social scientists attempting to study the existence, extent, and nature of such problems. This is not to say there is evidence *against* the existence of market failures. But the mere fact that a microentrepreneur does not use credit is not sufficient cause to argue that a credit market failure exists. Several steps must be shown empirically in order to identify a market failure, form well-reasoned policy prescriptions, and finally to assess the welfare implications of resolving the market failure.

The simplest evidence of a credit market failure comes from evaluation of interventions which show increased use of credit after some shift in supply. This has been shown both for microenterprises (Banerjee, Duflo, Glennerster, & Kinnan, 2009 in India and Karlan & Zinman, 2009a in the Philippines), consumers (Karlan & Zinman, 2009b in South Africa) and small and medium enterprises (Banerjee & Duflo, 2008 in India). Had these studies showed instead a substitution from one source of credit (presumably more expensive on some margin) to another, the evidence of credit constraints *per se* would not have been as clear.

However, noting the mere presence of credit constraints is not sufficient for policy. The following section attempts to illuminate four essential questions regarding a presumed credit market failure: (1) What is the exact nature of the presumed market failure? (2) Is there a particular policy prescription that would mitigate a particular type of information asymmetry? (3) Does this particular policy solve, or diminish, the market failure? And (4) does it lead to a welfare improvement?

The previous section addressed the question of whether or not the lack of credit we observe in developing countries is economically efficient. Specifically we explored whether entrepreneurs are borrowing less than optimally because of a lack of profitable opportunities or whether, instead, there are other constraints at play which limit demand for credit. We presented evidence from experiments in developing country settings to suggest that returns to capital are in fact often high on average, but that heterogeneity in returns, due to different levels of social connections and human capital,

may lead to low demand by some for entrepreneurial credit. An important area for further research is to understand the nature of heterogeneity in returns and the relevance of such heterogeneity for assessing the extent of market failure.

The interaction between informal and formal markets will recur as a theme in the discussion later. Merely observing a formal market failure need not lead to inefficiency if the informal market is complete. Examining the mechanics of the informal market is crucial for two reasons. First, the strength of the informal market is important for measuring and predicting how effective specific formal sector interventions will be. Second, lessons learned in the informal markets can help shape policy in the formal markets. Group lending, for example, is based largely on lessons learned from observing risk-sharing and credit and savings associations in informal markets.

A second important theme is the application of the theoretical literature on information asymmetries to consumption loan markets. While theories on information asymmetries are often applied to discussions of entrepreneurial credit markets, the same theories are relevant to consumer credit as well for several reasons. As a start, the line between entrepreneurial “investment” and consumption “smoothing” is rarely evident for small, closely held businesses. Money is fungible. Empirical evidence from Johnston and Morduch (2008, Table 6) shows how even entrepreneurs report using loan proceeds to pay for consumption. In data from Indonesia, clients of BRI reported that about half of their loans were being used for nonentrepreneurial activities. More generally, asymmetric information problems as applied to risky “projects” have natural and close analogs for consumption loan borrowers. Just as entrepreneurs may respond to interest rates according to unobservable fixed characteristics of the return structure of their “project,” so too may the elasticity of demand depend on unobservable fixed personality traits (e.g., trustworthiness) or behaviors (e.g., the probability of incurring bad shocks). These unobservables may have significant impact on the likelihood of repayment. Similarly, if entrepreneurs unobservably change effort levels or repayment choices in response to interest rates, individuals may change their effort in activities such as maintaining wage employment or securing alternative sources of cash in the event of a bad shock. And, of course, individuals may also default strategically.

5.1 Nature of frictions and policy examples

Understanding how and why markets fail to provide credit access for the poor is critical to designing and replicating innovations that improve access. For example, if the problem is adverse selection, then policies that help reveal hidden information should help lenders screen and tie prices appropriately to risks. Such policies can be at the product or process level within the firm, or at the regulatory level, with credit bureaus. For example, Ghosh and Ray (2001) and Drugov and Macchiavello (2008) show theoretically how small, initial “tester” loans can provide information to lenders useful for assessing risk on subsequent, larger loans. Karlan, Mobius, Rosenblat, and Szeidl (2008)

show in a field experiment how social networks can be used to generate referrals of “good” clients, improving loan repayment.

Identifying the specific nature of a market failure is difficult, however. Take, as an example, information asymmetries generated through the interest rates on credit. Even if a lender were to randomize its interest rates across borrowers, merely observing that high rates lead to higher default does not disentangle selection from moral hazard and income effects. In a study of consumer loans in South Africa, individuals were marketed a randomly assigned interest rate and then, following the loan take-up decision, the lender randomly assigned a contract interest rate (lower than the initially offered rate), and a dynamic incentive (Karlan & Zinman, 2009b). This double-tiered randomization (before and after selection into credit) provided for a separation of adverse selection from moral hazard,⁷ and perhaps most importantly, provided a roadmap for using experimentation to test theories in which selection is useful to separate from *ex post* incentives. Karlan and Zinman found weak evidence for adverse selection, only statistically significant for females, and stronger evidence for moral hazard, in particular the dynamic incentive (lower interest rate on future loans).

In some cases, the hidden information is not hidden *per se*, but rather just not used. Interventions that work with lenders to improve their data processing can also improve efficiency. This has been shown in both insurance markets (Finkelstein & Poterba, 2002) and credit markets (Karlan & Zinman, 2009b). Credit bureaus also provide institutional and market level mechanisms for revealing hidden information. Theoretically, it is also the case that introduction of credit bureaus can mitigate moral hazard, as the consequences of default increase if lenders share information with each other (de Janvry, McIntosh, & Sadoulet, 2008). On the other hand, public credit bureaus can have adverse effects, leading to coordination failures (Hertzberg, Liberti, & Paravisini, 2008). We will discuss these in details below.

If the problem is moral hazard, on the other hand, solutions may differ. Again, one could categorize solutions at the business or product level, and separately at the institutional or market level. At the firm level, policies include contracts with dynamic incentives, tighter enforcement, or even business or financial training to improve the outcomes of the micro-enterprises or cash management of the borrowers. Note that *ex ante* moral hazard, that is, choice in effort, is often construed as an analytically derived, forward-thinking “decision” by the agent. A set of behavioral stories also fit into the *ex ante* moral hazard category, albeit without a “conscious” decision to shirk. For example, someone who succumbs to temptation (hyperbolic, quasihyperbolic) and thus finds herself without sufficient cash to repay a loan would fall under *ex ante* moral hazard. Thus interventions aimed at addressing cash management of borrowers should be categorized as interventions to address *ex ante* moral hazard. At the institutional level, interventions include improving legal institutions for enforcement as well as improving internal banking information and payment systems so that banks can engage in enforcement activities more efficiently.

We now review the three main categories of market failures. We note however that although the theory as currently written clearly separates these mechanisms, in practice markets may, and likely do, contain elements of all three. For example, high interest rates may attract individuals who intend to exert less effort (as compared to the traditional adverse selection story that high interest rates attract individuals with risky projects). We now discuss such interactions in more detail.

5.1.1 Adverse and advantageous selection

Lenders set interest rates to maximize returns and may charge higher rates to riskier borrowers to compensate for higher anticipated default. Stiglitz and Weiss (1981) demonstrate that under imperfect information (in particular about borrowers' production functions) higher interest rates can lead to adverse selection, as only borrowers with riskier projects (and higher returns in a positive state) will be willing to pay such prices. Moreover, by lowering returns in all states, higher interest rates can induce all borrowers to undertake riskier activities.

Advantageous selection is also a possibility, and if such models accurately describe credit markets, would lead to very different policy implications. de Meza and Webb (1987, 1989) discuss the mechanics of such a model, and the difference between their prediction and that of Stiglitz and Weiss (1981). The key difference lies in the assumption about the relationship between risk and return. The de Meza and Webb papers assume that entrepreneurs with higher intrinsic quality have higher returns that first-order stochastically dominate lower quality entrepreneurs. This leads to the opposite result of adverse selection: as a bank raises its interest rate, the marginal client that drops out is a low-quality client.

Boucher, Carter, and Guirkinger (2005, 2008) develop a model demonstrating that imperfect information can lead to a form of rationing of credit in which would-be borrowers are deterred based on the terms of the contract, rather than price. Under this model of "risk rationing," by raising collateral requirements lenders shift so much contractual risk to borrowers that even borrowers with the collateral required to qualify for a loan choose not to borrow under these terms for fear of losing vital collateral. Note that "risk" to the borrower is both about explicit risks such as loss of collateral, whether physical or social, and also is about "ambiguity," where "ambiguity" describes the borrower's fear of dealing with formal institutions with which borrowers have little experience. Ambiguity here may also describe the fear of being stigmatized by formal institutions for being poor. Breaking through these barriers may require significant bravery on the part of borrowers. This model assumes that the risk-averse clients, the ones who refuse to borrow at offered terms because they fear the mental anguish of default or the risks of unknown or ambiguous legal and social consequences, are also the low-risk clients in terms of the projects they pursue. As a result, high-return entrepreneurship is limited, particularly among

the less wealthy, and welfare is reduced, as risk-averse agents choose safer but lower paid employment or lower return entrepreneurial activities.

5.1.2 *Ex ante* moral hazard (effort)

A second source of friction in credit markets is *ex ante* moral hazard, or effort. This can take on many forms. By lowering the net profits for a given business, higher interest rates, combined with limited liability, reduce the incentive to invest extra effort in production. Reduced effort diminishes returns to borrowers even when projects are successful and also increases the probability of failed projects and, by reducing the probability of sufficient returns, increases the likelihood of default.

Note that increased default at higher rates is not necessarily due to traditional entrepreneurial shirking or to adverse selection as discussed earlier. While the lender sees only whether or not the loan is repaid, there may be more nuanced explanations for defaulters' lack of sufficient cash flow to make payments. In household finance, higher interest rates may affect borrowers' effort to retain or obtain employment, to tap alternative sources of cash (e.g., borrow from a family member) in the event of a bad shock, or to manage consumption in order to retain sufficient funds for loan repayment. These examples do not relate to a firm's production function, but the same logic applies: at higher interest rates, consumers may exert lesser effort to make sure they have the cash available to repay the loan.

These examples all pertain to *conscious* shirking. However, other mechanisms may be at play, which exhibit the same reduced form prediction: insufficient cash is available at the time of repayment, and the lack of cash is a result of decisions of the borrower. For instance, temptation models predict that, under uncertainty, some types of consumers will consume more in the future than they consider optimal at the time of their decision to borrow. Thus they will have less money available to repay the loan.

Lack of entrepreneurial skills is another source of information asymmetry. If entrepreneurial skills are fixed but unobservable characteristics, one could argue this is also appropriately labeled adverse selection since the lender cannot tell *ex ante* which individuals will likely put in effective effort. The problem is that the lender here cannot enforce "effort" (e.g., being creative, being assertive), similar to the discussion earlier regarding Karlan and Zinman (2009b) with respect to interest rates and information asymmetries.

One way to generate insight into whether entrepreneurial skills are fixed is to try to teach these skills. Karlan and Valdivia (2009) conducted a field experiment in which microcredit borrowers were randomly selected to receive free business training on topics including cash management, business selection, and marketing. Results were mixed. Treated individuals earned greater profits in their businesses, and in particular greater profits in bad months, although this result was not robust to alternative econometric specifications. Along with marginally improved repayment, client retention

increased significantly, suggesting the clients placed great value on the training they received. These institutional benefits led training to be profitable for the microfinance institution, even with no cost sharing on the part of clients.

5.1.3 *Ex post* moral hazard (voluntary default)

Voluntary default is distinct from *ex ante* moral hazard because here the borrower has the ability to repay and chooses not to do so. Thus the problem lies strictly with enforcement of repayment. In the simplest construction, with respect to interest rates, borrowers are more likely to default voluntarily at higher rates than at lower rates since the cost of repaying rises, but the benefits of repaying remain the same. Much of the rhetoric around lending to the poor, including Yunus's strategy with respect to the Grameen Bank, assumes that voluntary default is extremely rare and that entrepreneurial individuals will repay as long as they can.

Demonstration of the presence or absence of *ex post* moral hazard requires identifying situations in which, despite the availability of cash, individuals default on loans by choice. Provided good survey measures, observational data on borrowers should allow us to identify whether default occurs even among clients who have the funds to repay. One could also identify *ex post* moral hazard indirectly, by showing that "trustworthy" individuals are more likely to repay loans. In Karlan (2005), individual borrowers from FINCA, a Peruvian microlending program, played a "trust" game and completed the General Social Survey, which contains questions about trust, fairness, and helping others. Both the survey and the "trust" game were intended to measure individuals' trustworthiness. The "trust" game was conducted with two players and an administrator. First, each player received three Peruvian soles coins. Player A is then instructed pass to Player B any or all of her three coins, which the administrator then doubles. Player B then chooses how much money to return to Player A in response (the administrator does not double the return). Individuals who return more when playing as Player B than when playing as Player A are considered trustworthy. Repayment records from the entrepreneurs' microfinance association show that individuals who show trustworthiness in this game exhibit lower default, lower dropout, and higher savings balances. The survey yielded similar results, showing that individuals who respond more positively to General Social Survey questions about trust, fairness and helping others, are more likely to repay their loans 1 year later.

Drawing such inferences about personality types can lead to a gray area between moral hazard and adverse selection that the theoretical literature has yet to delineate. In the classic case of adverse selection individuals with riskier production functions are more likely to borrow at higher rates and less likely to repay. The effort borrowers invest in ensuring the success of their projects might normally be associated with moral hazard, but if higher interest rates attract borrowers who are inherently apt to expend

less effort in their businesses the issue is really selection. Moral hazard is here inter-linked with adverse selection.

5.2 Interventions and mechanisms

5.2.1 Group lending in theory

Group lending is perhaps the first and most oft-discussed “solution” to information asymmetries in developing countries. Adverse selection and moral hazard are dealt with by effectively shifting the responsibility of screening, monitoring, and enforcement from the lender to clients. Group liability requires that if one group member defaults, her fellow group members will be responsible for her payment. Under group liability schemes then, clients have an incentive to screen other clients so that only trustworthy individuals are allowed into the program. Ghatak (1999) describes how group lending can, in theory, mitigate adverse selection through group formation. Potential clients, selecting fellow borrowers with whom they will be jointly liable for loans, will exploit information known to borrowers but not to banks to screen out bad credit risks. Groups segregate according to riskiness, with riskier borrowers joining with other risky borrowers. Thus, the cross-subsidization of risky customers by safer customers that is presumed to be inherent in individual liability schemes is mitigated. Group lending also addresses moral hazard by providing incentives for clients to employ peer pressure to ensure that funds are invested properly and effort exerted until the loans are repaid in full. By lowering default, the expected total cost of borrowing for borrowers can be reduced, improving welfare especially for households without collateral. Still, because clients face the potential of default from fellow group members, group liability could also be seen as a tax, effectively increasing the net interest rate on safer borrowers (Stiglitz, 1990).

Just as social networks can facilitate informal insurance mechanisms between households, borrowers, who often know each other, can serve as mutual insurance pools for fellow borrowers, covering loan payments when a fellow group member is unable to. Moreover, their familiarity should allow for mutual monitoring to avoid making payouts due to moral hazard. If social networks are important in fostering repayment under group liability we should see higher repayment where borrowers are more connected to each other. Wydick (1999) provided early evidence of this in Guatemala, testing the effect of three different types of social cohesion: peer monitoring, social ties, and borrowing group pressure. Wydick finds monitoring to be the most important of the different forms of cohesion, with little effect from social ties. Ahlin and Townsend (2007), with data from Thailand, find social ties (measured by sharing among nonrelatives, cooperation, clustering of relatives, and village-run savings and loan institutions) actually *reduce* repayment rates, though stronger social sanctions improve them.

The conclusions one can safely draw from these findings are limited, however, because borrowers self-select into groups. Group formation is determined by factors

endogenous to the forces of the liability scheme itself. If, for instance, borrowers with successful enterprises have greater social networks, there may be a correlation between social ties and repayment that has nothing to do with group monitoring, enforcement, or cohesion, but rather is due to success in entrepreneurship. Karlan (2007) uses quasi-random variation in the group-formation process at a Peruvian microfinance institution to show that groups with greater levels of social connection (ethnic ties and geographic proximity) have lower default and higher savings rates. Greater knowledge of default status among fellow borrowers, and deterioration of relationships after default, suggests both monitoring and enforcement play a role.

The structure of group-liability contracts produces an ambiguous effect on risk-taking in entrepreneurial investments and thus on returns (Fischer, 2008). The insurance aspect of joint liability creates an incentive to take risks by lowering the cost of default, while the monitoring and enforcement (group pressure) aspects increase the cost of default. To generate empirical data on the balance of the effect, Fischer ran a series of investment games in which microfinance clients in India “borrowed” and “invested” according to different types of contracts where players could share risk by making income transfers to their partners. Returns were randomized and payouts from the contracts were in real money of up to a typical week’s income. Each treatment was conducted under both complete information, where all actions and outcomes were observable, and limited information, where players observed only whether their partners earned sufficient income to repay their loans.

Several key implications of joint liability emerged from the results. First, joint liability produced free riding: risk-tolerant individuals (as measured in a separate risk game) made significantly riskier investments under limited information, without compensating their partners for the insurance they provided. However, under complete information joint liability did not encourage greater risk-taking. Peer monitoring via approval over partners’ investment decisions mitigated *ex ante* moral hazard by discouraging risky investment choices.

Fischer’s results are broadly consistent with those in Giné, Jakieła, Karlan, and Morduch (2009), generated from games conducted with entrepreneurs in Lima, Peru. Here, too, joint liability encourages free riding, but the mutual insurance it provides prevents default from increasing. Moral hazard is reduced by allowing clients to form their own two-person groups, but interestingly there is no apparent pattern of matching between risk-averse partners. Instead the effect seems to be one of fairness: though safe-risky investment pairings are viable over the long term, and generate higher returns, both partners opt to either both choose risky investments or both choose safe investments. This leads to suboptimal risk-taking overall. The policy conclusions from these findings depend on the returns to real-world investment choices, but the assumption that riskier projects would be rewarded seems reasonable.

Conning (2005) develops a model to compare outcomes under both individual and joint liability and finds no clear winner: each is optimal under different circumstances, depending on the type of borrower. Interestingly, his conclusions do not rely on an information advantage of borrowers over bank delegates. The microfinance industry collects a great deal of information on the financial performance of its institutions, and these data are helpfully broken down by institution type, including lending methodology. Drawing useful inferences from the cross-sectional data about which specific interventions, like group lending, mitigate information asymmetries can be difficult because of the difference in target markets served by each lender type. Individual loans, for instance, are most often offered to borrowers at the upper tier of the microfinance market. These borrowers are likely to have levels of assets (including working capital, human capital, and collateral) and production functions distinct from the type of entrepreneurs typically served by group-lending institutions. Interest rates, too, are correlated with poverty levels: the fixed costs of processing loans imply that interest rates will be higher on smaller loans.

Cull, Demirgüç-Kunt, and Morduch (2007) use a dataset comprising 124 microfinance institutions across 49 countries to analyze MFI performance and outreach. They find patterns in the data in line with the theoretical literature on adverse selection and moral hazard, but only for certain types of institutions. Specifically, lenders making individual loans show higher portfolio-at-risk (PAR) as interest rates increase, and after a certain point (60% annualized percentage rate) profits fall off. Reduced demand at higher rates likely contributes to this effect. This trend does not hold, however, for group lenders, suggesting that they are better able to mitigate problems stemming from information asymmetries.

5.2.2 Group lending and related mechanisms in practice

While lenders have been attracted to the intuitive appeal of relying on borrowers to monitor themselves, through group enforcement, recent successes of several MFIs offering collateral-free individual-liability loans to poor borrowers have caught the attention of the microfinance industry. While individual-liability schemes cannot overcome the problem of information asymmetries between client and lender, the methodology does avoid other problems typical of group lending, such as strategic default among group members. Nevertheless, the strong financial performance of a few lenders is not sufficient evidence to advocate the end of group lending. The low default rates among existing individual-liability borrowers may be indicative of superior management skills among these few programs, heroic (or coercive) efforts by field staff to ensure repayment, a culture of repayment in these regions (notably Bangladesh), or credit constraints among clients (who want to maintain access to future loans).

Armendáriz and Morduch (2000) report on microfinance institutions in Eastern Europe and Asia that use a combination of mechanisms, such as rebates, larger loan sizes, and faster loan approval for safe clients, making individual lending possible and profitable. Whether the bank employs incentives or sanctions, in theory the bank will extract payoffs up to the borrower's opportunity cost of not repaying. The size of the opportunity cost will depend on borrowers' outside options for credit. It can therefore be difficult to predict borrower behavior under different contracts. In two experiments with a microlender in the Philippines, Giné and Karlan (2009) first randomly assign groups of joint liability clients to either remain under joint liability or to have their loans converted to individual liability, and, second, randomized before the initial creation of lending groups whether they would be group or individual liability lending groups. All other aspects of the loan contract—interest rate, payments, and term—remained the same. For the first experiment, follow-up of the study participants after 3 years revealed that converting from group to individual liability had no adverse affect on default rates, despite the fact that the conversion reduced peer monitoring of loans (note that group meetings remained intact, just the group liability was removed). Furthermore, results showed greater client retention among individual-liability borrowers, making the switch appear profitable for the bank and appealing to clients. The second experiment then also allowed for selection effects from the group liability, by preannouncing before the creation of groups whether or not the village was eligible for group or individual liability loans (again, keeping group meetings intact, and just varying the liability rules). Similar to the first experiment, repayment rates remained high, near 100%. However, the credit officers in this second experiment were less likely to enter certain villages and create groups, thus indicating fear of default from the bankers' perspective. This could either be because credit officers know the relative risks, or due to the lack of learning by the credit officers that the group liability is in fact not a necessary component to generate high repayment.

5.2.3 Repayment schedules

Rutherford (2000a) argues that “financial services for poor people are largely a matter of mechanisms that allow them to convert a series of savings into usefully large lump sums.” The difference between credit and savings is simply whether the lump sum is taken at the beginning or the end of the cycle. The deposits are small, of necessity, while the lump sums are used for all sorts of expenditures: life-cycle events (births, school fees, home building, deaths), emergencies, and investments. By way of describing several types of financial institutions for the poor, from a homegrown “merry-go-round” savings scheme to an NGO-created MFI, Rutherford maps out the relationship between the complexity of the institution and the flexibility of the products they offer. The more formal the institution, the more likely it is to be able to turn savings into a lump sum at a time that is convenient for clients. With simple savings,

club members must wait for their turn at the lump sum handed out each week, and often the club must be reformed at the end of every cycle. But there can be tradeoffs on the other end too; the local moneylenders and deposit collectors of the informal market provide services with convenience, a measure on which formal sector services have fallen short.

Seen in this way, credit can be a useful mechanism to help people *save* toward the purchase of an asset. Microfinance programs emphasize small, frequent, regular payments, and create incentives for clients to make those payments. For those who have trouble making regular savings deposits, borrowing can be a way to commit to making those payments (albeit at greater total cost to the borrower). But are bank contracts sufficient, or do these savers need the social pressure of group loans to continue the payments? Basu (2008) develops a model showing hyperbolic discounters, who place a greater value on present consumption than future consumption (and therefore find it harder to set aside savings), will prefer to remain in rotating savings and credit associations (ROSCAs) even when there are no sanctions for absconding with their loans—because it is preferable to have the commitment device of the ROSCA over the long run than free money today. Following a similar logic, Bauer, Chytilova, and Morduch (2008) use data from south India to show that hyperbolic discounters borrow more and save less than others, conditional on household characteristics, and that they are especially likely to borrow from microcredit programs which offer helpful structure and social support.

Rutherford emphasizes flexibility in addition to stability and security as important factors in clients' take-up decisions, but it is unclear which institution types are best in this regard. While some MFIs offer a range of loan, savings, and insurance products, others offer only one: for example, a 4- or 12-month enterprise loan with a weekly payment schedule that may not match the cash flow of the poor. It is striking how many MFIs require loan repayments beginning only 2 weeks after disbursement—and how few microfinance account officers believe their clients' investments will start to pay off by that time (Armendariz de Aghion & Morduch, 2000). Moreover, there is a substantial transaction cost to conducting these meetings, for both the clients and the MFI. However, many MFI managers consider weekly payments essential for several reasons: (1) the smaller repayment amounts are easier for clients to manage; (2) it creates a culture or discipline of repayment for those not used to formal payment deadlines; and (3) the regular meetings prevent attention problems (e.g. forgetting when the repayment meeting is supposed to take place). In a theoretical paper, Jain and Mansuri (2003) give another possible reason: that the frequent repayment schedules force borrowers to turn in part to the informal sector, which is beneficial if the banks can then piggy-back on monitoring by moneylenders.

Empirically, the question of what is the optimal repayment frequency for borrower and lender remains unanswered. Field and Pande (2008) conducted a simple field experiment in India in which they randomly assigned new borrowers either to

traditional weekly payments or to monthly payments. They find no difference in repayment, although the results are preliminary (after 1 year), pertain to small-sized loans only, and hold for new borrowers, not long-term borrowers. McIntosh (2008) extends the Field and Pande result with a study of a Ugandan MFI in which the bank offered its village banks a choice between weekly or biweekly repayment. The choice had to be made by unanimous vote of each village bank's members. A simple comparison of the repayment performance of weekly to biweekly banks would be laden with two sources of bias: geographic selection by the MFI, and self-selection by the clients. McIntosh constructs a comparison group by asking clients in ineligible regions to decide whether they would want to switch from biweekly repayment, had it been offered to them. By making two comparisons: groups that switch to biweekly payments versus ineligible groups, and would-be switchers versus those who would not choose to switch, McIntosh is able to eliminate some of the bias in the estimate. He finds no drop in repayment (actually a slight improvement) and a large increase in client retention (dropout is reduced by 40%).

5.2.4 Credit scoring and credit bureaus

Adverse selection problems can be mitigated by lenders through better screening processes, a process which lenders have been continually refining. Credit scoring has become more sophisticated over time, but most developing countries still lack credit bureaus (and in many cases, unique identification numbers for citizens). It is often difficult to evaluate implementations of national programs but de Janvry et al. (2008) exploit a natural experiment in which a large MFI in Guatemala installed new credit reporting hardware in waves, allowing the researchers to compare early adopting branches to late adopters. Findings from the study show that with the new technology the lender screens out substantially more applicants, but also makes more new loans. New borrowers have higher repayment and take larger loans. In related research the same authors worked with an MFI to randomly assign some clients to receive training on the importance of credit bureaus to borrowers' credit opportunities. The clients were informed both that late payment with one lender will harm their access to credit at other lenders and that paying on time gives them greater access to credit at potentially lower rates. The authors find that the training led to higher repayment rates by their clients, but also led their clients to borrow elsewhere after establishing a good credit record.

While credit bureaus can clearly play a valuable role in leveling information asymmetries, they nevertheless are only capable of providing data. Lenders must still interpret the data and decide whether to approve loans, and at what terms. Working with a consumer-credit lender in South Africa, Karlan and Zinman (2009a, 2009b) show that there is room for improvement in the screening process. By extending loans to randomly selected applicants who would have been rejected under the bank's own scoring system, they show an increase in outreach and profit for the lender, despite

the fact that the broader client base carries with it higher default. By comparing the welfare of these randomly approved marginal clients to a control group of identically marginal applicants who remain denied, the researchers find that the loans are significantly welfare enhancing for borrower households.

5.2.5 Dynamic incentives

Dynamic incentives have also been credited with helping to solve moral hazard problems. Theoretical work beginning with Bolton and Scharfstein (1990) has shown the conditions under which dynamic incentives, for example, the threat of termination of credit, can generate sufficient incentives for the borrower to repay loans. Whereas this work established a clear theoretical understanding of the potential role for dynamic incentives, in practice specific methods of implementing dynamic incentives may or may not work, and evidence is needed to understand how this theory can be employed to improve product design. For example, progressive lending, in which initial loan sizes are capped low and progressively larger loans are only available after successful repayment, has become the norm for many microfinance lenders.

Clear evidence on dynamic incentive components of the contract structure is more difficult to show, as most lenders only vary contract terms endogenously, for selected clients. In one study in South Africa, discussed earlier as well, Karlan and Zinman (2009b) worked with a lender who randomly offered some clients a dynamic incentive, a discount on future loans assuming successful repayment of their current loans. This offer led to a 10% reduction in default (from a base default rate of 15%), and the responsiveness was proportional to the size of the incentive. This simple experiment naturally opens the door to further questions, such as whether the framing of the incentive matters (e.g., loss versus gain framing), and whether a dynamic incentive can continue to work over the long term or will instead collapse over time, either due to capacity constraints of the lender or to changes in responsiveness to incentives over time by the borrowers.

Controlled laboratory experiments can shed insight into questions that may be difficult to answer through a natural field experiment (for a taxonomy of field experiment methodologies, see Harrison & List, 2004). Lenders are often resistant to making the changes to operations required of a field experiment. Few lenders will, for example, agree to a field experiment in which further loans are made after default, or the reverse, to pledging not to lend any more even to good borrowers. For such tests, we turn to controlled laboratory experiments. Several papers have examined the power of dynamic incentives, on topics far afield from microfinance. Relevant papers include Charness and Gennicot (2007), and, as described earlier, on microfinance by Giné et al.(2009) and Fischer (2008). In each of these papers, the dynamic incentive led to the largest effects on reducing moral hazard problems.

As a general point, dynamic incentives on lending contracts are not much different than conditional transfer programs. Like conditional transfer programs, dynamic

incentives provide a future incentive (e.g., larger or cheaper loans rather than cash) in exchange for a change in behavior now (e.g., lower moral hazard rather than higher school attendance). The interesting questions here are not whether people respond to incentives, but rather how to structure the incentives to be socially optimal, how to frame the incentives so as to maximize their immediate and long-term effectiveness, and how to structure and time the incentives so as to maximize social welfare. On the last point, an excellent example comes from Colombia where the government tested the importance of timing in a conditional cash transfer program by randomly assigning some individuals to receive their conditional cash transfers at the time school fees were due rather than before. This is a classic example of Thaler and Sunstein (2003, 2008) employed in developing countries and provides a clear example of how the structure (in this case, the timing) of the incentive, not just the incentive itself, can affect social outcomes.

5.3 Impacts from solving credit market failures

For a poverty intervention as widespread as microfinance, with an estimated 154 million clients worldwide (Daley-Harris, 2009) and over 5 billion dollars invested each year (Forster & Reillie, 2008), there is surprisingly little rigorous evidence of the impact of microfinance on household welfare. In part this is due to the difficulties of measuring impact. There is ample reason to think that selection bias may be especially problematic with nonexperimental evaluations of microfinance programs. Microfinance clients are likely to possess a special determination and ability to improve their welfare and therefore comparing their outcomes to the outcomes of nonclients (presumably without this drive) will overstate the impact of microfinance. Unfortunately, personal characteristics like “entrepreneurial ability” or “drive” are either difficult to measure or unobservable.

A related challenge is bias from program placement, in which outcomes in program villages are compared to outcomes in nonprogram villages. The problem with this method is that programs choose where they operate for a reason. They may target the poorest villages, for instance, or they may start cautiously with better-off clients before expanding their outreach. The bias from nonrandom program placement, therefore, can go either way, depending on whether the evaluation compares program villages to nonprogram villages that may be (even unobservably) better or worse off.

It will not be clear how serious a problem these omitted variables are until rigorous impact assessments of credit are completed, side by side with alternative assessments, and the results can be compared. As a first step, Coleman (1999) shows how important selection bias can be in a study of microfinance borrowers in northern Thailand. By forming a group of prospective microfinance clients who signed up a year in advance to participate with two village banks, Coleman was able to create a comparison group mostly free of selection bias, as both the borrowers and the nonborrowers had selected

into the program at the same point in time. Coleman then generates two estimates of the impact of the program: an unbiased estimate using the clients who signed up in advance as the comparison group; and a “naïve” estimate using a group of seemingly similar nonparticipants (as in typical nonrigorous evaluations). Comparing his unbiased impact estimate to the estimate he would have calculated had he naïvely compared program participants to a group of nonparticipants Coleman finds the “naïve” estimate substantially overstated the gains from participation on several outcomes (especially women’s landholding) because participants turned out to be initially wealthier than nonparticipants.

In an ambitious early attempt to solve these identification problems, Pitt and Khandker (1998) surveyed 1798 member and nonmember households of three Bangladeshi MFIs (Grameen Bank, BRAC, and RD-12), and exploit eligibility criteria (landholding totaling less than one-half acre) to obtain measures of impact of credit. While there should be no discontinuity in income between people who own just over or just under a half acre of land, participation in the MFIs would be discontinuous because those who were above the cutoff would be rejected from the programs. They find huge impacts: every 100 taka lent to a female borrower increased household consumption by 18 taka. However, in a reanalysis of the data Morduch (1998) challenges the econometric models and identification assumptions in Pitt and Khandker (1998). Using a difference-in-difference model, he finds little evidence for increased consumption but does find reduction in the variance in consumption across seasons. He argues that in the cross-sectional setup, nonrandom program placement is only addressed through restrictive assumptions and that reliable inferences require additional data.

Khandker (2005) refined the earlier model with the benefit of panel data, finding lower impact estimates but greater total impact (from current and past borrowing in the survey rounds conducted in 1991–1992 and 1998–1999) and substantially lower marginal impact from new borrowing. Poorer clients are found to have larger impacts than the less poor, and money lent to men is not shown to have any impact at all.

Roodman and Morduch (2009) attempt to find closure to the issue by returning to the data and rebuilding the analysis from scratch. They are unable to replicate results from Pitt and Khandker (1998) or Khandker (2005). In fact, their estimates carry the opposite sign. Rather than concluding that microcredit harms borrowers, however, they unearth a raft of identification issues which are not solved with panel data. Their revised analysis casts doubt on *all* of the findings from the related set of papers, including Morduch’s (1998) finding on consumption smoothing.

Karlan and Zinman (2008b) conducted a randomized controlled trial to measure the welfare effects on borrowers at the same time that they worked with the South African consumer-credit lender to deepen outreach. They find the loans are substantially welfare enhancing for the borrowers, who are 11% points more likely to remain employed, are less likely to encounter periods of hunger or to score as poor, and had higher

income and improved credit reports. It should be noted this is the impact on a particular set of borrowers: those who would have been rejected under the lender's standard approval requirements.

Burgess and Pande (2005) use the introduction and eventual repeal of an Indian social banking law to evaluate the impact of access to finance on the rural poor. Between 1977 and 1990 Indian banks wishing to open new branches in locations already served by any commercial bank were required to open four new branches in unserved locations. Since the policy required banks to focus on locations with the lowest level of financial access the authors use the pre-1977 levels of financial intermediation as an instrument for the states that will be most affected by the policy. They find a 1 percentage point increase in the share of credit disbursed by rural branches reduces rural poverty by 1.5 percentage points, while a 1 percentage point increase in the share of savings held by rural banks reduces poverty by 2.2 percentage points. The opening of one bank branch per 100,000 people in a rural unbanked location reduces rural poverty by 4.7%. As optimistic as these results are, the authors caution that default rates during the expansion period were in the 40% range, leading to the abandonment of the program. As always, appropriate mechanisms must be designed for credit to be sustainable. The authors' caution also creates uncertainty about whether the results are due to access to finance *per se* or from the fact that a large share of customers received money as loans but never repaid them (creating implicit cash transfers).

Several randomized studies are being conducted on both individual lending and randomized program placement designs, similar to the Progresa evaluation in Mexico. These differ from the earlier mentioned South Africa experiment in that they are traditional "microfinance" credit programs that target entrepreneurs (rather than customers looking for consumer loans). Results from the first two trials, from urban India and the Philippines, challenge popularly held assumptions about the use of credit. Other studies underway in rural India, Morocco, rural Peru, urban and rural Mexico, and China will help paint a more complete picture of whether or not microfinance is welfare improving, by what measures, for whom, and under what conditions.

Karlan and Zinman (2009a) uses a credit-scoring methodology to evaluate the impact of loans to microentrepreneurs in urban Philippines. The methodology used is similar to Karlan and Zinman (2008a, 2008b), earlier, however, there the focus was on loans made to employees. Here the effects are much more muted, and some findings cast doubt on the traditional microfinance narrative. Business owners' profits increase, but not through investment in productive assets or working capital. Moreover, the treatment effects are stronger for groups that are not typically targeted by microlenders: male and higher income entrepreneurs. There is evidence that treated businesses actually shrink in size and scope, including the shedding of paid employees. The results suggest that borrowers used credit to reoptimize business investment in a way that produced smaller, lower cost, and more profitable businesses. The question remains as to how credit enabled this

change: why did business owners need to borrow to reduce staff? One potential explanation is household risk management: treated individuals substitute out of formal insurance products, while also reporting a greater ability to borrow from friends or family in an emergency. It is possible that before credit entrepreneurs were retaining unproductive employees as a kind of informal mutual benefit scheme. Those employees, even if unprofitable, were an additional place to turn in times of need.

In urban India, Banerjee et al. (2009) evaluate the impact of a nonprofit group-lending microfinance program in the slums of Hyderabad. The researchers randomly assigned 120 slums to either treatment or control. After 15–18 months the households from the treatment slums were compared to the households in the control slums. The results show impacts on a number of dimensions, though not critically, on average consumption. The treatment slums have greater investment in business durables, increases in the number of businesses started, and in the profitability of existing businesses. Among households that did not have existing businesses at the start of the program, those with high propensity to become entrepreneurs see a decrease in consumption,⁸ while those with low propensity to become entrepreneurs increase consumption. Likely this difference is explained by investment in durable goods among those likely to become business owners. While the short-term impacts are clear, this result makes it difficult to anticipate the long-term impacts. As the authors speculate, these investments may pay off in future consumption in the coming years. The increase in consumption among nonbusiness owners has an even more ambiguous future: if these households used credit to temporarily increase consumption they will have to reduce future consumption to pay down debts. Alternatively, if they used the credit to pay down high-cost moneylender debt, then their current consumption should remain high.

6. THE ECONOMICS OF SAVING

A large macroeconomic literature exists to understand national savings rates, their determinants, and implications. Policymakers often strive for target rates, but the variance in national savings rates is remarkable in both developed and developing countries (Gersovitz, 1988; Horioka, 2006). As Rosenzweig (2001) and Collins et al. (2009) argue, focusing on the saving rate places emphasis on asset levels at a given point in time, but that misses the value of savings for many poor households. Many poor households may be actively saving even if their assets at any given moment are low; instead, they are building up lumps of money and spending them within a year. To see this, we need to turn to microdata.

Microhousehold and individual data reveal much about personal savings rates, the decision-making process at the individual and household levels, and the impact on individuals and households from access to different savings services. We will focus specifically on how informal institutions (e.g., ROSCAs) and formal services influence

savings decisions. We will then conclude with a discussion about measuring the impacts from deepening access, both in terms of the quantity (e.g., lower transaction costs) and the quality of access to savings.

Where households in developed countries, including the poor, may have many products available (savings accounts, automatic transfers, savings bonds, certificates of deposit) to them to help them save and build assets, the poor in developing countries face a much more limited menu of options. Those who are able to save are often forced to invest in risky assets like jewelry or animals or to use informal savings arrangements (e.g., ROSCAs). It is easy to imagine that households in poor countries would save more if they were given access to a broader array of quality savings products. Such access would enable the building of safety nets to smooth shocks and greater accumulation for the purchase of indivisible goods. Much of the discussion on how products influence decisions will build from knowledge and innovation from developed countries, with discussion and examples of applications to developing countries—and a focus on the complexities, risks, and infrastructure unique to the poor in developing countries.

6.0.1 Revisiting “the poor can't save” assumption

The historical emphasis on credit in the microfinance movement implicitly assumes that the poor cannot save up for investments and instead must borrow at relatively high interest rates in order to make investments. Criticism of this view is not new, and was argued strongly by the rural finance group at Ohio State University in the 1970s and onward, following McKinnon (1973) and Shaw (1973). The argument has been revisited by some donors, notably the Bill and Melinda Gates Foundation, with a renewed focus on encouraging the promotion of savings in rural settings (Guth, 2008). Adams (1978) decries the lack of focus on expanding rural savings capacity and argues that only a handful of countries stress mobilization of voluntary household savings, arguing that policymakers have too quickly assumed that rural households are too poor to save, that there is no excess cash flow from income, and that those that do acquire additional income spend it on consumption or ceremonial expenses.

Yet evidence on rural savings behavior from various economies has shown impressive propensities to save among rural households. Adams cites the remarkable savings rates achieved in post-World War II East Asia: the average propensity to save in 1973–1974 increased to 0.31 in Taiwan (compared to 0.19 in 1960), 0.22 in Japan (compared to 0.10 in 1950), and 0.33 in Korea (compared to 0.04 in 1965)—a product, Adams argues, of prosaving public policy. The argument continues that while rural households have a substantial capacity for voluntary saving, their capacity is adversely influenced by rural financial markets that tend to discourage savers through subsidized (but unreliable and unsustainable) credit.

The focus on household decision making is not new either. For example, von Pischke (1978) argues that the rural poor can save, but that they have specific needs due to low account balances, seasonal income, remote location, noncash assets, and high transaction operations. The development challenge in encouraging savings lies in the design of financial technology to serve these specific needs.

6.0.2 Problem with concepts of poverty-destitution

National income accounting and purchasing power parity (PPP) calculations tell us that people in poor countries, on average, do not have a lot of money. But of course what this implies for policy requires much more information. Even for those at the bottom of the income distribution, themselves in countries at the bottom of the global distribution, the fact of earning very little money is insufficient for making reliable inferences about the types of products and services that can be afforded, much less valued.

Take, for instance, Ethiopia, where 23% of the population lives below \$1 per day (PPP) (UNDP). A family of seven at this international poverty line would consume roughly \$210 worth of goods and services per month—if they were purchased at US prices. PPP measures are debated (Deaton, 2006), with particular questions about the transferability for the bundle of goods that are purchased by the poor (vs. that of the median consumer, for example). However, even if these measures are underestimated by a factor of two, this still implies monthly household consumption of the equivalent of \$420 in the United States. These amounts buy so little in terms of meeting an entire family's needs in a rich country that it can be hard to fathom how they get by at all. Going further requires knowing more about the finances of the poor: what do they spend on food, housing, and healthcare? And do they have any money left over at the end of the month?

Banerjee and Duflo (2007) tackle this question: “how actually does one live on less than 1 dollar per day,” by assembling a dataset comprising existing detailed survey data (mostly World Bank LSMS surveys and Rand Family Life Surveys) representing the expenditures of poor households (including households living under \$1 per day and \$2 per day) in 13 countries. The surveys were conducted between 1988 and 2005. They find among those living under \$1 per day between 56% and 78% of household income is spent on food (slightly less in urban areas). Only around 2% of income is spent on education, and a bit less on healthcare. Housing does not seem to be a major expense, perhaps because many very poor households effectively own their own land (though their holdings are quite small).

Though the very poor often fail to meet minimum caloric requirements they choose to make expenditures on many items like alcohol, tobacco, festivals, and radios. As if to reemphasize that these households are truly poor despite their disposable income the authors describe the sample from Udaipur, India, for whom they have the most detailed asset data: among the extremely poor most have a bed or a cot but only 10% own a chair and 5% a table. Less than 1% has an electric fan, a sewing machine, or a bullock cart and no one has a phone. Clearly the very poor are choosing

to spend their resources in particular ways. Like households everywhere they have to make choices between consumption, investment, and savings.

Having the right financial products conveniently available might help them make the choices they prefer over the long term. Though some of the surveys predate the global explosion of microfinance, the lack of financial access is nonetheless striking. While borrowing is quite common among the extremely poor (ranging from 11% of households in rural East Timor to 93% in Pakistan), little of it is conducted with a formal institution. By and large the extremely poor borrow from relatives, shopkeepers, and fellow villagers. The pattern is consistent with recent data from India reported by Banerjee and Duflo (2007), where only 6.4% of borrowing is from a bank or cooperative, even when there is a branch nearby. This informal credit is expensive: borrowers pay nearly 4% per month. Collins et al. (2009) report similar data in their samples from Bangladesh, India, and South Africa.

Savings is no better. Except for one notable outlier (Cote d'Ivoire, where 79% of extremely poor households have a savings account) the fraction is below 14% in the other countries (Banerjee & Duflo, 2007). That may be even more problematic than the lack of credit since, as the data suggest, even the very poor can use help resisting the temptation to spend money on immediate consumption.

6.1 Basic models of saving

Browning and Lusardi (1996) review nine models used to explain motivations to save: precautionary, life cycle (to provide for anticipated needs), intertemporal substitution (to enjoy interest), improvement (to enjoy increasing expenditure), independence, enterprise, bequest, avarice, and downpayment. The authors then review the major economic theories of savings, before reviewing empirical evidence on these theories. The data provide a valuable description of who saves and how saving rates have changed over time, but no unique theory explains why people save. We refer the reader to this and other reviews (e.g., Armendáriz & Morduch, 2010; Rosenzweig, 2001), and instead here focus on the key motivating and classic models, and then turn to the evidence from developing countries to help understand household decision-making processes.

The life-cycle hypothesis (Ando & Modigliani, 1963; Modigliani, 1986) remains the most influential model of savings. The life-cycle hypothesis (LCH) framework articulates the relationship between consumption, income, wealth, and savings, over the life of individuals. Its central insight is that households have a finite life and a long-term view of their income and consumption needs. They therefore increase their wealth during their working life and use it to smooth consumption during retirement. Wealth itself can come from the accumulation of savings (the difference between “permanent” and “transitory” income) or from bequests. The life-cycle hypothesis was one of the first models used to explain savings; it is supported by many empirical analyses in rich countries and is robust to varying assumptions.

As discussed by Deaton (1997), the life-cycle hypothesis requires adaptation to fit the multigenerational households commonly found in developing countries. If, for example, the social norm is that a household unit has grandparents, children, and grandchildren, the need to save over one's lifetime is diminished. Rather, intergenerational transfers replace the need for savings and borrowing over one's life. Naturally, demographic transitions (e.g., AIDS epidemics leading to premature deaths of income-earning generations) can wreak havoc on the ability of intergenerational households to transfer wealth, thus making it important to consider how such trends influence life-cycle savings decisions as the LCH suggests.

For poor households, precautionary savings models are often a better fit (Deaton, 1997). The models capture the fact that for many poor households the volatility of income and the inability to borrow to smooth consumption is potentially just as damaging as a persistently low level of consumption. Rutherford (2000b) puts forward a simple prediction, one that also falls out of most models of savings that generate a preference for smooth consumption: the poor need mechanisms to make small deposits and large withdrawals. The idea is that poor people can save and that they want to save in order to meet life-cycle needs, cope with emergencies, acquire assets, and develop businesses. Most of these needs come in lump-sums, however, whereas income often comes in little installments (cash labor income, or entrepreneurial income). One exception is agricultural income, with a small number of harvests per year, and we will now discuss this.

Putting these points together: the poor have uneven cash flows (thus the need to save) and they have available income (thus an ability to save). What they require is a safe and convenient place to keep their money and a structure with which to discipline the accumulation of lots of small sums and their transformation into a large sum. Rutherford (2000a) eloquently articulates that the poor can and do save using a variety of homegrown mechanisms including ROSCAs and deposit collectors who *charge* the poor to take their savings. Each option presents tradeoffs in convenience, risk, price, and simplicity. The importance of the last dimension, simplicity, should not be underestimated: as programs and products get more complex, they have a greater potential to meet the specific financial needs of the poor—but they also become harder for clients to understand and manage. Good research can help programs identify optimal combinations of all these dimensions by determining which factors “sell” the best to clients and by providing an understanding of the literacy necessary for a product to succeed (both in terms of take-up and proper usage). However, as we discuss in the following section, even the “best” combination of convenience, security, and price may not be enough: just because people make decisions does not necessarily mean they make good decisions—and by “good” we mean decisions about actions that people say *they want to make*—like building a safety net for hard times or saving up for profitable investments.

Features such as transaction costs, liquidity, and interest rates influence the take-up and usage of financial services. But other factors matter, perhaps in some cases even more. In an experiment in South Africa, Bertrand, Karlan, Mullainathan, Shafir, and Zinman (2010) tested the relative importance of interest rates, marketing features, and choices on a direct mail solicitation to consumer borrowers from a regulated for-profit microlender. They found that interest rates mattered, but that simple marketing ideas mattered even more. For instance, giving consumers only one choice on loan size, rather than four, increased the take-up of loans just as much as if the lender reduced the interest rate by about 20%. Even more striking, they found that if there was a photograph of a woman on the direct mail solicitation, this drove take-up, among both men and women, of the loans just as much as if the lender reduced the interest rate by about 33%. These are large effects, particularly given the industry focus on “important” characteristics, such as price.

What does this tell us? We need to pay attention to more than just the pure economics of the choices being offered. The way offers are presented can have just as much to do with take-up and usage as do the terms of the account. Recent work by Richard Thaler and Cass Sunstein has documented a plethora of examples of “choice architecture” in which the way choices are presented and structured may matter just as much, if not more, as the choices themselves (see Thaler & Sunstein, 2003, 2008).

Given that the presentation of choice matters, targeted research can help answer the question of how products and processes can be designed to most effectively assist poor people to accomplish their stated goals. Are there certain temptations, for instance, that individuals would actually prefer not to have? For example, do individuals find they purchase items, such as entertainment goods, that they later regret having purchased (when, for example, a health shock occurs)? Commitment savings accounts can help prevent consumption of goods that are later regretted. Altering “default” settings is another example of how product design can *nudge* individuals toward decisions they say they prefer. In most decisions, something must be established as the action that occurs if no alternative action is taken. For example, should a proportion of a remittance or paycheck be set aside automatically into a savings account?

In order to understand how product design might matter, we need to first understand something deeper about the psychology of the decision-making process and the household and societal constraints that lead to certain decisions and coping strategies. Furthermore, employing one design over another requires recognizing that there is rarely a “neutral” option: choices in product design will affect how the product is used and by whom. Some users may benefit more than others in particular ways. As such, every choice about how to design and offer a product requires taking a normative stand on what will improve outcomes for individuals.

6.2 Constraints to saving

If people do not save as much as they wish they would, why not: what are the constraints? We think about these as demand-side and supply-side constraints, although clearly there is a relationship between the two.

It is useful to first note the high price the poor are willing to pay for savings services that they can trust. This clearly shows the high value they place on savings more generally, and thus the welfare improvement possible from identifying policies that provide savings options for the poor at lower costs. For example, a ROSCA is a common method of savings that provides zero interest income and significant loss risk. Wright and Mutesasira (2001) document these risks for informal savings. Through a study of MicroSave data, they indicate that people with access to the formal sector reported much higher savings than those without access. The percentage of those reporting losses from formal mechanisms (15%) was much lower than reported losses in the semi-formal (26%) and informal (99%, albeit loosely defined) sectors. On average, clients in the informal sector lost 22% of their savings. Note that this includes informal savings such as livestock, which can die but also provide a potential return if they live. Thus this is not the ideal analysis to show that, holding returns constant, informal savings are riskier. In this light, the issue of allowing microfinance institutions to offer savings products should be considered in terms of relative risk, since savings are relatively safer at microfinance institutions than in informal mechanisms. Rather than dictate the decision of where the poor save, it is key to help them make informed choices by helping them understand the relative risk of semiformal institutions.

The popularity of deposit collectors also offers clear evidence of the demand for savings and offers insight into some of the potential behavioral constraints on savings. In Ghana, for instance, individuals pay for informal deposit collection services through susu collectors who travel to individual homes or businesses at regular intervals in order to collect savings deposits. In some cases, payment for this service is high enough that individuals' real return on savings is negative. Aryeetey and Steel (1995) document this and discuss the basic structure of these services. The susu collectors collect deposits from customers (mainly women operating market stalls) every day and return the accumulated savings at the end of the month. They neither pay nor charge interest on the deposits but they keep one deposit per month as a fee. This 3.3% fee works out to a negative annual return of 54% for those who withdraw their deposits at the end of each month.⁹

6.2.1 Self and spousal control: Arguments for commitment

These issues lead to a reevaluation of the two views that dominate thinking around the financial behavior of the poor (Mullainathan & Shafir, 2008). One view positions the poor as rational individuals who are methodical and calculating in their financial decisions; the other positions them as impulsive and misguided. Mullainathan and Shafir

present an alternative and more realistic perspective, that the poor are neither completely rational nor irrational. Just like everybody else they make good decisions some of the time, and rash or uninformed decisions at other times. They are subject to the same psychological biases as the wealthy; the main difference is that these behaviors have a more profound effect on the poor because of their narrower margins of error due to their adverse financial condition. For example, if a poor individual makes a mistake, it may lead to their telephone being cut off (or in a developing country, to not having sufficient funds for buying “load” on their cellphone). Not having telephone access may then lead to a problem at work, in which the worker is unavailable to call in. This then leads to loss of income, which then leads to further problems. A wealthier individual may have made the same initial cognitive error, but it simply did not reverberate and escalate to further problems in that person’s life.

Putting psychology into conversation with economics changes the way we think about some basic relationships. Despite a preponderance of arguments for lowered transaction costs as a panacea to savings policy (Robinson, 2001), for example, behavioral theories suggest that higher transaction costs can, in some cases, be welfare enhancing. The problem lies with limited commitment problems, in which individuals want to make certain future decisions but face constraints from their own weaknesses, or constraints from others. Beginning with Strotz (1955) and Phelps and Pollak (1968), theoretical models have been put forth that describe outcomes that arise when there are inconsistencies between current temporal tradeoffs and future temporal tradeoffs. These models often incorporate hyperbolic or quasihyperbolic preferences (Ainslie, 1992; Frederick, Loewenstein, & O’Donoghue, 2001; Laibson, 1997; O’Donoghue & Rabin, 1999), theories of temptation (Gul & Pesendorfer, 2001, 2004), or dual-self models of self-control (Fudenberg & Levine, 2005) to generate this prediction. They all share one consistent prediction: individuals should exhibit a preference for restricting their future choice set, and being able to do so will increase their *ex ante* welfare. Naturally, this requires that individuals are self-aware enough to volunteer for such restrictions (much like Odysseus tying himself to the mast to avoid the tempting song of the sirens).

Laibson (1997) looks at decisions made by a consumer with access to illiquid assets—these are assets that generate substantial benefits in the long run, but no immediate benefits (“golden eggs”). The theory suggests that financial innovation, which increases liquidity and reduces implicit commitment opportunities, could have led to the lower savings rates in the US. The model implies that financial market innovation may reduce welfare by providing “too much” liquidity. We revisit the implications of this work below when we discuss technological innovation that lower transaction costs for savings.

Evidence of such preferences is often difficult to show, but a growing literature is demonstrating that demand exists for such restrictions, and in some cases firms are

responding by offering such services. In the United States, Christmas Clubs, popular in the early twentieth century, committed individuals to a schedule of deposits and limited withdrawals, typically with zero interest. In more recent years, defined contribution plans, housing mortgages, and tax overwithholding now play this role for many people in developed economies (Laibson, 1997). For example, on the Earned Income Tax Credit, many individuals do not take advances, effectively an interest-free loan to the government, potentially as a costly commitment device to save (see Jones, 2009).

In developing countries, informal institutions have arguably played this role for years. Many theories exist to explain the presence and structure of rotating savings and credit organizations (ROSCAs), but one commonly held belief is that they provide a form of commitment from your future self (Gugerty, 2007) or from your spouse or extended family (Anderson & Baland, 2002). In Gugerty's work from Kenya (2007), qualitative and quantitative evidence from 70 ROSCAs in rural Kenya was consistent with a self-control commitment story: nearly 60% of the ROSCAs had an explicit spending agreement, for which members were required to identify in advance their purpose for the pot of funds when their turn is up. Members would verify each other's purchases to ensure they adhered to the intended purpose (though most used the money for more than one purchase). Drawing causal conclusions from the data is difficult because in practice new members had limited choice in selecting a ROSCA. While spending-agreement ROSCAs show higher savings rates, they also tend to be composed of wealthier members, often with formal-sector income. Much of the qualitative evidence revealed individuals discussing their inability to save alone. Married women were *not* more likely to participate than nonmarried, nor were women whose husbands lived on their compound more likely to join. This suggests that the save alone reference is fundamentally about one's own ability to save, or about claims made by nonspouse family or neighbors.

ROSCAs also have been shown to be spousal control tools, not just self-control tools. In Anderson and Baland (2002), we learn from Kenya (although a more urban setting than the Gugerty paper discussed earlier) that women with some but not all of the household income are more likely to participate in a ROSCA than those with all or none. The paper works through the following model: it assumes a husband and a wife differ only in their preference for an indivisible good, for which the wife values the good, and the husband does not. Assuming that the husband will respect the ROSCA institution (i.e., a man cannot punish a woman for joining the institution, presumably due to a societal norm), then women who need to extract funds from the household in order to save up for an indivisible good will find the need to join ROSCAs. Women who earn all of the income presumably have power, and thus under this model, no need to join a ROSCA.¹⁰ Women who have none of the household income, on the other hand, have no control over any income flows, and likewise are unable to join ROSCAs because they are unable to commit to any future deposits.

This parabolic relationship is exactly what Anderson and Baland find when trying to predict which type of wives join ROSCAs. Naturally, these two stories are not mutually exclusive, and commitment devices in general can work for very different reasons. Understanding the relevant importance of these models may have important policy implications, such as how to design marketing and account access rules.

The need for commitment from others is not limited to spouses. Platteau (2004) examines the way in which egalitarian norms may inhibit personal savings and thus growth. In some societies successful individuals are called upon to transfer some of their own wealth (directly and indirectly through transfers and favors) to benefit poorer community members and kin. In cases where the wealthy do not consider these to be legitimate contributions, this even serves as a disincentive to work hard. These individuals may also resort to a number of alternate strategies to hold their wealth that involve high transaction costs, in an effort to reduce demands on their income and wealth. These behaviors result in inefficiency that inhibits economic progress and reduces saving levels.

Whereas these studies argue that indigenous institutions have been formed in order to satisfy a demand for commitment, they leave open the question of whether stronger, more formalized commitment devices could succeed in both attracting the right individuals and helping them to adhere to their stated preferences. In the context of Thaler and Sunstein (2008), formal institutions can “nudge” individuals via product framing and design toward decisions they claim to want. For example, Ashraf et al. (2006b) designed a commitment-savings product, called SEED, that provided clients with a commitment to not withdraw their funds until a goal was reached. SEED clients voluntarily restricted their right to withdraw any funds in their own accounts until they reached a self-specified goal. Clients could opt to restrict withdrawals until a specified date (e.g., in a month when school fees were due), or until a specified savings amount was reached (e.g., a certain amount of money for a new roof). The clients had complete flexibility to choose which of these restrictions they would like on their account. However, once the decision was made it could not be changed, and SEED clients could not withdraw funds from the account until they met their chosen goal amount or date.

To evaluate the impact of this new product the bank implemented a randomized control trial where it assigned individuals to either receive an offer to open the SEED account or not, or a third group which received a marketing pitch about the importance of setting goals for savings (but no access to a new commitment savings account). Among those offered the account, 28% opened one, and, importantly, those who exhibited inconsistent time preferences in survey questions about hypothetical alternatives were more likely to open an account.¹¹ After 12 months, average balances increased by 80% in the treatment group (i.e., those who got the SEED offer) compared to the control group. The account offer was also associated with a significant

increase in women's decision-making power within the household (as measured both qualitatively by asking who has power over certain decisions, and more objectively by observing the gender-bias of durable goods purchased) (Ashraf, Karlan, & Yin, 2010).

A question remains, however, to what extent such commitments are about binding one's behavior, or are in fact merely about creating structure. The previous section described how those wanting to save might opt for credit simply because it provides a defined schedule for making regular deposits. Karlan, McConnell, Mullainathan, and Zinman (2009) test the effects of simply making savings more salient by sending clients simple reminders to make deposits. They find even with no commitment, the reminders can be successful in increasing savings rates (by 6%) and helping clients meet savings goals (a 3% increase in the likelihood of reaching one's goal). Similar positive impacts on savings were found by deposit collection services tested in Ashraf, Karlan, and Yin (2006a), as well as Dupas and Robinson (2009).

These ideas are gaining acceptance within academia. They are also having impacts in practice. Grameen Bank of Bangladesh, for example, has launched a successful commitment savings device, a "pension" product that requires monthly deposits in fixed amounts and returns savings (with interest) after 5 or 10 years (depending on the product). The savings account is a "pension" in name only, and while it is used to pay for old age, households also use it to accumulate for housing improvements, wedding expenses, migration and the like (Collins et al. 2009, chap. 6).

6.2.2 Planning and financial literacy

A preponderance of evidence shows that financial illiteracy is prevalent around the world, and is correlated with low savings. However, this begs the policy question of whether interventions intended to increase financial literacy can in fact lead to changes in behavior of importance. Evidence is clear that people, everywhere, are financially "illiterate" by many definitions, both described and measured by lack of basic numeracy (e.g., simple compounding), financial knowledge (familiarity with financial products, including credit, savings, and mortgages), and financial planning (e.g., saving for retirement). Lusardi (2007) shows that financial illiteracy even in the US population is widespread, and particularly acute for specific demographic groups, such as those with low education, women, African-Americans, and Hispanics. Moreover, close to half of older workers do not know which type of pensions they have and the large majority of workers know little about the rules governing Social Security benefits. Notwithstanding the low levels of literacy that many individuals display, very few rely on the help of experts or financial advisors to make saving and investment decisions.

Naturally, the mere correlation of financial illiteracy with outcomes such as savings decisions does not imply that financial illiteracy is the cause of the low savings. Several endogeneity issues could be at play, including the simplest, that omitted variables such as motivation to succeed are the true cause of both financial illiteracy and low savings.

Or, reverse causality: if one is incapable of saving in quantity then one would be unlikely to invest in knowledge of savings vehicles. Examining the causal impact of financial education programs requires effective methods for establishing the counterfactual.

Two studies on financial literacy examine this very issue and draw strikingly different conclusions. This demonstrates the difficulty of establishing attribution in evaluation of public policies. The first paper, Bernheim, Garrett, and Maki (2001), uses variation in state-mandated financial education to measure the treatment effect of financial literacy training on household savings. Using data from Merrill Lynch, and a telephone survey of 3500, the authors employ a difference-in-difference approach and assume that timing of the introduction of state-mandated financial education is exogenous. They conclude that the mandates were effective in teaching basic financial literacy and led to a 1.5 percentage point higher saving rate. However, it turns out there is evidence against their key identification assumption. Cole and Shastry (2008) extend this study with more data and conclude that the Bernheim et al. result was spurious. Census data, and thus a larger dataset, allow for the inclusion of state fixed effects to control for unobserved, time-invariant heterogeneity in savings behavior across states, as well as nonparametric identification of the treatment effect itself (rather than a linear measure of years-since-mandate-began employed by Bernheim, Garrett, and Maki). Once these three enhancements are implemented, all treatment effects fall to a precisely estimated zero, thus both eliminating the conclusion that financial literacy as implemented under this program had any effect, and demonstrating a key endogeneity issue that plagues this literature (that rollout of programs is responsive to demand and thus extensive work must be done to create convincing counterfactuals).

Savings, as we have seen, is important to poor households, important enough that they are willing to pay for the service. This commitment to save has led some practitioners to conclude that savings, rather than credit, is the more practical strategy to promote, especially for the very poor. This is surely true for some households, but is strongly contested by the evidence in Collins et al. (2009).

There are some settings where the poor do not save, even where it is clearly advantageous for them to do so. Why have not more households saved their way out of credit constraints? Such puzzles suggest that large-scale savings promotion may require a more sophisticated strategy than imagined initially: it will involve better research to understand why the poor do not save when they could, and which strategies can help overcome these barriers.

A related puzzle concerns entrepreneurs who borrow persistently, not for a one-time business expansion but for routine working capital. This is very expensive. In extreme examples vegetable vendors in India are known to borrow small sums each day to purchase vegetables, repaying each afternoon from their daily sales. They pay rates as high as 10% per day. Among a sample of vendors in Chennai, 50% claimed

to have been engaging in this type of borrowing for at least 10 years! What is so interesting about such borrowing cycles is that it is easy to show that in principle by saving merely one rupee (a few cents) each day (and borrowing that much less) the vendor could be debt free, and able to finance her own working capital, in just 50 days. From there her returns are enormous: her daily profit margin is boosted by the 10% she was paying to the moneylender. So why do the vendors persist in borrowing at such rates? The answer remains unclear but financial education seems a good place to start. If they understood the true cost of their borrowing, they might well be convinced to save.

In India and the Philippines, Karlan, Mullainathan, and Zinman (2008) explore the impact of debt payoff and financial education on the persistence of high interest debt. These experiments address the aforementioned endogeneity concerns by randomly offering indebted vendors different experimental treatments. Specifically, some vendors are offered an unexpected debt payoff, others financial literacy training emphasizing the utility of savings to finance business expansion and the cost of debt (the training itself was a brief but focused 30-min session), and a third group are offered both debt payoff and financial training. A fourth group serves as a control group and thus received no experimental treatments. The evidence from the Philippines suggests that debt payoff can have immediate but then dissipating effects on vendors' reliance on high-interest loans: vendors who received one of the payoff treatments were 31 percentage points less likely to have money lender debt almost 1 month after the payoff and 17 percentage points less likely to have taken a high-interest (defined as greater than 5% per month) loan after 3 months. Preliminary results showed that the effect dissipates over time, however, as individuals gradually go back into debt.

Of course a measured reduction in household consumption is not necessarily a good thing. More research will be required to look deeper into the specifics of how household consumption decisions change as a result of greater financial literacy. Are households choosing to save rather than spend excess income? Are they making immediate sacrifices in the interest of financing longer term business goals? Another question that remains unanswered is whether the impacts observed in the debt payoff and financial literacy training experiment can be multiplied by providing high-interest borrowers with an effective savings mechanism. These and other questions require further research to disentangle how product offerings can change the cash management and thus savings and consumption decisions of the poor.

6.2.3 Pricing

Just as we discussed earlier with respect to credit, many have assumed that because the poor are willing to actually pay to save, this implies that their demand is inelastic with regard to price. Naturally one does not imply the other, since willingness to save at zero or negative real interest rates simply means that at that price demand is positive, but at higher prices of course demand could still be higher. Hirschland and Owens (2005)

provide a useful overview of a typical practitioner's perspective on the considerations for how to set price for savings, including competitive analysis and cost drivers. Clear information on elasticities, however, would help tremendously.

In Karlan et al. (2008), just as in earlier work on credit elasticities, the authors conducted a series of field experiments in which the interest rate on the savings account was randomized, and collected data on how this influenced the decision to open an account as well as the volume of savings held in the account. Specifically, the treatments tested are high, low, and "reward," in which the high rate was 1.0% per annum higher than the low, and the bonus was also 1.0% but only awarded if the individuals savings goal amount was reached by their goal date. In the second experiment, the design was slightly modified, and the increase in interest rates and bonus was 1.5% per annum instead of 1.0%. While they find that the product take-up rates among the high-interest group are higher than that of the low-interest group in both experiments, and both lead to 3% higher savings balances, still after 7000 offers (with an average take-up rate of 23% and average balance of about US\$10), neither of those results are statistically significant.

6.2.4 Loss aversion and mental accounting

The concept of loss aversion is well established in our understanding of human behavior but not well adopted in the design of savings products. The 2002 Nobel Prize in Economics was awarded to Daniel Kahneman for, among other ideas, the simple demonstration of this. The canonical classroom experiment involves mugs (Kahneman, Knetsch, & Thaler, 1990). Half of the students are randomly selected in a lottery to each get a mug. Everyone in the class is then given the opportunity to trade. In an "efficient" world of no loss aversion, the half with the highest valuations should end up with the mugs, irrespective of whether or not they "won" the lottery. But winning the lottery changes one's valuation of the mug. Why? Because once the mug is won, the reference point is shifted. Giving up the mug is now a loss, whereas for everyone else acquiring the mug is a gain. Losses loom larger than gains, and few people trade. Those who won the mugs mostly fail to find someone willing to pay them enough to make them part with their moments-ago-won mug. Instead, if the professor asked for students to announce their valuation *before* receiving the mugs, then handed them out and made everyone adhere to their stated valuations, one would expect, on average, half of the mugs to be traded.

Now let us apply loss aversion theory to the decision to make a deposit into a savings account. What is a savings deposit in a mental accounting system? It is trading off some current (salient) consumption for some future (nebulous) consumption. Ignore the timing for a moment (more on that below), and what does one have? A sure loss, in exchange for an unclear gain. Thus loss aversion may drive the individual to consume, rather than save.

How does one tackle this? Can product design effectively convert the savings deposit into a gain? Some of the problem is driven by the vagueness of the future gain. Perhaps making the future gain more salient, one can convert the savings deposit into, at a minimum, a “neutral” and, at best, a “gain.” Ashraf et al. (2006a, 2006b), in the study in the Philippines discussed earlier, employed a “placebo” treatment involving marketing only, in which bank marketing agents visited homes of prior and current clients to encourage them to consider setting savings goals. This treatment was intended as a placebo, in order to make sure the treatment effect from the commitment savings product was due to the mechanism design of the product, and not the labeling and goal setting promotion. In fact, the impact of the goal-setting treatment (referred to as “marketing” treatment) was positive, although not statistically significant. Labeling can also matter. In developed countries, we have seen accounts “labeled” through marketing (and tax advantages): savings accounts for *education*, savings accounts for *health*, savings accounts for *retirement*. These ideas do not require fancy infrastructure, but rather mere marketing and packaging. Does account labeling allow one to make the later gain from saving more salient in the present, and thus cancel out the loss from foregoing current consumption? This is an open question.

More generally stated, one could imagine many methods for converting the “loss” of the foregone consumption due to making a deposit be converted to a gain. Doing so relies heavily on the *lack* of fungibility of money, a violation of most traditional models in economics. Mental accounting, as put forward by Thaler (1985) provides a framework to interpret such behavior.

6.2.5 Default options

Extensive evidence exists to show that default options matter. This is true in many facets of life, and savings behavior is no exception. Madrian and Shea (2002) document this clearly for individuals making retirement decisions in the USA. There, Madrian and Shea show that setting automatic enrollment as the default participation in retirement plans leads to a 50 percentage-point increase in the likelihood of participation due to automatic enrollment, and similarly large effects are found on the type of investments individuals choose.

A more complex example comes from Thaler and Benartzi (2004). Here, the authors implemented a program called “Save More Tomorrow™” (SMarT) in which individuals weakly commit (by “weakly,” we mean they can reverse this decision at any time) a portion of their *future* salary increases toward retirement savings. The paper cited reports findings based on evidence from the first implementation of the SMarT program, through four raises. Key findings are (1) a high proportion (78%) of those offered the plan joined, (2) the majority (80%) remained in the program through the fourth raise, and (3) the average savings rate increased from 3.5% to 13.6% over 40 months. This program has now been adopted by many corporate retirement plans in the United States. In a

survey of 146 employers in the United States, Hewitt Associates found that 31% of plans have an automatic escalation program, and of the 69% that do not, 42% said they plan to incorporate such a component into their plan (Hewitt, 2007).

Adapting these ideas to developing countries has tremendous promise, and also could provide ample opportunities for learning *why* this works more precisely, as several factors were employed at once in the design of SMarT. First, since upon sign-up, the default future decision would be to save more, action had to be taken to change this decision. Second, by framing the future increase as “coming out of your next pay increase” (even if the next pay raise was merely an inflation adjustment), money illusion may have led more to sign-up than would otherwise have occurred. Third, the method of presentation, and the skills of the advisors, may have influenced the decision of individuals to participate.

6.2.6 Marketing

Often design and discussion about products by academics focus on the terms and structure and risks of a product. Yet, in many cases, the presentation, framing, and promotion style influences the outcomes. For example, in Landry, Lange, List, Price, and Rupp (2006), in a door-to-door marketing of a fundraising appeal, the authors found that the physical attractiveness of the door-to-door salespeople was far more important than the lottery that was being offered to some but not others. Similarly, in a developing country setting, Bertrand et al. (2010) find that adding a photo of a woman to a direct mail solicitation increases the likelihood of borrowing by just as much as dropping the interest rate by about 30%. Al-Bagadi and Cracknell (2005) discuss this more generally in the context of marketing and promotion of savings in developing countries. They argue that microfinance institutions must translate ideas about why potential clients should want to deposit with a particular institution into a message that motivates them to do so. We suggest that future studies that examine product features as discussed earlier do not forget to think about the promotional strategy and ideally incorporate the promotional strategy as an integral part of any study.

6.2.7 Social networks and peer effects

If the poor have limited knowledge of the benefits of saving, then learning about these benefits through social networks could potentially have important impacts on the poor’s savings behavior. Despite its relevance, not much research exists on the role of peer effects in savings decisions. Peer effects are notoriously hard to identify empirically. In many instances individuals’ decisions within a social group are correlated for reasons that have nothing to do with social learning or social imitation. Behavior of individuals may be correlated simply because individuals in the same group have similar unobserved characteristics—for example, a common propensity to save—or share a common environment. Manski (2003, 2005) lays out some of the difficulties involved in identifying peer effects.

In an empirical contribution, Duflo and Saez (2002) investigate the role peer effects play in retirement savings decisions in the United States. They study whether the decisions of employees of a large university to enroll in a university-sponsored Tax Deferred Account plan are affected by the decisions of colleagues in the same department. Instrumenting average participation within peer groups by average salary or tenure and looking at subgroups of peers within departments, Duflo and Saez find evidence that the individual participation decision is influenced by the decision of one's peers.

Assessing the importance of peer effects in savings decisions matters for the design of policy interventions. If innovations in savings behavior spread through social networks, the impacts of financial education efforts can be much larger than the impacts on the financially educated. An initiative that aims to make effective use of learning from peers to promote savings is Oxfam's Saving for Change program, based on a model used in Nepal and now replicated in Africa. Saving for Change groups are informal self-managed saving and credit groups consisting of 25–30 women. Group members meet weekly to save a predetermined amount and the collected funds are used to make loans to group members. Since loans are repaid with interest, the fund grows over time. At the end of the cycle the fund is divided according to each member's share in the savings. The program has been found to rapidly reach considerable numbers of women. The first group in a village is typically trained by an external agent with the members of the first group then forming and training subsequent groups in the village. The role of social networks and peer effects in the spread of these saving and loan groups is the object of ongoing research.

Peer influences need not work merely through one-on-one interaction. For example, as Garon (2004) points out, East Asian states played a key role in inculcating savings habits and thrift among their citizens. Asian values of savings and consumption were created by a regionally adopted model of state promotion of savings, with Japan's colonial presence in these countries acting as one of the major catalysts to these state efforts. The history of the "Japanese Model" includes national campaigns to promote savings, postal savings banks, and a Central Council for Savings Promotion. This model was exported to varying degrees to South Korea, Singapore, and Malaysia. Garon argues that though thrift and savings are not timeless or unique Asian values, many Asian people have come to embrace these as part of their national identities. This begs the question, though: are savings levels cultural, driven by social norms and macrolevel policies, or simply the product of one huge omitted variable, for example, institutions.

6.2.8 Technology

There is a massive effort underway to harness technology to overcome the transaction costs that have prevented the poor from accessing banking services, especially in rural areas. Technology also appears to be revolutionizing the remittance industry, as

technology solutions are making remittances easier and cheaper than ever before. As discussed earlier, remittances historically, and still today in many countries, are expensive as a proportion of money sent, particularly for small amounts. Yet unlike credit there are no information asymmetries or default risks to explain these costs, and hence technology solutions are proving to have big impacts on the costs consumers are offered.

The donor consortium CGAP has launched a campaign to bring mobile banking services to 25 million low-income people by 2012. It is too early to tell what form these services will ultimately take, but right now attention is focused on point-of-sale devices installed at retail agents (e.g., local stores) who can accept deposits or payments, as well as withdrawals; and banking via mobile handsets, enabling customers to transfer funds electronically. The explosion of accessible technology, no doubt, will change whether and how people interact with banks. This may have important unintended consequences. As we discussed earlier, liquidity can actually be undesirable for some. Could electronic banking lead to further problems with self-control and spousal, familial, or community pressure? If hidden savings are not offered in conjunction with increased access to funds through technology, such pressures could lead to worse outcomes. As transaction costs drop, this will become an even more important area for future research and innovation.

Furthermore, much of the advent of microfinance has been through innovations that remove layers of information asymmetries so that banks can profitably lend to the poor (and for microinsurance, so that they can insure the poor). Removing the human touch from banking may have unintended adverse consequences, in that individuals will reach for the “easy” technology but in the process lose the interaction with the credit officer that is necessary for the financial services firms to establish relationships and lend and insure. The personal relationship with a banker may reduce information asymmetries both by allowing the bank to have more information for screening, and also by reducing moral hazard, by instilling a sentiment of reciprocity or personal loyalty in the mind of the borrower, making them less likely to engage in moral hazard.

Thus, technological innovations in the user-interface are promising for radical reductions in cost of accessing savings services, but we need to better understand how behavior will change as a result of this easier access. In a quasieperiment from the field, albeit not in this context, consumption on a particular good (sorting laundry into two washing machines or pooling into one) increased by 50% as the payment mechanism shifted from cash to prepaid cards (Soman, 2003). How will mobile “minutes” be treated, as cash to save or money to spend? This simple mental categorization could have serious implications for the long-term impact on consumption and savings from such technologies.

Early experiences with branchless banking are already showing clients using the new technologies largely for payment transactions, rather than savings or credit. To some extent this may be the result of marketing campaigns by mobile operators which focus on the transfer of money rather than the storage of money—apparently because the operators are concerned about appearing to market themselves as banks and thus attracting notice by regulators. Despite the great potential for branchless banking to expand financial access, thus far mobile banking customers in developing countries have been wealthier customers in urban areas (Ivatury & Mas, 2008).

6.3 Impacts of saving

The impact of savings programs can be difficult to measure, both because savings is hard to capture in survey data and because it is hard to isolate savings from other financial services: few institutions offer *only* savings. Three factors particularly complicate measurement: size, timing, and diffusion. Unlike credit inflows, which can be sizable relative to household income, savings flows can be quite small, and balances accumulate slowly. Also, the timing of the change in behavior and outcome is less clear. For households, savings develops slowly through a small reduction in consumption over time, with a large inflow later. At some point the household will have built up enough savings to protect themselves from shocks (like sickness or unemployment), to pay school fees, or to start a business. But when is that point? It may not be a simple question of waiting for savings to accrue: household cash flows may vary over time. Researchers need to measure savings balances at multiple points in time, often over several years.

The last issue, diffusion, is perhaps the most important. Poor households save in all sorts of different ways. In addition to a formal savings institution they may save cash at home, with deposit collectors, within a savings club (such as a ROSCA), or by lending to family members. They may also save in noncash assets such as jewelry or livestock. An evaluation that failed to capture these different savings vehicles could underestimate savings rates, potentially leading to incorrect conclusions about the impact of the program. Or, an evaluation which focused on just one savings channel may incorrectly conclude that net savings goes up, whereas in fact the only impact was a shift from one type of savings to another.

As with credit, it is problematic to compare savers to nonsavers. Savers may be better educated or have more disposable income. They may be better planners, or more risk averse. There might also be reverse causality: healthy people might be able to earn more, or spend less on hospital visits, increasing savings balances. Given these issues it is unsurprising that there are few rigorous evaluations of savings. Burgess and Pande (2005), described in the previous section, solve the identification problem, showing that financial access reduces poverty, but they are unable to separate the effect of savings from credit.

Aportela (1999) evaluates the impact of a government savings program in Mexico, Patronato del Ahorro Nacional (Pahnal). In 1993 Pahnal decided to expand its savings operations through the postal office network. By using the postal offices, Pahnal was able to locate close to many people who were otherwise inaccessible, but without the fixed cost of opening a vast network of offices. As part of this expansion, Pahnal offered two savings options: (1) a fixed-term instrument (12, 24, or 36 months) with compulsory monthly deposits of 5 dollars, and inability to withdraw until maturity, and (2) a liquid savings account, with a minimum balance of just over 5 dollars, with no fees but lower interest than the fixed term. Aportela uses the partial expansion of the program to compare the change in outcomes for those in communities that received the Pahnal expansion to the change in outcomes for those in communities that did not receive the Pahnal expansion. The analysis uses data from the 1992 and 1994 Mexican Household's Income and Expenditure surveys. The advantage of these data, rather than using client data, is that there is no individual selection bias. Given that the program chose its own expansion path, however, it is critical to verify that pre-existing savings rates or poverty rates are not correlated with the treatment communities. Aportela finds no correlation, nor any evidence of an operational plan that could have confounded the analysis (e.g., an initial expansion into districts that had expressed high demand for the program). The expansion appears to have more closely followed Pahnal's operational convenience, that is, proximity to its prior branches. Hence an issue remains as to whether Pahnal's prior branches were located strategically such as to create a preexisting trend and thus selection bias.

The impact on formal-sector savings balances was noticeable: the average savings rate increased by 3–5 percentage points. For low-income individuals it was even higher: an increase of 5.7–8 percentage points. The emphasis of the analysis in the study is on formal sector savings volume, broken down by different income levels. The attempt to measure aggregate savings, including informal savings, fails to reach decisive conclusions because of a lack of statistical precision and data limitations: the analysis is unable to rule out substitution from other savings vehicles. Moreover, it does not allow one to isolate which of the two products, or within either product which particular feature, led to the impact. Whether the increase in formal savings represents an increase in net savings or a shift from informal to formal savings it can be considered a positive impact in that the formalization of savings implies safer savings for individuals.

Impact from specific product trials is more limited. The evidence discussed earlier, from Ashraf et al. (2010), showed a potentially good relationship between savings and empowerment. The commitment savings product there led to a significant increase in women's decision-making power within the household (measured by an index of decision-making authority over various types of purchases, as well as family planning and children's education), and an increase of the purchase of female-oriented durable

goods. The impacts were particularly strong for women who have below median decision-making power in the baseline data.

Dupas and Robinson (2009) worked with a community bank in rural Kenya to provide incentives to open a savings account to randomly selected entrepreneurs, for whom the researchers paid the fee to open the account and provided the minimum account balance. The control group received no incentives but were not barred from opening an account. The incentives were strong enough that 89% of the treatment group opened an account, while only three individuals in control group did so. The researchers find remarkable impacts despite substantial transaction fees charged by the bank (\$0.50 or more) and the fact that many never used the account after opening it. In contrast to the Karlan and Zinman (2009a) study of the impact of credit in the Philippines and the de Mel et al. (2008b) study of returns to capital, here the impacts are found only among female entrepreneurs. Four months after opening the account women show 40% increases in productive investment, and after 6 months daily consumption is approximately 40% higher than the control group. In this study however women have different enterprises than men, and hence the gender difference may be properly interpreted as a difference generated by occupational choice. Further research and expansion can help understand this important result.

7. RISK MANAGEMENT AND INSURANCE

To be poor in most of the developing world is also to be disproportionately vulnerable to risk (Dercon, 2004a; Morduch, 1994). Among sources of vulnerability are the high correlation of poverty and ill health (e.g., Case, Lubotsky, & Paxson, 2002; Dercon & Hoddinott, 2004), the riskiness of agricultural occupations (Dercon, 2004b), employment instability within the informal sector (Lund & Nicholson, 2004), and the broad insecurities that arise from weak legal protections (e.g., Field, 2007).

One accounting of links between poverty and vulnerability is offered by Dercon (2004b, Table 1.1), who reports on a survey of rural households in Ethiopia who were asked to look backward on the two decade span 1974–1994. Most households reported having suffered seriously due to harvest failure brought on by drought, flooding, frost, or pests (78%). Others suffered due to illnesses and death, both of family members and of livestock (about 40%). Like the harvest failures, the health losses are potentially insurable. Others serious losses suffered, however, arose from forces that would be near impossible to insure with traditional insurance products, notably policy shocks resulting in forced labor, new taxes, and migration bans.

Collins et al. (2009, Table 3.1) similarly report on major losses suffered by small samples of poor and “near-poor” households during a single year in rural and urban Bangladesh, India, and South Africa. Half of the Bangladesh sample suffered due to

serious injury and illness during the study year, as did 42% of the India sample. These health crises were accompanied by losses to income and property. For example, in Bangladesh, the urban individuals suffered from slum clearing by police and construction workers, and in rural India individuals suffered from a particularly bad harvest. When researchers in Bangladesh returned to their three urban sites in 2005, 5 years after the original year-long survey, all three sites had been wholly or partly destroyed.

Fully addressing such vulnerability requires policymakers to use their powers to reduce insecurities rather than exacerbate them, and, with an active stance, to expand social security programs, improve health infrastructure and disaster management, and create stronger property rights. Collins et al. (2009), for example, find that in South Africa health losses are substantially mitigated by the presence of free public clinics and the generosity of the pension system, based around a system of monthly government grants that gives households flexibility to cope with health losses (Case & Deaton, 1998). Implementing South African-style safety nets universally, however, would surely run up against budget and administrative constraints in most poor countries. While public sector solutions may be part of the picture, households' main forms of risk coping will no doubt continue to be private.

It is not surprising then that providing low-income households with access to reliable and reasonably priced insurance mechanisms is increasingly taken to constitute a key part of "inclusive" financial sectors. The field of "microinsurance" (a term that encompasses insurance products targeted to poor and low-income consumers) holds promise, but the field is young and no approaches have emerged so far that offer breakthroughs akin to the original group-lending innovations that ignited the global explosion of microcredit (Morduch, 2006). Basic financial products like loans and savings accounts will remain critical devices for risk management in poor communities, just as they are in richer communities.

7.0.1 The persistence of shocks

The project of expanding insurance cover and coping mechanisms is made more urgent to the extent that temporary shocks translate into long-term losses. Collins et al. (2009) provide a string of stories of households hit by serious illness. Over time most of the households end up depleting their financial assets in their attempts to pay for medicines and doctors, and the households' asset depletion translates into reduced earning potential over the long term. The most severe losses occur when the shocks themselves play out over time, as with a worsening case of tuberculosis that requires repeated visits to doctors and extended courses of medicine.¹²

One sense of these dynamics emerges from papers that relate risk to health outcomes. Dercon (2004b) reports on a broad group of studies that link income shocks to health outcomes. Rose (1999), for example, correlates the incidence of bad rainfall realizations in rural India and increased infant mortality rates. She finds that the inability to cope with the temporary loss of income leads to choices that directly harm children,

a finding that emerges in a sample of landless households (i.e., those with the most limited ability to self-insure) but not in the sample of households with substantial assets. The link between vulnerability to temporary income shocks and increased rates of child mortality is the most extreme example. Other examples, drawing on evidence mainly from poor populations in South Asia and sub-Saharan Africa, describe links between vulnerability to income swings and outcomes that fall short of death but which nevertheless generate lasting deprivations; they include low school attendance (Jacoby & Skoufias, 1997); increased child labor (Beegle, Dehejia, & Gatti, 2003); reduced physical stature of children (Alderman, Hoddinott, & Kinsey, 2006; Foster, 1995); and diminished school performance of young children (Alderman et al., 2006; Foster, 1995). As with Rose's (1999) study, these are not generic findings but hold chiefly for households with few assets, either of land or livestock, and limited means to self-insure.

The link between vulnerability and the profitability of enterprises is a staple of the theory of risk aversion (i.e., the notion that investors and entrepreneurs trade off average financial returns for a lower variance of expected returns), but it is not yet well established in practice. The most promising avenue will likely involve studies that relate access to insurance to measures of productivity (e.g., the profitability of household business).

Dercon and Hoddinott (2004) sum up their evidence from Zimbabwe and Ethiopia by stressing the way that temporary setbacks turn into permanent deprivations: they find that the inability to cope with transitory shocks from droughts and other serious crises has long-term consequences, especially for children, for whom reductions in stature and schooling outcomes tend to diminish future employment prospects and productivity. Drawing on the empirical regularity that "taller (and better educated) women have, on average, taller (and healthier) children," they conclude that "the impact of these transitory shocks may well be felt for several generations." (Dercon & Hoddinott, 2004, p. 134).

Jalan and Ravallion (2004) tackle this question of transience and permanence of effects in the structural estimation of income processes using a 6-year panel (1985–1990) of Chinese households in Guangdong, Guangxi, Guizhou, and Yunnan provinces. They investigate whether the error structure in income equations is consistent with poverty traps arising from temporary income shocks. They model the lagged dependent variable as a cubic function in a dynamic panel-data income model, looking for evidence of nonconvexities in patterns of autocorrelated income. While income is found to be autocorrelated, such that a bad shock this year makes another bad shock more likely next year, Jalan and Ravallion do not find generalizable evidence that temporary shocks create poverty traps. They do, though, find large differences in the speed of recovery from shocks, with poor households taking far longer to bounce back than their better-off neighbors. The picture, as with the evidence earlier, is one in which risk is widespread and consequences are long lasting for those who lack the means to cope.

7.1 Why insurance markets fail

The findings earlier help to show why risk management is important for poor households, yet the studies also point to a fundamental challenge: both the supply and the demand of private insurance tend to be low in low-income communities, especially relative to the take-up of new credit and saving products.

The supply-side difficulties are stubborn, starting with the well-known problems of adverse selection and moral hazard (e.g., Besley, 1995). As the classic papers of Arrow (1963) and Pauly (1968) detail, information asymmetries can drive failures in the commercial provision of insurance. Systems of deductibles (households pay the cost of initial losses before insurers begin paying for the bill) and coinsurance (households cover a fraction of total expenses) can help, but practical difficulties remain when insuring outcomes that are heavily effort dependent, such as crop failure and livestock health. When insurers cannot observe effort, nor observe the inherent riskiness of customers, contracts generally yield suboptimal outcomes. In light of the theory, it is little surprise that there is so little profitable, large-scale commercial insurance coverage in low-income communities. Crop insurance accounts for a particularly notable gap given how important agriculture is in much of the developing world. Information problems make crop insurance a relatively less enticing product line for most commercial insurers (Morduch, 2006), but political imperatives make supporting farmers a priority policy initiative (Carter, Galarza, & Boucher, 2007). The combination leads to a landscape presently dominated by subsidy-dependent efforts.

In a theoretical treatment of the general insurance problem, Banerjee and Newman (1993) examine the market for risk from the viewpoint of the poor. While the Banerjee-Newman model focuses on broad relationships between risk-bearing and the income distribution, an important insight emerges on the working of insurance markets. The logic of using deductibles and coinsurance to improve insurance markets hinges on exposing customers to enough risk such that they have incentives to work hard to limit bad outcomes. Providing complete coverage undermines such incentives. Banerjee and Newman (1993) show that in this setup, it is the poor, rather than the rich, who—all else the same—will receive closer-to-complete insurance coverage. This is because poor households, who by definition live close to subsistence levels, are assumed to be relatively sensitive to variations in consumption levels when compared to richer households—an observation captured by the common assumption of declining absolute risk aversion (i.e., that as people get richer, they tolerate more risk). A given-sized deductible or coinsurance rate will thus deliver a greater dose of beneficial incentives for poor households than richer households. Equivalently, incentives can be delivered through the use of smaller deductibles and lower coinsurance levels when transacting with poor households than with rich households. Optimal contracts for poor households thus, in principle, deliver a greater extent of insurance cover relative to contracts for richer households.

Banerjee (2004) cites the literature on informal village insurance to argue that the poor may in fact be quite well insured (for an overview, see Deaton, 1997; Morduch, 2006; Empirical studies include Dubois, 2000; Fafchamps & Lund, 2003; Grimard, 1997; Jalan & Ravallion, 1999; Ligon, Thomas, & Worrall, 2002; Morduch, 2004; Townsend, 1994, 1995; Udry, 1994). But our review of the literature at this juncture suggests that the view is too optimistic, and that poor households remain substantially exposed to risk. Still, the polar extreme—in which household's simply consume what they earn without smoothing ups and downs—is also a poor characterization. The literature so far shows that households do manage to self-insure and arrange collectively to share risk.

As Townsend (1994) suggests, there are several potential ways in which villagers might deal with risk on an informal basis, including: (1) diversification of a given farmer's landholdings into various spatially separated plots and into various crops, (2) storage of grain from one year to the next, (3) purchases and sales of assets such as bullocks and land, (4) borrowing from village lenders or itinerant merchants and borrowing/lending more generally, and (5) gifts and transfers in family networks (Townsend, 1994, pp. 539–540).

In any given year villagers might well use more than one or even all of the mechanisms on this list. As Townsend notes, each of the mechanisms is by itself nontrivial to evaluate. Using survey data from high-risk villages in semiarid India, Townsend (1994) looks holistically at the extent to which consumption among individual households covaries with average consumption in their villages. That is, through some combination of these or other mechanisms, do villages share risk? If markets for risk are complete, consumption should move together. Townsend finds that risk sharing, while not perfect, is quite good. In particular, credit and gifts are important in smoothing consumption; in some cases, the volume of loans and gifts exceeds average consumption (although Morduch, 1994 finds weaker evidence of extensive gift giving). Townsend goes on to look for evidence of regional risk-sharing but finds the data inconclusive.

Udry (1994) focuses in on credit as source of consumption smoothing in the absence of insurance markets. By collecting a dataset of credit transactions in four villages in northern Nigeria, Udry provides insight into the specific nature of the mechanism behind the observed outcomes. The data are interesting even in their basic description: while over 75% of households lent money and nearly as many borrowed, the area was served neither by a formal financial institution nor specialized money-lender. Nearly all loans (97%) were between neighbors and relatives. Loans were made without witnesses or even written records, and though the repayment amount was negotiated, explicit interest rates were never discussed. Only 3% of the loans were backed by collateral. Whether borrower or lender, 82% of those surveyed were able to enumerate the farm activities of the party on the other side of the transaction. It is

this flow of information, Udry argues, that provides the basis for risk sharing. Most tellingly, payments are shown to respond to the financial circumstances of the *lending* household (through shorter payment periods and/or higher interest payments), a situation for which there is no provision in formal finance. (The reverse is seen as well: adverse shocks among borrowing households are met with lower payments over longer terms.) Nonetheless, Udry finds that the mutual insurance system provided by this arrangement is not sufficiently complete to insure all idiosyncratic risk faced by households.

Fafchamps and Lund (2003) build on Udry's work by also considering gifts and transfers made outside of the context of loans, as well as savings and labor market participation. These steps have the benefit of increasing the scope for capturing transfers and responses to shocks, and they generalize findings to the village population, rather than just borrowers and lenders (though Udry finds they comprise the majority of the population in his sample). The authors' panel data collected from the Philippines show the majority (71% in value) of credit transactions are conducted between relatives and neighbors. More than 80% of loans are made within the same village, with virtually the rest occurring between neighboring villages. As in Udry's data, lenders and borrowers are very familiar with each other's activities: more than 85% of respondents were able to provide a complete accounting of the wealth holdings and demographic characteristics of their loan partners.

Gift giving is universal in the sample households: all households gave or received at least one gift, and 94% did so in each of the three survey rounds. 92% of households borrowed and 61% lent money, and 80% of informal loans carry no interest. For both gifts and informal loans the most common use of the funds is immediate consumption (a finding that squares with recent evidence collected by Johnston and Morduch (2008) showing that on average half of microfinance loans issued to a sample of Bank Rakyat Indonesia customers were used for nonbusiness purposes). Loans respond to shocks (as in Udry, 1994), and gifts appear to as well, but with only borderline significance ($p = 0.13$).

Fafchamps and Lund (2003) reject a village-level full insurance model in favor of one that works through networks of friends and relatives. Not all shocks are insured, and households respond to shocks in part by drawing upon financial savings (not livestock or crops), but not by increasing labor. As with Ligon et al. (2002), who revisit the data from Townsend (1994), Fafchamps and Lund find that the fit of their model is improved by taking account of limited commitment—that is, through a model of insurance predicated on the notion that contracts in the informal sector are not enforceable and any party can withdraw from the scheme at any time.

Focusing specifically on illness, a major source of shocks for poor households, Gertler and Gruber (2002) find considerable exposure to risk in Indonesia. A key distinction drawn by Gertler and Gruber is the size of the shock: even if they take

Townsend's (1994) results as given and assume that minor health-related fluctuations in consumption can be smoothed over time, they argue that there may be less frequent, more serious health shocks that households will be unable to insure against through informal mechanisms. They test for this by looking at consumption patterns in a dataset incorporating a measure of the severity of illnesses. They find that while households are able to fully insure minor illnesses (those that do not limit physical functioning), they are only able to insure 71% of the economic costs (the cost of health care plus lost income) of moderate illnesses, and 38% of the cost of illnesses that severely limit physical functioning. The authors suggest the introduction of informal disability insurance as a potential remedy.¹³

Much of the earlier literature focuses on the village as a “natural” insurance unit, and the question posed by researchers centers on the degree to which villagers set up arrangements by which idiosyncratic risk is insured collectively. The structure of estimating equations captures the focus on idiosyncratic risk defined as income variability around the mean village income in a given year—that is, doing better or worse than your neighbors. It is a helpful starting place, but the tests remain mute on the ability to cope with the kinds of major regional risks described by Dercon (2004a): the broad devastation brought by wide-scale shocks like droughts, floods, and economic crises.

In addition, villages are in many ways *not* natural insurance groups, especially relative to families and kinship groups that often extend geographically (Morduch, 2004; Munshi & Rosenzweig, 2007; Rosenzweig, 1988). Debate about the role of villages as risk-sharing collectivities goes back to Scott (1976) depiction of collectivist communities, contrasted with Popkin (1979) characterization of generally individualistic and inefficient relationships. The debate remains unresolved decades later, as evidence has accumulated on both sides. Goldstein, de Janvry, and Sadoulet (2004), for example, depict patterns of inclusion and exclusion in community-level risk-sharing arrangements in Southern Ghana, drawing on insights into incomplete informal insurance (for more on fragmented collective relationships, see Fafchamps & Lund, 2004; Genicot & Ray, 2003; Plateau, 2000). This strand of literature shows how and why households may remain vulnerable even to idiosyncratic risks, in addition to broad aggregate risks. Grimard (1997) looks for evidence of risk sharing across regions in Côte d'Ivoire. He, too, rejects complete insurance but finds evidence of partial insurance between members of the same ethnic groups, especially in regions with the lowest availability of formal financial products. As Grimard suggests, the lack of full regional insurance might not come as a surprise given the difficulty in monitoring and enforcement over long distances. Munshi and Rosenzweig (2007) provide complementary data on intracaste and intrafamily insurance in India.

Taking this literature as a whole, it becomes clear that insurance provided by the formal sector should be seen in the context of a broader array of risk-coping mechanisms employed by households, some deployed preventatively and some used aftershocks have

occurred. Formal sector insurance includes a “loading factor” that includes taxes and administrative costs, and, as a starting point, insurance will be demanded only if compared to other options it is both relatively effective and relatively cheap.

7.2 Partnership models and index-based insurance

Taken as a whole, the empirical literature on informal collective insurance shows that poor households remain substantially without insurance (especially when aggregate risk is considered) while richer households tend to be better insured. The literature on information asymmetries suggests that the gaps in coverage are not just inequitable but apt to be inefficient as well.

The literature on informal credit and saving has offered guidance for developing commercially implementable credit and saving products, but parallels are harder to find with regard to insurance. One hurdle for commercial insurers is posed by the logistical challenge of collecting small-sized premia from customers and needing to assess claims for losses that may loom large for small households but which are relatively small for major insurers (Morduch, 2006). The pursuit of profitability makes serving the poor with current technologies a low commercial priority.

Two new approaches are starting to change the equation. The first is partnership models, whereby commercial insurers partner with microfinance institutions to deliver commercially viable products. The insurers bear the major risks and maintain responsibility for actuarial calculations and pricing strategies, while the microfinance institutions use their existing relationships with customers to sell products and handle claims. The model has been especially successful in delivering “credit-life” insurance, in which microfinance customers receive term life insurance coverage during the life of their loans. The product is relatively simple to administer, and premia are typically collected as extra fees on top of the interest rates paid for loans. A disadvantage is that this pricing may be opaque to borrowers and many products appear to be costly relative to the value of coverage provided. Also coverage extends to borrowers only—coverage lapses if customers choose to stop borrowing. Still, the principle is robust, and the promise remains for developing an improved array of insurance products.

The second new approach is index-based insurance (Carter et al., 2007; Skees, Varangis, & Larson, 2004). As high transaction costs, moral hazard, and adverse selection have thwarted attempts to provide crop insurance on a commercial basis and wide scale, new approaches have sought to sidestep those problems by shifting from insuring crop losses to insuring bad weather realizations instead. The insight is that if the correlation between crop losses and bad weather is high enough, substantial insurance can be provided through index-based weather insurance. Farmers are powerless to change the weather; thus, moral hazard and adverse selection disappear. Transactions costs also fall since claims do not need to be verified and products can be standardized around a given weather station.

An example is given by a rainfall insurance product offered in Andhra Pradesh, South India studied by Giné, Townsend, and Vickrey (2007b). To get a sense of the product, it is worth reviewing contract details. The insurance contract divides the cropping season into three parts, roughly corresponding to sowing, podding/flowering, and harvest phases, and farmers can purchase separate contracts for each part. The risk in the early phases is that rainfall will be insufficient, so, in this example, the contract pays nothing if rainfall exceeds 70 mm. If accumulated rainfall is less than 70 mm, the policy pays 10 rupees for each millimeter of rainfall below the cutoff, paying out a fixed amount (1000 rupees) when the season is extremely dry. In the third (harvest) phase, problems emerge when rainfall is excessive, so the policy reverses itself; it now pays out when rainfall exceeds 70 mm and pays nothing below the threshold. A policy covering all three phases is inexpensive enough to be accessible to low-income farmers (coverage costs 200–300 rupees or US\$5–6; Giné, Townsend, & Vickrey, 2007a).

In principle, even villagers who are not farmers can purchase contracts. While crop insurance is marketed only to farmers, there is nothing stopping the sale of weather insurance to anyone in the region who wants protection from the ups and downs of weather-related demand and supply fluctuations. Since the risks are correlated locally, prudence requires that rainfall insurance be offered in partnership with a reinsurer who can help local retailers spread risks across regions. The large global reinsurers like Munich Re and Swiss Re are playing a role in spreading risks across countries and broad regions.

7.2.1 Low demand

The idea of rainfall insurance makes sense. As a reality, though, the product in Andhra Pradesh described has not been embraced enthusiastically by farmers, despite its relatively low cost and the major costs of drought in the region (Giné et al., 2007a). The reasons why most people refused to buy it are hard to pin down, but basis risk between insurance payouts and the risk insured is a major determinant.

The value of rainfall insurance depends on a high correlation between incomes and rainfall as measured at the local rain gauge or weather station; the divergence is so-called basis risk. Two forces combine to create basis risk. First, there may be an insufficient number of rainfall gauges to adequately capture weather variations in a region. Second, the impact of weather on a given plot may be affected by the characteristics of the plot, including its slope, soil quality, and the availability of alternative water sources; the contract's pricing structure thus may not make sense for all farmers equally, even when they face the same weather patterns. Not surprisingly, Giné et al. (2007a) find that take-up of the Andhra Pradesh rainfall insurance product decreases with basis risk.

They also find that demand increases with household wealth and decreases with the extent to which credit constraints bind; again neither is surprising although since it is the poorest households that are generally least able to bear risk, the opposite finding

(that take-up decreases with wealth) would also be plausible. One clearly surprising result from Giné et al. (2007a) is that take-up falls with risk aversion. Since those who exhibit the greatest degree of risk aversion should be most eager to obtain insurance, the result falls outside the benchmark model. The most likely explanation is that it is uncertainty about the product itself (Is it reliable? How fast are pay-outs? How great is basis risk?) that drives down demand.

The evidence on low demand is repeated with a similar product in a very different setting. Giné and Yang (2008) investigate the demand for a hybrid credit–rainfall insurance product in Malawi. Their hypothesis is that risk-averse farmers will be unwilling to adopt new agricultural technologies—in this case, high-yielding varieties of maize and improved groundnut seeds. Coupling insurance with a credit product should then, in principle, increase adoption. The study employs a randomized field experiment involving about 800 maize and groundnut farmers; half the farmers were offered credit to buy the new seeds, while the other half were offered a similar credit product coupled with a weather insurance policy. The insurance policy was priced at actuarially fair rates, providing some protection against low rainfall outcomes. (Though, we note that given that there is basis risk, a rate that is actuarially fair in terms of weather realizations at the rainfall gauge may not be actuarially fair on a given farmer's plot.) A third of the farmers who were offered just the credit contract accepted it. Surprisingly, take-up was 13 percentage points lower for the half of the sample offered the combined credit–insurance product. One possible explanation discussed in the working paper version is cognitive: insurance is not a simple concept and the terms of the policy can be hard to weigh, a contention supported by the finding that more educated farmers were more likely to take up the combined product in the Malawi experiment. Increased exposure to the product, financial literacy training, or a new marketing strategy may all play a role in raising demand levels. Another explanation is that farmers already received insurance implicitly through loan contracts: in case of a serious drought, farmers do not have to repay loans (by virtue of limited liability); this, in itself, can limit demand for the insurance product. As with the case of the Indian product, it is also critical that the product itself be desirable in more basic ways (with reasonable administrative costs, reliably speedy pay-outs, and a sufficient spread of rainfall gauges to ensure limited basis risk—factors that are not always in place). In the scale-up of this product, the firms in Malawi no longer sell directly to farmers, and are instead insuring the lenders and large firms directly.

Carter et al. (2007) argue that a variation on index-based agricultural insurance can sometimes do better. This is area-based yield insurance, in which payouts are based on measured average yields in a region (rather than on yields on the policymakers own plot). Basing pay-outs on average yields in, say, a valley, effectively eliminates moral hazard and adverse selection in the same way that weather-based insurance does (Miranda, 1991). The advantage is that coverage can extend beyond risks due to

weather only—a finding stressed in the Carter et al. (2007) application in northern Peru. A disadvantage is that, as with traditional crop insurance, yields need to be measured, a sometimes expensive and time-consuming process. And, as with rainfall insurance, basis risk remains a problem for customers.

These studies provide a start at unpacking the reasons that the demand for insurance tends to be low. Additional reasons include the fact that the effectiveness of informal insurance mechanisms may, in some cases, be sufficient to limit the net impacts of formal insurance contracts (Morduch, 1999), and the expectation that in severe disasters the public sector will step in with aid, again limiting the net impact of private insurance. The importance of marketing in determining demand for credit (Bertrand et al., 2010) also highlights the urgency for insurers to identify new modes of marketing to enhance take-up of insurance. Given that the prevalence of risk is widely taken to be a fundamental element in the lives of poor households, it is notable that such fundamental questions about the demand for insurance remain; the topic is a priority on the research agenda.

7.3 Health insurance

Health risks loom large among the risks faced by poor households, and out-of-pocket health spending is high. In the United States, the percentage of health spending that is out-of-pocket (i.e., not paid by insurance) is about 12% (Pauly, Blavin, & Meghan, 2008). In Bangladesh, the comparable figure is 64%; in Ghana, 59%; in India, 78%; in Paraguay, 55%; in the Philippines, 47%; and in Vietnam, 62%. The figures are from the 2002 World Health Survey, a World Health Organization household survey covering about 4000–6000 households per country (cited in Pauly et al., Table 1).¹⁴

As with crop insurance, few commercially successful health insurance programs exist that serve poor communities (Morduch, 2006). And, as with the other types of insurance discussed earlier, moral hazard and adverse selection pose substantial barriers for insurers. In line with theory, copayments can help. Grameen Kalyan, a Bangladesh-based health insurance scheme that is part of Grameen Bank offers coverage for preventative and curative health services, and requires copayments for the curative services. Copayments turned out to help Grameen Kalyan not only reduce overuse of medical services, but they helped to effectively signal the quality of care as well (given that customers judged quality by price; Radermacher, Dror, & Noble, 2006, p. 78). Elsewhere, though, copayments have been viewed skeptically in poor populations. In Mali, for example, copayments were feared to be too onerous for poor customers, and a system of cooperatives was founded to help members pay copayments for health services (Radermacher et al., 2006, p. 78). It is unclear whether the problem here is with requiring copayments *per se* or with the levels at which they are set—a problem ripe for research.

One of the themes of the chapter is that design matters, and the sentiment is echoed here with regard to health insurance. Above all else, insurers sell their reputations: the

promise companies will reliably and promptly deliver pay-outs when crises hit (in the exact amounts and with the exact timing specified in contracts). Such basic reliability cannot be taken for granted in either the public or private sectors (Das, Hammer, & Leonard, 2008). If doctors, nurses, and pharmacists are unavailable or of uncertain quality, a reliable health insurance system is hard to envision. Yet experiences with hospitalization insurance in South India suggest that this conclusion may put things backward. In the health insurance program of BASIX, for example, the organization of villagers into insurance groups, facilitated by their membership in BASIX's microfinance program, created a large enough block of customers that it was possible for the insurer to certify and contract with high-quality doctors directly. This "preferred provider" system benefited doctors and medical personnel who gained from the steady demand. It was thus the existence of the health insurance program, and the financial resources it aggregated, that helped fix quality deficiencies in healthcare quality faced by customers. In other programs, like that of BRAC and Grameen Kalyan in Bangladesh, nearly all health care is provided within the programs' own clinics (Radermacher et al., 2006, pp. 86, 91).

Another concern rests with cost. Without reliable data, insurers face a series of risks in pricing insurance, most importantly, misestimating the probabilities of loss and fluctuations in health care costs (Radermacher et al., 2006, p. 90). But even if it is possible to set rates precisely, it is unclear that customers are willing to shoulder the full costs. Product design features appear to help. Some institutions lend money to customers to help them pay premiums: FINCA Uganda, for example, and Karuna Trust in Karnataka India. Others break the premiums into small-sized installments that can be easier for poor households to handle (though when given a choice, BRAC customers sought monthly rather than weekly installments). But most of the health insurance programs assessed by Radermacher et al. (2006) nevertheless lost money. The unanswered question is whether, if offered a truly effective high-quality health insurance policy, households would willingly pay the required price. As it is, many customers fail to renew their contracts at the end of a given year—Radermacher et al. (2006, Table 10) show BRAC's renewal rate at 51%, Grameen Kalyan's at 54%, and VimoSEWA's (part of a long-established community institution in Ahmedabad, India) at 51%. Better marketing and rising comfort levels will likely help raise renewal rates, but the evidence on renewal rates suggests a *prima facie* case that success will also require improving service quality for the price. This is an area that will surely benefit from the kinds of experimental economic research that has energized the literatures on credit and saving.

8. GOVERNMENTS, BUSINESSES, AND NONPROFIT INSTITUTIONS

Yunus (2008), in defense of his strategy in founding Grameen Bank, argues for expanding financial access by launching "social businesses." In this case, this means banks that earn modest profits and that do not distribute them to investors; instead profits are

reinvested in institutions to further social goals. Such social businesses are necessarily dependent on donors and other “social investors” who value the institutions’ social aims and who are willing to sacrifice financial returns in order to obtain greater social returns. Economists in the tradition of Friedman (1970) argue instead that the optimal path entails maximizing pure profit. Surpluses can then be distributed to improve social welfare. In this vision, there is no role for social businesses.

Resolving these conflicting visions has been hampered by the lack of evidence on the impacts of subsidy, allowing ideological positions (on all sides) to go unchecked. The specter of inefficient, heavily subsidized public banks hangs over the microfinance movement, and it is partly responsible for a lack of enthusiasm for embracing donated funds. Microfinance advocates, wary of subsidies, have energetically sought to move microfinance to a commercial basis, free of long-term subsidies. Their goal is to expand scale while not sacrificing mission. There is thus a general consensus among microfinance leaders, that problems tend to arise when the government becomes a direct lender (United Nations, 2008); the main problems include inefficiency and the mistargeting of subsidy.

Conning and Morduch (2007), though, argue that while commercial investors play important roles, by virtue of their philanthropic objectives, social investors can improve social welfare and total surplus in ways that commercial investors by themselves often cannot. Their theoretical model is built around a nested moral hazard problem and draws on approaches to corporate finance (e.g., Dewatripont & Tirole, 1994; Tirole, 2005). At one level is the well-known moral hazard problem in credit markets, in which borrowers lack collateral and thus contracts are maximized subject to a limited-liability constraint. In these second-best contracts, optimal levels of effort are generally impossible to elicit (i.e., effort to ensure that risky ventures work out well). A second layer of moral hazard exists between investors and institutions (Hansmann, 1996). Just as lending to villagers entails risk, investing in microlenders also entails risk, and such risk is also not generally backed by collateral.

In this setting, Conning and Morduch (2007) show cases in which philanthropic giving can raise output. Thus, not only does it reduce poverty but it can also increase an economy’s overall productivity. The possibility arises because social investors, by definition, are willing to absorb costs in order to bring gains. (This is, of course, their reason for existence.) They can then do more than shift resources: they can also help increase the total sum of resources available.

The result emerges when social investors expand the contract space by being willing to subsidize the “limited-liability rents” that often stand in the way of trades in the commercial loan market. (Limited-liability rents refer to the returns that must be left with borrowers, or the loan officers that monitor them, in order to provide incentives to work hard even when limited liability constraints restrict incentives created by threatened punishments.) When such rents become large relative to expected project

gains, lenders may opt not to make the loans, leaving socially valuable projects unfunded. A small subsidy in such contexts may be enough to make lending profitable. If enough small loans become profitable, a new microfinance institution might emerge where otherwise it might not have.

In the Conning and Morduch (2007) framework, profitability is neither necessary nor sufficient for attracting commercial capital given that risky investments remain unsecured. The framework implies that the most effective uses of social investment are either “high” or “low” when directed to poor communities. In the “high” case, optimal subsidy involves allocations that either go to institutions serving relatively well-off (but still poor) customers, where some collateral is obtainable and the potential for leverage is maximized (i.e., commercial funds can be drawn in). In the “low” case subsidies go to institutions serving the poorest customers where social gains are greatest. There is less “bang for the buck” in supporting a middle range of institutions that are self-sufficient with their own resources but for which leverage (and potential scale) is limited. The Conning–Morduch framework thus embraces a range of strategies to promote financial access.

In practice, though, the microfinance industry has divided around competing strategies. Cull et al. (2009b) shed light on the trade-offs between those competing strategies. They employ a dataset that includes 346 of the world’s leading microfinance organizations and covers nearly 18 million borrowers. One strength of the dataset is that the data are adjusted to show the roles of both explicit and implicit subsidies (generally delivered as soft loans).¹⁵ The data suggest that, while commercialization is a powerful trend, commercial banks and those that are avowedly “social businesses” like Grameen Bank are not substitutes. The data show that microfinance is not taking a single path, nor that it should. The authors put forward a series of basic findings to frame debates:

1. Licensed commercial banks with for-profit status serve a growing share of low-income customers and garner media attention, but the share remains relatively small. Globally, microfinance continues to be dominated by nongovernmental organizations (NGOs), government-owned banks, and “nonbank financial institutions” that are a cross between banks and NGOs.
2. Being a nonprofit institution does not mean being unprofitable. A large share of microfinance institutions with “nonprofit” status in fact earn steady profits—but (as “social businesses”) they reinvest their profits in the institution and cannot legally distribute earnings to shareholders. Earning profits (and thus limiting dependence on subsidies) and becoming a commercial entity are distinct activities. Neither implies the other.
3. Commercial microfinance banks, as a group, make loans that on average are about four times larger than loans from NGOs. Since poorer customers generally

demand smaller loans, average loan size is a rough proxy for the poverty level of customers. On average, commercially oriented microfinance banks thus tend to serve a substantially better-off group of borrowers than do NGOs. These microfinance banks, as a group, also serve fewer women as a share of their customers.

4. Most microfinance institutions charge inflation-adjusted interest rates between 20% and 40% per year. NGOs as a group charge interest rates that are roughly double the size of the average charged by commercial microfinance banks. Thus, the poorest customers tend to pay the highest interest rates on loans. The high interest rates are necessary to cover the added costs of making small loans, and the NGOs' record of expansion and high loan repayment rates over time suggests that customers value the services, even at high costs.
5. Despite the high interest rates, most of the institutions serving the poorest customers earn profits too small to attract profit-maximizing investors.
6. Thus, subsidies and noncommercial funding continue to be important to non-governmental organizations, while banks rely mainly on social investment and commercial sources of capital.

The findings are generally in line with Conning's and Morduch's framework, which suggests that debating about a single, correct vision for microfinance—be it a nonprofit approach or a commercial model—misses the reality that microfinance flourishes thanks to a diversity of strategies. One key question, that remains unanswered, hinges on whether subsidized microfinance institutions serving the poor are likely to unfairly compete against commercial institutions. The Cull et al. (2009b) evidence suggests that this is unlikely since being subsidized does not mean being inexpensive; as noted earlier, even with subsidies, poor households pay more for credit than richer households.

The role for governments remains unclear. While state-run banks have a generally poor record as lenders, government savings banks and postal savings schemes have proven successful around the world. Beyond that direct role, policymakers find an important role in setting appropriate regulation and creating a supportive environment for expanding financial access. As Ahlin and Lin (2006) suggest in cross-institution regressions, macroeconomic environments featuring steady growth and low inflation not surprisingly improve the performance of microfinance institutions.

Cull, Demirgüç-Kunt, and Morduch (2009a) argue that regulation is critical, especially for deposit-taking microfinance banks, but they demonstrate that regulation carries costs. They draw on a database that combines financial data on 245 microfinance institutions with data on levels of prudential supervision. Their basic regressions show that regulatory supervision is negatively associated with profitability. Once the nonrandom assignment of supervision is addressed via treatment effects and instrumental variables regressions, supervision appears to be associated with substantially larger average loan sizes and less lending to women than in the least squares regressions, though it

is not significantly associated with profitability. Cull et al. (2009a) conclude that, as a group, profit-oriented microfinance institutions absorb the cost of supervision by curtailing outreach to market segments that tend to be more costly per dollar lent. Cull et al. (2009a) do not argue against supervision, but argue that policymakers should take trade-offs into account.

9. CONCLUDING COMMENTS

In the past decade, research on financial access for the poor has made a transition that parallels the success of practitioners in bringing successful microfinance programs to communities around the world. The earlier mode for research on financial access centered on understanding the financing constraints and opportunities of poor households. This work laid out the case for innovations, and researchers aimed to quantify the potential gains from relaxing financing constraints. For the most part, the scenarios were hypothetical given that actual banking institutions capable of providing reliable, professional financial services in poor communities were scarce. In focusing on understanding household needs, research on financial access had much in common with research on healthcare and education in poor parts of developing countries. Researchers provided insight into households' coping mechanisms and helped draw the outline of optimal solutions.

In the past decade, the financial landscape has transformed in villages and slums worldwide, and so have important strands of scholarship. Gaps in financial access remain wide, but viable institutions dedicated to serving poor and low-income communities have penetrated thickly in Bangladeshi villages and are making substantial inroads in rural India. In Latin America and Africa, most of the action has been urban so far, but innovation has been rapid. The spread of mobile telephones, for example, heralds the possibility of expanding financial services widely via mobile banking in sub-Saharan Africa. Customers in parts of the Philippines, Peru, and Bolivia can now choose between a range of different microfinance providers actively seeking their business.

This chapter illustrates how much is being learned from this expansion of activity—and how much remains to learn. If there is a single message of the chapter, it is that mechanisms matter. The design of products (including their prices, term structure, flexibility, and marketing) affects adoption and usage—and ultimately economic and social impacts. That insight is made visible in part through a shift in research strategies. Much of the new work described earlier involves researchers creating partnerships with existing financial providers, both commercial banks and public or nonprofit institutions. The partnerships generally center on testing the effects of systematically varying delivery mechanisms or introducing new services. The lessons are thus not hypothetical but are based on actual products delivered by actual institutions. Most of the lessons are, by definition, directly implementable.

This shift in research strategy puts a natural focus on innovation, with practitioners seeking to expand existing approaches or test competing strategies. The practice of medical and consumer product and marketing trials has given the new research much of its direction, including the focus on randomized controlled trials. While trial-based approaches are not the only way that researchers are analyzing trends and possibilities, they have become an important part of the mixture of methodologies. The trial-based methods have been particularly valuable in testing long-held assumptions and opening the door to behavioral perspectives that introduce elements of psychology into the economic framework. One important lesson that the trials yield, and one which applies as well to sectors like health and education, is that modes of implementation matter. The new research orientation complements the traditional more observational approach to empirical economic research by working with functioning institutions, identifying needs that they might fill through innovation and adaptation, and designing studies to test theories fundamental to individual decision making and functioning of markets.

End Notes

- *. Karlan and Morduch thank Nathanael Goldberg and Catherine Burns for assistance in writing this chapter. We acknowledge the Bill and Melinda Gates Foundation for support through the Financial Access Initiative. Their funding supported research cited here and the writing of the chapter. Karlan thanks the National Science Foundation CAREER SES-0547898 and the Alfred P. Sloan Foundation for support. The opinions herein are those of the authors and not those of any of the funders.
- 1. The term microcredit was later expanded to “microfinance” by Yunus’ followers to be inclusive of other financial services for the poor, including savings and insurance.
- 2. We do not focus on remittances in this chapter, except briefly in the technology section as it relates to interventions such as electronic banking. However, we note the growing role of remittances in international capital flows (and in poor households’ financial lives). The World Bank reports that remittance flows to developing countries totaled \$251 billion in 2007, more than doubling since 2002 (Ratha, Mohapatra, Vijayalakshmi, & Xu, 2008). Part of the increase in remittances may be due to improved data recording and the depreciation of the US dollar (Savage & Harvey, 2007), but, even so, the figures likely underestimate total flows (Orozco, 2007). In particular, these data only capture officially recorded transfers, and a large portion of remittances appear to flow through informal channels and thus remain unrecorded. Estimates that impute unrecorded flows yield levels of remittances that are greater than foreign direct investment flows and more than twice the level of official aid received by developing countries (Mohapatra, Ratha, Vijayalakshmi, & Xu, 2006).
- 3. Randomized trials have proved especially valuable when evaluating the working of new mechanisms, a focus of the present chapter. Other econometric approaches have been (and will continue to be) valuable in assessing directions for financial access and relationships between economic and financial variables.
- 4. The “poorest” is defined as being below the World Bank UN international poverty line of \$1 a day in purchasing power adjusted dollars, or being in the bottom half of a country’s poor population (as defined by the local poverty line).
- 5. Accessed from <http://www.promujer.org/meetourclients.html> on June 2, 2008.

6. See also Banerjee and Duflo (2005) for a parallel survey of the literature on returns to capital in developing economies.
7. “Adverse selection” in this context also includes “lower anticipated effort” which may be generated through the same pooling process that traditional adverse selection on risk-type generates. Such an “anticipated effort” effect requires moral hazard to be possible, since if effort is perfectly observable this effect would not occur. However, hidden information must also be present for this mechanism to be relevant, since otherwise firms would price according to anticipated effort by the borrower.
8. Characteristics with explanatory power are: whether the wife of the household head is literate, whether the wife of the household head works for a wage, the number of “prime-aged” (18-45) women in the household, and the amount of land owned by the household.
9. Despite the high cost Susu services are extremely popular: Aryeetey and Gockel (1991) reports 78% of market women in Ghana’s largest cities using Susu collectors to save. Susu collectors lend too, but the risk to depositors seems limited: the collectors surveyed by Aryeetey and Steel lent only small amounts to a small portion of their client base. Borrowers are screened based on regular savings history, and the Susu collectors typically lend only half of a month’s deposits to a given client. On average less than 10% of their portfolios were exposed to lenders.
10. Two alternative stories are worth mentioning. Women who have all of the income may also be too busy for the ROSCA, if the ROSCA is time consuming, and women who earn all of the income in the household may also be better financial managers, less likely to have self-control problems, and thus also less likely to join a ROSCA.
11. Interestingly, this preference for commitment was strongest among women. The study did not provide data to help understand heterogeneity across gender, and thus future work on this issue would be fruitful. It is suggestive of an important interaction between gender and the preference and need for commitment contracts, as discussed here.
12. A sense of the magnitude of the problem is given by Deaton’s (1992) simulations of optimal asset accumulation under borrowing constraints. Deaton begins with a dynamic stochastic choice problem in which a household builds up and draws down assets in order to dampen the variability of consumption in the face of income swings. Since by assumption the household cannot borrow, assets are used as buffers, and a relatively low level of assets can be used to smooth independent, identically distributed shocks. When the pattern of shocks has even a moderate autoregressive component, a much higher average level of assets is required to smooth consumption.
13. One of the limits to this literature is given by difficulties in interpreting results. Consider the finding of Jalan and Ravallion (1999) on risk-sharing in China. They find that on average 40% of idiosyncratic income shocks translate into consumption shocks for the poorest households. Readers are left unclear as to whether the finding means that all poor households are vulnerable to 40% of shocks—or might, say, half fully insure while the other half suffers 80% of the loss? Moreover, are positive shocks handled differently from negative shocks? These kinds of questions are less important if the focus is on testing the basic fit of a benchmark model of full consumption insurance as in Townsend (1994), but they are critical in evaluating and developing policy responses.
14. Most of the spending is attributable to spending on prescription drugs. Turning to the same countries, the percentages are: Bangladesh, 84%; Ghana, 48%; India, 55%; Paraguay, 73%; the Philippines, 61%; and in Vietnam, 44% (World Health Survey, cited in Pauly et al., Table 3).
15. Cull et al. (2008) report that the adjustments include “an inflation adjustment, a reclassification of some long-term liabilities as equity, an adjustment for the cost of subsidized funding, an adjustment for current-year cash donations to cover operating expenses, an in-kind subsidy adjustment for donated goods and services, loan loss reserve and provisioning adjustments, some adjustments for write-offs, and the reversal of any interest income accrued on non-performing loans.”

References

- Adams, D. (1978). Mobilizing household savings through rural financial markets. In J. D. Von Pischke, D. Adams & G. Donald (Eds.), *Rural financial markets in developing countries: Their use and abuse*. Baltimore: Johns Hopkins University Press.
- Adams, D., & Pischke, J. D. V. (1992). Microenterprise credit programs: *deja vu*. *World Development*, 20, 1463–1470.
- Adams, D. W., Graham, D. H., & von Pischke, J. D. (1984). *Undermining rural development with cheap credit*. Boulder: Westview Press.
- Aghion, P., & Bolton, P. (1997). A theory of trickle-down growth and development. *The Review of Economic Studies*, 64, 151–172.
- Ahlin, C., & Lin, J. (2006). *Luck or skill? MFI performance in macroeconomic context*. BREAD Working Paper.
- Ahlin, C., & Townsend, R. M. (2007). Using repayment data to test across models of joint liability lending. *Economic Journal*, 117, F11–F51.
- Ainslie, G. (1992). *Picoeconomics*. Cambridge: Cambridge University Press.
- Al-Bagdadi, H., & Cracknell, D. (2005). Promotion: The final step in marketing. In M. Hirschland (Ed.), *Savings services for the poor: An operational guide*. Bloomfield: Kumarian Press.
- Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58, 450–474.
- Anderson, S., & Baland, J.-M. (2002). The economics of Roscas and intra-household resource allocation. *Quarterly Journal of Economics*, 117, 963–995.
- Ando, A., & Modigliani, F. (1963). The ‘life-cycle’ hypothesis of saving: Aggregate implications and tests. *American Economic Review*, 53, 55–84.
- Aportela, F. (1999). *Effects of financial access on savings by low-income people*. MIT Department of Economics Dissertation (Chapter 1).
- Armendáriz, B., & Morduch, J. (2010). *The economics of microfinance* (2nd ed.). Cambridge, MA: MIT Press.
- Armendariz de Aghion, B., & Morduch, J. (2000). Microfinance beyond group lending. *Economics of Transition*, 8, 401–420.
- Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53, 941–967.
- Aryeetey, E., & Gockel, F. (1991). *Mobilizing domestic resources for capital formation in Ghana*. African economic research consortium. Nairobi: Initiatives Ltd.
- Aryeetey, E., & Steel, W. F. (1995). Informal savings collectors in Ghana: Can they intermediate? *Finance and Development*, 31, 36–37.
- Ashraf, N., Karlan, D., & Yin, W. (2006a). Deposit collectors. *Advances in Economic Analysis & Policy*, 6, Article 5, 1–25.
- Ashraf, N., Karlan, D., & Yin, W. (2006b). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quarterly Journal of Economics*, 121, 673–697.
- Ashraf, N., Karlan, D., & Yin, W. (2010). Female empowerment: Further evidence from a commitment savings product in the Philippines. *World Development* (forthcoming).
- Banerjee, A. (2004). The two poverties. In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.
- Banerjee, A., & Duflo, E. (2005). Growth theory through the lens of development economics. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic growth*. Amsterdam: Elsevier.
- Banerjee, A., & Duflo, E. (2007). The economic lives of the poor. *Journal of Economic Perspectives*, 21, 141–167.
- Banerjee, A., & Duflo, E. (2008). *Do firms want to borrow more? Testing credit constraints using a directed lending program*. MIT Working Paper.
- Banerjee, A., Duflo, E., Glennerster, R., & Kinnar, C. (2009). *The miracle of microfinance? Evidence from a randomized evaluation*. Working Paper, MIT Department of Economics and Abdul Latif Jameel Poverty Action Lab.
- Banerjee, A., & Newman, A. (1993). Occupational choice and the process of development. *Journal of Political Economy*, 101, 274–298.

- Banerjee, A., & Newman, A. (1994). Poverty, incentives, and development. *American Economic Review*, 84, 211–215.
- Basu, K. (2008). *A behavioral model of simultaneous borrowing and saving*. Working Paper. University of Chicago.
- Basu, P. (2006). *Improving access to finance for India's rural poor*. Washington, DC: The World Bank.
- Bauer, M., Chytilova, J., & Morduch, J. (2008). *Behavioral foundations of microcredit: Experimental and survey evidence from rural India*. Working Paper, New York University.
- Beck, T., Demirguc-Kunt, A., & Levine, R. (2007). Finance, inequality and the poor. *Journal of Economic Growth*, 12, 27–49.
- Beck, T., Demirguc-Kunt, A., Peria, M., & Soledad, M. (2006). *Banking services for everyone? Barriers to bank access and use around the world*. Working Paper Series No. 4079. World Bank Policy Research.
- Beck, T., Levine, R., & Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics*, 58, 261–300.
- Beegle, K., Dehejia, R., & Gatti, R. (2003). *Child labor, crop shocks, and credit constraints*. Working Paper No. 10088. National Bureau of Economic Research.
- Bernheim, D., Garret, D., & Maki, D. (2001). Education and saving: The long-term effects of high school financial curriculum mandates. *Journal of Public Economics*, 80, 435–465.
- Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's advertising content worth? A field experiment in consumer credit market. *Quarterly Journal of Economics*, (forthcoming).
- Bertrand, M., Mullainathan, S., & Shafir, E. (2006). Behavioral economics and marketing in aid of decision making among the poor. *Journal of Public Policy and Marketing*, 25, 8–23.
- Besley, T. (1994). How do market failures justify interventions in rural credit markets? *World Bank Research Observer*, 9, 22–47.
- Besley, T. (1995). Savings, credit and insurance. In T. N. Srinivasan & J. Behrman (Eds.), *Handbook of development economics*. Amsterdam: Elsevier.
- Bhaduri, A. (1973). A study in agricultural backwardness under semi-feudalism. *Economic Journal*, 83, 120–137.
- Bolton, P., & Scharfstein, D. S. (1990). A theory of predation based on agency problems in financial contracting. *American Economic Review*, 80, 93–106.
- Boucher, S., Carter, M., & Guirkinger, C. (2005). *Risk rationing and activity choice in moral hazard constrained credit markets*. Working Paper.
- Boucher, S. R., Carter, M. R., & Guirkinger, C. (2008). Risk rationing and wealth effects in credit markets: Theory and implications for agricultural development. *American Journal of Agricultural Economics*, 90, 409–423.
- Browning, M., & Lusardi, A. (1996). Household saving: Micro theories and micro facts. *Journal of Economic Literature*, XXXIV, 1797–1855.
- Burgess, R., & Pande, R. (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review*, 95, 780–795.
- Carter, M., Galarza, F., & Boucher, S. (2007). *Underwriting area-based yield insurance to crowd-in credit supply and demand*. Working Paper, UC Davis 07-003. Giannini Foundation of Agricultural Economics.
- Case, A., & Deaton, A. (1998). Large cash transfers to the elderly in South Africa. *Economic Journal*, 108, 1330–1361.
- Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92, 1308–1334.
- Charness, G., & Genicot, G. (2007). *Informal risk sharing in an infinite-horizon experiment*. Working Paper. University of California-Santa Barbara and Georgetown University.
- Clarke, G., Xu, L. C., & Zou, H.-F. (2006). Finance and income inequality: What do the data tell us? *Southern Economic Journal*, 72, 578–596.
- Coleman, B. (1999). The impact of group lending in Northeast Thailand. *Journal of Development Economics*, 60, 105–141.
- Cole, S., & Shastry, G. K. (2008). *If you are so smart, why aren't you rich? The effects of education, financial literacy, and cognitive ability on financial market participation*. Working Paper.

- Collins, D., Morduch, J., Rutherford, S., & Ruthven, O. (2009). *Portfolios of the poor: How the World's poor live on \$2 a day*. Princeton: Princeton University Press.
- Conning, J. (2005). *Monitoring by delegates or by peers? Joint liability loans under moral hazard*. Working Paper.
- Conning, J., & Morduch, J. (2007). *Social finance*. Working Paper. New York University.
- Cull, R., Demirgüç-Kunt, A., & Morduch, J. (2007). Financial performance and outreach: A global analysis of leading microbanks. *Economic Journal*, 117, F107–F133.
- Cull, R., Demirgüç-Kunt, A., & Morduch, J. (2009a). Does regulatory supervision curtail microfinance profitability and outreach? Working Paper. New York University and World Bank.
- Cull, R., Demirgüç-Kunt, A., & Morduch, J. (2009b). Microfinance meets the market. *Journal of Economic Perspectives*, 23, 167–192.
- Cull, R., & Scott, K. (2009). Measuring household usage of financial services: Does it matter how or to whom you ask? In: World Bank Conference on Measurement, Promotion, and Impact of Access to Financial Services, Washington, DC.
- Daley-Harris, S. (2009). *State of the microcredit summit campaign report 2009*. Washington, DC: Microcredit Summit Campaign.
- Das, J., Hammer, J., & Leonard, K. (2008). The quality of medical advice in low-income countries. *Journal of Economic Perspectives*, 22, 93–114.
- Deaton, A. (1992). *Understanding consumption*. Oxford: Clarendon Press.
- Deaton, A. (1997). *The analysis of household surveys*. World Bank.
- Deaton, A. (2006). *Purchasing power parity exchange rates for the poor: Using household surveys to construct PPPs*. Working Paper. Princeton University Research Program in Development Studies.
- Dehejia, R., Montgomery, H., & Morduch, J. (2009). *Do interest rates matter? Credit demand in the Dhaka Slums*. Working Paper. NYU Wagner Graduate School.
- de Janvry, A., McIntosh, C., & Sadoulet, E. (2008). *The supply- and demand-side impacts of credit market information*. Working Paper. University of California-Berkeley and University of California-San Diego.
- de Mel, S., McKenzie, D., & Woodruff, C. (2008a). *Are women more credit constrained? Experimental evidence on gender and microenterprise returns*. Working Paper No. 4746. World Bank.
- de Mel, S., McKenzie, D., & Woodruff, C. (2008b). Returns to capital in microenterprises: Evidence from a field experiment. *Quarterly Journal of Economics*, 123, 1329–1372.
- de Meza, D., & Webb, D. (1987). Too much investment: A problem of asymmetric information. *The Quarterly Journal of Economics*, 102, 281–292.
- de Meza, D., & Webb, D. (1989). The role of interest rate taxes in credit markets with divisible projects and asymmetric information. *Journal of Public Economics*, 39, 33–44.
- Demirgüç-Kunt, A., Beck, T., & Honohan, P. (2007). *Finance for all: Policies and pitfalls in expanding access*. World Bank.
- Demirgüç-Kunt, A., & Levine, R. (2008). *Financial and economic opportunity*. Working Paper No. 4468. World Bank.
- Dercon, S. (2004a). *Insurance against poverty*. Oxford: Oxford University Press.
- Dercon, S. (2004b). Risk, insurance, and poverty: A review. In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.
- Dercon, S., & Hoddinott, J. (2004). Health, shocks, and poverty persistence. In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.
- Dewatripont, M., & Tirole, J. (1994). *The prudential regulation of banks*. Cambridge, MA: MIT Press.
- Drugov, M., & Macchiavello, R. (2008). *Self-discovery, lending and microentrepreneurship*. Working Paper. Oxford University.
- Dubois, P. (2000). *Consumption insurance with heterogeneous preferences. Can sharecropping help complete markets?* Econometric Society World Congress 2000 Contributed Paper 1597.
- Duflo, E., Kremer, M., & Robinson, J. (2006). *Understanding technology adoption: Fertilizer in Western Kenya evidence from field experiments*. Working Paper.
- Duflo, E., & Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics*, 85, 121–148.
- Dupas, P., & Robinson, J. (2009). *Savings constraints and microenterprise development: Evidence from a field experiment in Kenya*. Working Paper. National Bureau of Economic Research.

- Emran, M. S., Morshed, A. M., & Stiglitz, J. E. (2007). *Microfinance and missing markets*. Working Paper. George Washington University and Columbia University.
- Fafchamps, M., & Lund, S. (2003). Risk sharing networks in rural Philippines. *Journal of Development Economics*, 71, 261–287.
- Fafchamps, M., & Lund, S. (2004). *Rural poverty, risk and development*. Cheltenham: Edward Elgar.
- Feder, G., Lau, L., Lin, J., & Luo, X. (1990). The relationship between credit and productivity in Chinese agriculture: A microeconomic model of disequilibrium. *American Journal of Agricultural Economics*, 72, 1151–1157.
- Field, E. (2007). Entitled to work: Urban property rights and labor supply in Peru. *Quarterly Journal of Economics*, 122, 1561–1602.
- Field, E., & Pande, R. (2008). Repayment frequency and default in microfinance: Evidence from India. *Journal of the European Economic Association*, 6, 501–509.
- Finkelstein, A., & Poterba, J. (2002). Selection effects in the United Kingdom individual annuities market. *The Economic Journal*, 112, 28–50.
- Fischer, G. (2008). *Contract structure, risk sharing and investment choice*. Working Paper 24. Institute for Financial Management and Research Centre for Micro Finance.
- Forster, S., & Reillie, X. (2008). *Foreign capital investment in microfinance: Balancing social and financial returns*. CGAP Focus Note 44.
- Foster, A. (1995). Prices, credit markets and child growth in low-income rural areas. *Economic Journal*, 105, 551–570.
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2001). Time discounting: A critical review. *Journal of Economic Literature*, 40, 351–401.
- Friedman, M. (1970). The social responsibility of business is to increase its profits. *The New York Times Magazine*, September 13.
- Fudenberg, D., & Levine, D. (2005). *A dual self model of impulse control*. Working Paper.
- Galor, O., & Zeira, J. (1993). Income distribution and macroeconomics. *Review of Economic Studies*, 60, 35–52.
- Garon, S. (2004). The transnational promotion of saving in Asia. ‘Asian values’ or the ‘Japanese model’. In S. Garon & P. L. Maclachlan (Eds.), *The ambivalent consumer: Questioning consumption in east Asia and the west*. Ithaca: Cornell University Press.
- Genicot, G., & Ray, D. (2003). Group formation in risk-sharing arrangements. *Review of Economic Studies*, 70, 87–113.
- Gersovitz, M. (1988). Saving and development. In H. Chenery & T. N. Srinivasan (Eds.), *Handbook of development economics*. Amsterdam: Elsevier.
- Gertler, P., & Gruber, J. (2002). Insuring consumption against illness. *American Economic Review*, 92, 51–70.
- Ghatak, M. (1999). Group lending, local information and peer selection. *Journal of Development Economics*, 60, 27–50.
- Ghosh, P., & Ray, D. (2001). *Information and enforcement in informal credit markets*. Boston University, Mimeo.
- Giné, X., Jakielka, P., Karlan, D. S., & Morduch, J. (2009). *Microfinance games*. Working Paper.
- Giné, X., & Karlan, D. (2009). *Group versus individual liability: Long term evidence from Philippine microcredit lending groups*. Working Paper.
- Giné, X., Townsend, R., & Vickrey, J. (2007a). *Patterns of rainfall insurance participation in rural India*. Working Paper 4408. World Bank Policy Research.
- Giné, X., Townsend, R., & Vickrey, J. (2007b). Statistical analysis of rainfall insurance payouts in Southern India. *American Journal of Agricultural Economics*, 89, 1248–1254.
- Giné, X., & Yang, D. (2008). *Insurance, credit, and technology adoption: Field experimental evidence from Malawi*. Working Paper.
- Goldsmith, R. W. (1969). *Financial Structure and Development*. New Haven: Yale University Press.
- Goldstein, M., de Janvry, A., & Sadoulet, E. (2004). Is a friend in need a friend indeed? Inclusion and exclusion in mutual insurance networks in Southern Ghana. In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.

- Gonzalez, A., & Rosenberg, R. (2006). *The state of microfinance—Outreach, profitability and poverty (findings from a database of 2600 microfinance institutions)*. World Bank Conference on Access to Finance.
- Greenwood, J., & Jovanovic, B. (1990). Financial development, growth, and the distribution of income. *Journal of Political Economy*, 98, 1070–1107.
- Grimard, F. (1997). Household consumption smoothing through ethnic ties: Evidence from Côte d'Ivoire. *Journal of Development Economics*, 53, 391–422.
- Gugerty, M. K. (2007). You can't save alone: Commitment in rotating savings and credit associations in Kenya. *Economic Development and Cultural Change*, 55, 251–282.
- Guirkinger, C., & Boucher, S. (2007). *Credit constraints and productivity in Peruvian agriculture*. Working Paper, UC Davis, 07-005. Giannini Foundation of Agricultural Economics.
- Gul, F., & Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69, 1403–1436.
- Gul, F., & Pesendorfer, W. (2004). Self-control and the theory of consumption. *Econometrica*, 72, 119–158.
- Gurley, J. G., & Shaw, E. S. (1955). Financial aspects of economic development. *American Economic Review*, 45, 515–538.
- Guth, R. A. (2008). Giving a lot for saving a little (Gates foundation to invest in programs to help collect deposits for the world's poor). *Wall Street Journal*, July 31, A11.
- Hansmann, H. (1996). *The Ownership of Enterprise*. Cambridge: Belknap Press.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42, 1009–1055.
- Hertzberg, A., Liberti, J. M., & Paravisini, D. (2008). *Public information and coordination: Evidence from a credit registry expansion*. Working Paper.
- Hewitt Associates. (2007). Survey findings: Hot topics in retirement. Lincolnshire, IL: Hewitt Associates.
- Hirschland, M., & Owens, J. (2005). Pricing. In M. Hirschland (Ed.), *Savings services for the poor: An operational guide*. Bloomfield: Kumarian Press.
- Honohan, P. (2008). Cross-country variation in household access to financial services. *Journal of Banking and Finance*, 32, 2493–2500.
- Horioka, C. Y. (2006). Are the Japanese unique? An analysis of consumption and saving behavior in Japan. In S. Garon & P. L. Maclachlan (Eds.), *The ambivalent consumer: Questioning consumption in east Asia and the west* (pp. 113–136). Ithaca: Cornell University Press.
- Ivatury, G., & Mas, I. (2008). *The early experience with branchless banking*. CGAP Focus Note 46.
- Jacoby, H. G., & Skoufias, E. (1997). Risk, financial markets, and human capital in a developing country. *Review of Economic Studies*, 64, 311–335.
- Jain, S., & Mansuri, G. (2003). A little at a time: The use of regularly scheduled repayments in microfinance programs. *Journal of Development Economics*, 72, 253–279.
- Jalan, J., & Ravallion, M. (1999). Are the poor less well insured? Evidence on vulnerability to income risk in rural China. *Journal of Development Economics*, 58, 61–82.
- Jalan, J., & Ravallion, M. (2004). Household income dynamics in rural China. In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.
- Johnston, D., & Morduch, J. (2008). The unbanked: Evidence from Indonesia. *World Bank Economic Review*, 22, 517–537.
- Jones, D. (2009). *Information, preferences and public benefit participation: Experimental evidence from the advance EITC and 401(k) savings*. Working Paper. University of California-Berkeley.
- Kahneman, D., Knetsch, J., & Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98, 1325–1348.
- Karlan, D. (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, 95, 1688–1699.
- Karlan, D. (2007). Social connections and group banking. *Economic Journal*, 117, F52–F84.
- Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2009). *The salience of savings*. Working Paper. Yale University, California Institute of Technology, Harvard University, and Dartmouth College.
- Karlan, D., Mobius, M., Rosenblat, T., & Szeidl, A. (2008). *Measuring trust in Shantytowns in Peru*. Working Paper.
- Karlan, D., Mullainathan, S., & Zinman, J. (2008). *Behavioral savings experiments in Peru and the Philippines*. Working Paper. Innovations for Poverty Action.

- Karlan, D., Mullainathan, S., & Zinman, J. (2009). *Interest rate elasticities in the long run for microcredit borrowers in Mexico*. Working Paper.
- Karlan, D., & Valdivia, M. (2009). *Teaching entrepreneurship: Impact of business training on microfinance institutions and clients*. Working Paper. Yale University Economic Growth Center.
- Karlan, D., & Zinman, J. (2008a). Credit elasticities in less-developed economies: Implications for microfinance. *American Economic Review*, 98, 1040–1068.
- Karlan, D., & Zinman, J. (2008b). *Expanding credit access: Using randomized supply decisions to estimate the impacts*. *Review of Financial Studies* (forthcoming).
- Karlan, D., & Zinman, J. (2009a). *Expanding microenterprise credit access: Using randomized supply decisions to estimate the impacts in Manila*. Working Paper. Yale University, Dartmouth College, and Innovations in Poverty Action.
- Karlan, D., & Zinman, J. (2009b). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica* (forthcoming).
- Khandker, S. R. (2005). Micro-finance and poverty: Evidence using panel data from Bangladesh. *World Bank Economic Review*, 19, 263–286.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112, 443–477.
- Landry, C. E., Lange, A., List, J. A., Price, M. K., & Rupp, N. G. (2006). Toward an understanding of the economics of charity: Evidence from a field experiment. *Quarterly Journal of Economics*, 121, 747–762.
- Levine, R. (2005). Finance and growth: Theory and evidence. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth*. Netherlands: Elsevier Science.
- Ligon, E., Thomas, J. P., & Worrall, T. (2002). Informal insurance arrangements with limited commitment: Theory and evidence from village economies. *Review of Economic Studies*, 69, 209–244.
- Loury, G. (1981). Intergenerational transfers and the distribution of earnings. *Econometrica*, 49, 843–867.
- Lund, F., & Nicholson, J. (2004). *Chains of production, ladders of protection: Social protection for workers in the informal economy*. Durban: University of Natal School of Development Studies.
- Lusardi, A. (2007). *Household saving behavior: The role of literacy, information and financial education programs*. Working Paper No. 13824. National Bureau of Economic Research.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.
- Madrian, B., & Shea, D. (2002). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, 116, 1149–1187.
- Manski, C. F. (2003). Identification problems in the social sciences and everyday life. *Southern Economic Journal*, 70, 11–21.
- Manski, C. F. (2005). Optimal search profiling with linear deterrence. *American Economic Review (AEA Papers and Proceedings)*, 95, 122–126.
- McIntosh, C. (2008). Estimating treatment effects from spatial policy experiments: An application to Ugandan microfinance. *Review of Economics and Statistics*, 90, 15–28.
- McKenzie, D., & Woodruff, C. (2006). Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises. *Economic Development and Cultural Change*, 55, 3–42.
- McKenzie, D., & Woodruff, C. (2008). Experimental evidence on returns to capital and access to finance in Mexico. *World Bank Economic Review*, 22, 457–482.
- McKinnon, R. (1973). *Money and capital in economic development*. Washington: The Brookings Institution.
- Miranda, M. (1991). Area crop yield insurance reconsidered. *American Journal of Agricultural Economics*, 83, 650–665.
- Modigliani, F. (1986). Life cycle, individual thrift, and the wealth of nations (Nobel Address). *American Economic Review*, 76, 297–313.
- Mohapatra, S., Ratha, D., Vijayalakshmi, K. M., & Xu, Z. (2006). *Remittance trends 2006*. World Bank Migration and Development Brief 2.
- Morduch, J. (1994). Poverty and vulnerability. *American Economic Review (AEA Papers and Proceedings)*, 84, 221–225.
- Morduch, J. (1998). *Does microfinance really help the poor? New evidence on flagship programs in Bangladesh*. Working Paper. MacArthur Foundation Project on Inequality. Princeton University, draft.

- Morduch, J. (1999). Between the market and state: Can informal insurance patch the safety net? *World Bank Research Observer*, 14, 187–207.
- Morduch, J. (2004). Consumption smoothing across space: Testing theories of risk-sharing in the ICRI-SAT study region of South India. In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.
- Morduch, J. (2006). Micro-insurance: The next revolution? What have we learned about poverty? In A. Banerjee, R. Benabou & D. Mookherjee (Eds.), *Understanding poverty*. New York: Oxford University Press.
- Mullainathan, S., & Shafir, E. (2008). *Savings policy & decision-making in low-income households*. Working Paper.
- Munshi, K., & Rosenzweig, M. (2007). *Why is mobility in India so low? Social insurance, inequality and growth*. Working Paper. Brown University Department of Economics.
- O'Donoghue, T., & Rabin, M. (1999). Doing it now or doing it later. *The American Economic Review*, 89, 103–121.
- Orozco, M. (2007). *Estimating global remittance flows: A methodology*. IFAD.
- Patten, R., & Rosengard, J. (1991). *Progress with profits: The development of rural banking in Indonesia*. San Francisco: International Center for Economic Growth/HIID.
- Pauly, M. (1968). The economics of moral hazard: Comment. *American Economic Review*, 58, 531–537.
- Pauly, M., Blavin, F. E., & Meghan, S. (2008). *Is there a market for voluntary health insurance in developing countries?* Working Paper.
- Phelps, E. S., & Pollak, R. A. (1968). On second-best national savings and game-equilibrium growth. *Review of Economic Studies*, 35, 185–199.
- Pitt, M., & Khandker, S. (1998). The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter? *The Journal of Political Economy*, 106, 958–996.
- Platteau, J. (2000). *Institutions, social norms, and economic development*. London: Harwood.
- Platteau, J. (2004). *Egalitarian norms and economic growth, institutions, social norms and economic development*. Chur: Harwood Academic Publishers.
- Popkin, S. (1979). *The rational peasant: The political economy of rural society in Vietnam*. Berkeley: University of California Press.
- Radermacher, R., Dror, I., & Noble, G. (2006). Challenges and strategies to extend health insurance to the poor. In C. Churchill (Ed.), *Protecting the poor: A microinsurance compendium*. Geneva: Munich Re and ILO.
- Rajan, R. G., & Zingales, L. (1998). Financial dependence and growth. *American Economic Review*, 88, 559–587.
- Ratha, D., Mohapatra, S., Vijayalakshmi, K. M., & Xu, Z. (2008). *Revisions to remittance trends 2007*. World Bank Migration and Development Brief 5.
- Robinson, M. S. (2001). *The microfinance revolution*. The World Bank and Open Society Institute.
- Roodman, D., & Morduch, J. (2009). *The impact of microcredit on the poor in Bangladesh: Revisiting the evidence*. Working Paper. Center for Global Development and Financial Access Initiative.
- Rose, E. (1999). Consumption smoothing and excess female mortality in rural India. *Review of Economics and Statistics*, 81, 41–49.
- Rosenberg, R. (2002). *Microcredit interest rates*. CGAP Occasional Paper.
- Rosenzweig, M. R. (1988). Risk, implicit contracts, and the family in rural areas of low income countries. *Economic Journal*, 98, 1148–1170.
- Rosenzweig, X. (2001). Savings behaviour in low income countries. *Oxford Review of Economic Policy*, 17, 40–54.
- Rutherford, S. (2000a). *The need to save, the poor and their money*. New Delhi: Oxford University Press.
- Rutherford, S. (2000b). *The poor and their money*. New Delhi: Oxford University Press.
- Samphantharak, K., & Townsend, R. M. (2008). *Households as corporate firms: Constructing financial statements from integrated household surveys*. Working Paper. UC-San Diego IRPS and University of Chicago Department of Economics.
- Savage, K., & Harvey, P. (2007). *Remittances during crises: Implications for humanitarian response*. Humanitarian Policy Group of the Overseas Development Institute.

- Scott, J. C. (1976). *The moral economy of the peasant: Rebellion and subsistence in Southeast Asia*. New Haven: Yale University Press.
- Shaw, E. S. (1973). *Financial deepening in economic development*. New York: Oxford University Press.
- Skees, J., Varangis, P., & Larson, D. (2004). Can financial markets be tapped to help poor people cope with weather risks? In S. Dercon (Ed.), *Insurance against poverty*. Oxford: Oxford University Press.
- Soman, D. (2003). The effect of payment transparency on consumption: Quasi-experiments from the field. *Marketing Letters*, 14, 173–183.
- Stiglitz, J. (1990). Peer monitoring and credit markets. *World Bank Economic Review*, 4, 351–366.
- Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71, 393–410.
- Strøtz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23, 165–180.
- Thaler, R. (1990). Anomalies: Saving, fungibility and mental accounts. *Journal of Economic Perspectives*, 4, 193–205.
- Thaler, R., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112, S164–S187.
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4, 199–214.
- Thaler, R., & Sunstein, C. (2003). Libertarian paternalism. *American Economic Review (Papers and Proceedings)*, 93, 175–179.
- Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- Tirole, J. (2005). *The theory of corporate finance*. Princeton: Princeton University Press.
- Townsend, R. (1978). Intermediation with costly bilateral exchange. *Review of Economic Studies*, 45, 417–425.
- Townsend, R. (1983). Financial structure and economic activity. *American Economic Review*, 73, 895–911.
- Townsend, R. (1994). Risk and insurance in village India. *Econometrica*, 62, 539–591.
- Townsend, R. (1995). Financial systems in Northern Thai villages. *Quarterly Journal of Economics*, 110, 1011–1046.
- Townsend, R., & Ueda, K. (2006). Financial deepening, inequality, and growth: A model-based quantitative evaluation. *Review of Economic Studies*, 73, 251–293.
- Tripathi, S. (2006). Microcredit won't make poverty history (Small loans can empower women, argues Salil Tripathi—But don't expect them to help the very poorest, no matter what the Nobel judges say). *Guardian Unlimited* (online).
- Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in Northern Nigeria. *Review of Economic Studies*, 61, 495–526.
- Udry, C., & Anagol, S. (2006). The return to capital in Ghana. *American Economic Review*, 92, 388–393.
- United Nations Advisors Group on Inclusive Financial Services. (2008). *Key messages*. New York: United Nations Capital Development Fund.
- Von Pischke, J. D. (1978). Towards an operational approach to savings for rural developers. In J. D. Von Pischke, D. Adams & G. Donald (Eds.), *Rural financial markets in developing countries: Their use and abuse*. Baltimore: Johns Hopkins University Press.
- Wright, G., & Mutesasira, L. (2001). *The relative risks of savings to poor people*. MicroSave Briefing Note 6.
- Wydick, B. (1999). Can social cohesion be harnessed to repair market failures? Evidence from group lending in Guatemala. *Economic Journal*, 109(457), 463–475.
- Yunus, M. (2006). *Nobel lecture*. Norway: Oslo.
- Yunus, M. (2008). *Creating a world without poverty: Social business and the future of capitalism*. New York: Public Affairs.

Population and Health Policies*

T. Paul Schultz

Department of Economics, Yale University, Box 208269, New Haven, Connecticut 06520-8269, USA

Contents

1. Introduction	4786
2. Historical Transition of Death and Birth Rates and Population Policies	4789
2.1 International population policy assistance: Levels and trends	4792
3. Stylized Facts Related to Health, Fertility, and Development	4797
4. Framework for Studying the Determinants and Consequences of Health	4800
4.1 Estimating without bias health production functions with endogenous inputs	4802
4.2 Estimating the marginal product of health human capital	4804
4.3 Parent investment in children in response to their children's initial endowments	4805
4.4 Spillovers from health human capital beyond the individual and family	4810
4.5 Phases of the mortality transition and implications for health inequality	4812
5. Macroeconomics Evidence on Health Determinants and Economic Growth	4813
5.1 Macroeconomic evidence of the impact of income on health	4816
5.2 Alternative indicators of health status and their economic consequences	4818
6. Microeconomics of Health and Development: Individuals and Households	4820
6.1 Height, weight, body mass index, and birth weight as indicators of health stocks	4822
6.2 Heterogeneity of height as a measure of the stock of health human capital	4828
6.3 Pathways from fetal and early child development to adult productive life span	4840
7. Macroeconomics of Fertility and Development: Diminishing Returns to Labor, Human Capital and Savings, and Changing Age Composition	4844
8. Microevidence of Determinants of Fertility and Consequences of Policy on Fertility	4847
8.1 Income and price effects affecting the quantity and quality of children	4848
8.2 Family planning policies and fertility: Cross-country relationships	4850
8.3 Estimating the cross-effects of fertility variation on family lifetime outcomes	4853
8.4 A social experiment in family planning and reproductive health: Bangladesh	4855
9. Internal Migration, Demographic Transition, and Evaluation of Policy	4857
10. Conclusions and Policy Directions	4861
End Notes	4865
References	4871

Abstract

The program evaluation literature for population and health policies is in flux, with many disciplines documenting biological and behavioral linkages from fetal development to late life mortality, chronic disease, and disability, though their implications for policy remain uncertain. Both macro- and microeconomics seek to understand and incorporate connections between

economic development and the demographic transition. The focus here is on research methods, findings, and questions that economists can clarify regarding the causal relationships between economic development, health outcomes, and reproductive behavior, which operate in many directions, posing problems for identifying causal pathways. The connection between conditions under which people live and their expected life span and health status refers to "health production functions." The relationships between an individual's stock of health and productivity, well-being, and duration of life encompasses the "returns to health human capital." The control of reproduction improves directly the well-being of women, and the economic opportunities of her offspring. The choice of population policies may be country specific and conditional on institutional setting, even though many advances in biomedical and public health knowledge, including modern methods of birth control, are now widely available. Evaluation of a policy intervention in terms of cost effectiveness is typically more than a question of technological efficiency, but also the motivation for adoption, and the behavioral responsiveness to the intervention of individuals, families, networks, and communities. Well-specified research strategies are required to address (1) the economic production of health capacities from conception to old age; (2) the wage returns to increasing health status attributable to policy interventions; (3) the conditions affecting fertility, family time allocation, and human capital investments; and (4) the consequences for women and their families of policies which change the timing as well as number of births.

JEL classifications: D13, I18, J13, O12

Keywords

health
fertility and family planning
biology of health human capital
economic development

1. INTRODUCTION

Population policies are primarily a response to the anticipated consequences of fertility and mortality, and secondarily to internal and international migration that also modify the size, age composition, and regional distribution of the population. It is not novel that economists, especially macro- and micro-oriented economists who tend to examine empirically different types of data, country aggregates or individual and household observations, have not always reached a consensus on population issues. However, recognizing the differences between these research traditions and their findings in this chapter may help formulate working hypotheses that warrant more study, and may guide research on and empirical testing of these hypotheses in a common framework to close some of the existing gaps in our knowledge.

One underlying development that motivates much population policy research, directly or indirectly, is the demographic transition that involves first a decline in age-specific mortality rates, and then is generally followed by a decline in fertility rates.

During the interval between the onset of the decline in mortality and the sustained and substantial decline in fertility, the natural rate of population growth tends to increase and the age composition of the population changes. This may create opportunities or imbalances in the aggregate economy for which social welfare policies are sometimes proposed to improve prospects, typically through the introduction of subsidies, taxes, transfers and regulatory policies to deal with externalities. These changes in mortality and fertility are related to the reallocation of family resources over its life cycle, and these interdependent behavioral responses of women and men, parents and children, provide the core microeconomic issues for study. Without an understanding of how people respond to the provision of new health opportunities and means for controlling births, it is difficult to discuss the tradeoffs on which population policies seek traction. Without knowledge of the technical features of health production functions, and the marginal health gains associated with the use of health-related inputs and behaviors, policies affecting health and well-being will be difficult to evaluate. Biological issues also affect how reproduction is modeled and technological change in birth control is viewed from a policy perspective.

The average length of life in the world has approximately doubled from the start of the nineteenth to the start of the twenty-first centuries, from 30–35 years to 60–70 years, with the recent notable exception being several states in Southern and Eastern Africa where life expectancy has declined due to the HIV/AIDS epidemic. This increase in length of life has paralleled the spread of modern economic growth and the increase in household personal consumption per adult (Kuznets, 1966). But the connections between the economic conditions under which people live and their longer lives, on the one hand, and the relationship between the improved health status of adults and their economic productivity as workers, on the other hand, are two possible causal relationships underlying this covariation of life span and economic growth. Understanding these causal relationships could inform the choice of population policies related to health, family planning, and migration, and improve the basis for predicting future economic development. A goal of this chapter is therefore to describe what we are learning about these underlying causal relationships, and how social scientists are improving the empirical measurement of these causal relationships that could become the basis for better development policy.

Evidence on these relationships is accumulating in a variety of disciplines and sub-fields of economics, and a number of emerging hypotheses merit refinement and concerted empirical study to test the magnitude of behavioral and technical responses, to determine which biological and behavioral pathways are involved in these responses, and to assess longer-run consequences of programs and policies after individuals and families reallocate their lifetime resources. Household sample survey data are an important resource for this research, especially panel surveys which interview the same individuals and families repeatedly over time, to assess more rigorously causal relationships

affecting behavior and welfare. These empirical assessments are generally premised on exclusion restrictions embedded in implicit structural models. The biological and behavioral processes underlying the current improvement in health status are complex, with long-gestating lags linking the growth of the fetus and early childhood biological development, all the way to late life mortality, disability, and health status, as well as impacting intermediate observable outcomes such as cognitive achievements (IQ), schooling, productivity, fertility, and other forms of behavioral adaptation to local environments and policy conditions. The political economy governing the performance of social organizations affecting health and schooling may also be impacted. There is substantial uncertainty about even the short-run effects of social policy interventions on health, fertility, and population growth, and even less certainty regarding the longer-run effects as can only be inferred from the study of long time series and panel data, which are exceedingly rare.

The agenda for research on population policies requires a simultaneous description of the determinants of the *supply* of public produced health-related services and birth control, and the determinants of private household *demand*s for those services and technologies, including preventive or curative health services, social insurance, subsidies for family planning technologies, schooling attainment of boys and girls, the health and reproductive health content of schooling, etc. Given the complexity of these interdependent behavioral, organizational, and technological processes, some economists propose that social experiments are needed to evaluate reliably the long-run social consequences of population programs and policies. Randomized allocations of program and policy treatments can be especially informative in this field, but may not provide a general basis to forecast policy effects. Experimental studies cannot be extrapolated outside the range of observed samples, and program designs and technologies change over time and across societies. For programs to be extended to new populations, the sources of heterogeneous response to the program treatments become a key issue. Additional methods for policy evaluation may also be useful for analyzing naturally occurring “quasi-experiments,” which tend to impose restrictions on how behavioral and technological relationships occur, yet are necessary to recover estimates of the effects of policy treatments. These combined approaches, experimental and nonexperimental, may reveal the likely effects of population policy on mortality, morbidity, disability, labor productivity, labor supply, and fertility, as well as the life-cycle accumulation of physical and human capital that enhance private well-being, generate significant social externalities, operate as public goods, and spur economic growth. However, this area of population policy research is complicated and cannot yet deliver confident answers to settle many of these central questions confronting policy makers. Many empirical and policy questions related to fertility, women’s human capital, and child quality are addressed in a previous paper (Schultz, 2008a), and therefore the current chapter focuses first and more thoroughly on current health issues.

The outline of the chapter is as follows. Section 2 surveys the historical changes in demographic rates. Section 3 reviews some stylized facts about health and fertility, and economic development. Section 4 outlines a framework for studying health determinants and consequences. Section 5 surveys the macroeconomic evidence on health and growth, which has relied largely on cross-country comparisons, whereas Section 6 surveys the microeconomic evidence on health and development, where households and communities are the units of observation, and the primary goals are to estimate health production functions and the productive labor returns to health human capital. Section 7 turns to fertility, and focuses on the macroeconomic evidence of the consequences of fertility change and specifically its effect on economic growth, whereas Section 8 reviews briefly the microliterature on the determinants of fertility decline and the effects of policy-induced voluntary declines in fertility on the welfare of women, their families and communities that might be attributed to effective family planning and reproductive health programs. Section 9 notes the connections between the demographic transition, development, and internal migration, and the problems it raises for policy evaluation studies. Section 10 concludes.

2. HISTORICAL TRANSITION OF DEATH AND BIRTH RATES AND POPULATION POLICIES

The demographic transition occurred first in what are today high-income countries, when age-specific death rates started to decline gradually after about 1750, whereas sustained mortality declines in low-income countries may have only started in the 1920s, but proceeded much more rapidly after the Second World War. Consequently, there has been a catch up by the low-income countries, or a convergence in life span after 1930 which emerges more clearly after 1950, and persisted until the 1990s. The decline and convergence in fertility has tended to follow that of child mortality.

There are differences between the age-specific pattern of mortality decline experienced in various historical periods in high- and low-income countries. In the early-industrializing countries, the initial declines in mortality from 1750 to 1875 may have favored slightly adults. Only toward the end of the nineteenth century did the heavy mortality among infants and young children begin to fall rapidly. Where the early agricultural and industrial revolutions attracted a growing fraction of the population to work in the increasingly unhealthy cities, the economic growth associated with urbanization did not immediately lead to increased life spans, at least not in England, Scotland, or the United States (Tanner, 1982). By the end of the nineteenth century, it is widely believed that improvements in urban public health infrastructure was responsible for mitigating the urban health problems associated with rising population densities, specifically through public investments and regulation of sanitation and water supplies, while the growing acceptance of the germ theory of disease provided

individuals and institutions with a motivation to adopt more hygienic practices that could control the spread of communicable infectious diseases. Many factors are attributed a role in the salient declines in infant and childhood mortality rates, including the rising economic standard of living, and in particular the declining price of food relative to wages, the increasing quality and diversity of food in urban markets, the decreasing crowding of urban worker housing, and extending of welfare systems designed to provide a safety net for food and medical care for the poor, disabled, vulnerable widows, women with children to support, and the elderly. Although advances in medical knowledge and public health interventions were probably unimportant in reducing mortality before the beginning of the twentieth century, the impact of medical science became notable after the 1930s (Fogel, 2004; McKeown, 1976, 1979). Until the 1960s, most lives saved by these improvements in health technology, income, and the living environment were those of children, and only thereafter in the industrially advanced countries does the reduction in deaths among older persons, over age 50, add more to the population than the reduction in deaths among children.

As indicated above, a second phase in mortality decline in the early-industrializing countries occurred from roughly 1875 to 1960, as the proportion of children surviving birth and childhood diseases increased substantially, and these improvements in child health conditions from communicable diseases were often associated with a lag to the decline in fertility. Observers hypothesized that fertility was in part responding to the increase in child survival in a homoeostatic manner controlling population growth (e.g., Freedman, 1975). Ultimately, fertility fell from roughly 6 to little more than 2 children per woman, more than enough to offset the effect of the decline in child mortality on population growth, and the number of surviving children per woman declined from approximately 5 to 2 as the rate of natural population growth subsided in a few generations to approximately zero by the end of the twentieth century in Europe, Japan, and areas settled by Europeans.

The post-Second World War baby boom in some countries could be in part attributed to shifts in the timing of fertility, births delayed during the Great Depression and the Second World War occurred during the postwar economic expansion. Completed fertility for a sequence of birth cohorts fell gradually toward replacement levels of about two children per woman in most developed countries. Differentials in fertility across developed countries and within these countries by region and socioeconomic classes gradually narrowed. Life expectancy at birth and lifetime fertility have become more homogeneous across and within more developed countries by the end of the twentieth century.

Infectious diseases were first controlled by isolating the sick (i.e., quarantine), and then by the use of preventive health measures and changing personal hygiene to limit the transmission of disease, and eventually by the development and application of sulfa and antibiotic drugs, whose use to control infectious and parasitic diseases was

accelerated by the involvement of the armed forces in health research during the Second World War. After the war, new public health technologies diffused rapidly to the rest of the world. By the 1960s, medical science increasingly grappled with the non-communicable chronic diseases, slowly reducing death and disability due to cardiovascular disease, respiratory diseases, some forms of cancer. The reduction in fertility has also been associated with a number of other changes in coordinated family life-cycle behavior, including increased female schooling (relative to male), increased female participation in the market labor force, increased human capital investments in children's schooling, and perhaps increased physical savings and life-cycle accumulation of wealth. With the declines in fertility and continuing declines in mortality rates among the elderly, national populations have become older, first in the high-income countries, and now emerging as an issue in a growing number of low-income countries where birth rates first started to fall in the 1960s and 1970s.

Beyond Europe, Japan and countries settled by Europeans the secular declines in mortality occurred later but evolved thereafter more rapidly, though incomplete registration of death rates often leaves in doubt the precise timing of these changes, how cause-specific deaths evolved, and the economic, social, and institutional developments that governed this epidemiological transition. The sustained slow declines in mortality in low-income countries begins to be evident after the flu epidemic of 1918–1919, and probably accelerated in the late 1940s as public health measures were coordinated to control infectious and parasitic diseases by WHO, and supported by technical assistance programs, such as Truman's "Point Four program for economically undeveloped countries." These declines in mortality in low-income countries initially benefited largely infants and children, leading to more rapid population growth and more youthful populations in the 1950s, without clearly associated improvements in labor productivity or growth in national income (Acemoglu & Johnson, 2007). By comparison, the early gradual declines in mortality in Western Europe seem to have been related to the initial agricultural and industrial revolutions, which raised wages relative to the price of food, improved nutrition, adult health, and probably increased labor productivity (Fogel, 2004).

If the increase in life expectancy is accepted as a summary measure of survival and population health, this indicator of life span has been increasing in the countries with the longest life span and most reliable mortality statistics by about a quarter of a year with the passage of each calendar year. This trend can be traced backwards for almost 200 years (Oeppen & Vaupel, 2002). In low-income countries life expectancy at birth has been increasing at twice this rate, or about a half year per calendar year, from 1940 to 1990, after which this pattern of convergence in life span between high- and low-income countries has been interrupted by the HIV/AIDS epidemic in sub-Saharan Africa (Strauss & Thomas, 2008).

Longer-lived parents may need more support in their old age, and since they have chosen to have fewer surviving children to provide such support, this development

may lead to an increase in life-cycle savings in the form of physical or financial assets, or to increased pension systems and retirement funds managed by governments or offered through employers. Since there is little exogenous variation across individuals within countries in public policies affecting the tax-transfer incentives to work and save, microestimates from individual data of how the changing retirement program incentives affect labor supply behavior, or crowd out private savings by workers, is not resolved. Comparisons of tax-transfer incentives across a handful high-income country suggest high taxes on the earnings of elderly and generous pensions could contribute to the decline in elderly labor force participation, or in other words the reduced age of retirement despite the increase in health, adding to the fiscal burden of most public pension and health care programs (Gruber & Wise, 1995, 2004).

Becker and Barro (1988) postulate that more generous retirement pensions are responsible for the decline in fertility in more developed countries in the twentieth century. But this correlation could also signal that populations that want fewer children also demand pensions from their governments. In either case, funding public pension system from taxes on the earnings of workers has often led to “pay as you go” systems (i.e., Ponzi scheme), in which the earnings of younger generations of workers are taxed to support part of the pensions of the elderly. This is one reason advanced for nations to subsidize childbearing and investment of public resources in the health and schooling of children, in order to increase the size and productivity of future generations, and thereby reduce the future tax rate on labor required to finance “pay as you go” old-age pensions (Gruber & Wise, 2004). A reduction in the tax rate on labor is often assumed to encourage more work, effort, and savings among the young, but the empirical documentation of the magnitude of these behavioral responses remains uncertain in high- or low-income countries. In low-income countries, population policies are associated with subsidies for birth control and health. International development assistance in the population field has recently increased, predominantly for health, and much of this chapter deals with research on health and its implications for development policy, and fertility is discussed later and in another review (Schultz, 2008a).

2.1 International population policy assistance: Levels and trends

According to UNFPA/NIDA (2009) resource flows project, the funding of population assistance activities by the OECD countries in family planning, reproductive health, and other health activities in less developed countries has increased in the last 11 years from \$1883 million (1993 US dollars) in 1996 to \$5290 million in 2006. Population assistance as a percent of total official development assistance increased from 2.5% to 6.1% in this period. The functional composition, geographic focus, and institutional delivery system have all changed appreciably. In 1996, 37% of population assistance was allocated to family planning services, which increased to 43% in 1998, and thereafter declined to 5% in 2006, leaving donor real expenditures on family planning at

about a half the level they were in 1996. Reproductive health services as a share of population assistance have gradually declined from 33% in 1996 to 20% in 2006, yet they more than doubled in 1993 dollars. Research, program evaluation, and data collection, such as the demographic health surveys (DHS), have declined from 14% of the total to 5%, holding approximately constant in real dollar terms. The remainder of population assistance is devoted to the prevention and treatment of sexually transmitted diseases and HIV/AIDS and presumably other unspecified diseases, and it increased as a share of population assistance from 16% to 70%, growing 15-fold in 1993 dollars. Sub-Saharan Africa received 28% of population assistance in 1996, and its share increased to 35% by 2006, whereas Asia and the Pacific received 24% in 1996 and 14% by 2006, and Latin America and the Caribbean received 13% in 1996 and 5% in 2006, with the regional remainder to Western Asia and North Africa, Eastern and Southern Europe, while the global/interregional share increased from 26% to 40% of population assistance. NGOs were the largest channel of distribution of population assistance in 1996, whereas the importance of bilateral and multilateral organizations was the dominant source of funding by 2006 (<http://www.resourceflows.org/index.php/articles/288>, Tables 1-6; 3 March 2009).

Comparable figures for expenditures on health and family planning activities by the governments in less developed countries and by private households and local NGOs in these countries are not readily available to complete the accounting of social resources allocated to these population activities. WHO and World Bank report data for total public (government) and private (household) expenditures on health, some of which are reported in Table 1 for 1990 and 2000. There is much variation across countries in the fraction of GDP allocated to health, and the share of all health expenditures in the public (vs the private) sector differs markedly across countries at all levels of development, for example, India, China, USA, and Norway. It would be informative to disaggregate public expenditures on health among public health (preventive) services, out patient care, and hospital (curative) services, as reported in the IMF *Government Financial Statistics*. But these annual figures from IMF generally omit state and local public health programs, and even for the central government expenditures these distinctions are missing for many countries. To account for private expenditures on health, a comprehensive representative household survey including consumption and expenditure data is required, such as the living standard measurement surveys of the World Bank (<http://www.worldbank.org/LSMS/>). But these surveys are only available for an unrepresentative subset of countries and years. Out of pocket private expenditures on health as estimated by WHO (2003) excludes health payments to enterprises which deliver medical benefits to their employees, insurance premiums, or other prepayments and reimbursements to third-party payers. Moreover, it is rarely possible to allocate private or public health expenditures according to function, for example, family planning versus

Table 1 Expenditures on health as percent of GDP and percent spent in public sector: 1990 and 2000

Region/country	Percent of GDP spent on health		Percent of health in public sector	
	1990	2000	1990	2000
<i>1. South Asia</i>				
Bangladesh	3.2	4.1	44	37
India	6.0	4.9	22	18
Sri Lanka	3.7	3.7	49	49
<i>2. Latin America and Caribbean</i>				
All	4.0	NA	60	NA
Brazil	4.2	8.3	67	41
Chile	4.7	3.2	72	78
Cuba	NA	7.1	NA	86
Mexico	3.2	7.3	50	42
<i>3. East Asia</i>				
China	3.5	5.4	60	37
Indonesia	2.0	2.7	35	22
Thailand	5.0	3.7	22	57
<i>4. Sub-Saharan Africa</i>				
All	4.5	NA	56	NA
Cote d'Ivoire	6.2	2.8	52	36
Ghana	3.5	4.1	49	54
Kenya	4.3	8.8	63	27
Nigeria	2.7	1.7	44	29
<i>5. Developed countries</i>				
All	9.2	NA	61	NA
Canada	9.2	9.0	74	72
Norway	7.4	7.6	95	86
Germany	8.0	10.6	73	75
Japan	6.5	7.7	74	77
United States	12.7	13.1	44	44

NA, not available from source.

Sources: World Bank (1993), Table A-9 and United Nations Development Programme (2000), Table 6.

reproductive health, or for specific diseases. Finally, no estimates were found of the value of consumer time spent obtaining public and private health services, or the opportunity cost of the time of household member who care for their own sick (UNFPA/NIDA, 2009).

Another perspective on trends and differentials in population policies may be inferred from what governments indicate are their attitudes and priorities. The United Nations (2003) collected questionnaires from governments on population policies and their responses are summarized in Table 2 by a country's level of development

Table 2 Percentage of countries by development level and year according to their views on population policies

1. Fertility policies				
Objective	Raise	Maintain	Lower	No intervention
<i>Least developed countries (N = 49, 1996)</i>				
1976	2	5	14	79
1996	0	6	65	29
2004	0	8	69	22
<i>Less developed countries (N = 146, 1996)</i>				
1976	5	10	34	50
1996	8	10	56	26
2004	8	11	57	24
<i>More developed countries (N = 48, 1996)</i>				
1976	21	21	0	59
1996	33	8	2	56
2004	44	10	0	46
2. Access to contraceptive methods				
Objective	Limits	No support	Indirect support	Direct support
<i>Least developed countries</i>				
1976	10	33	14	43
1996	0	6	6	88
2004	0	2	10	88
<i>Less developed countries</i>				
1976	6	21	9	64
1996	1	10	8	82
2004	0	5	10	85

Continued

Table 2 Percentage of countries by development level and year according to their views on population policies—Cont'd

<i>More developed countries</i>				
Objective	Major	Minor	Not a concern	
<i>Least developed countries</i>				
1996	48	27	24	
2004	63	24	12	
<i>Less developed countries</i>				
1996	53	28	19	
2004	61	29	10	
<i>More developed countries</i>				
1996	27	36	36	
2004	33	48	19	
4. Accept mortality levels				
Objective	Acceptable		Unacceptable	
<i>Least developed countries</i>				
1976	5		95	
2004	4		96	
<i>Less developed countries</i>				
1976	24		76	
2004	35		65	
<i>More developed countries</i>				
1976	79		21	
2004	69		31	

Source: United Nations (2004), Tables 6–8, 10, 17, 19.

and over time. Virtually all least and less developed (LLD) countries view mortality rates as unacceptably high, as they do infant, child, and maternal mortality (not reported here), whereas this attitude toward mortality is more muted among the more developed (MD) countries. The majority of LLD countries have in the last

decade sought to lower their fertility, while 44% of the MD countries now consider raising their fertility as beneficial. Policies that support directly or indirectly access to contraceptive methods are approved by 95% of the LLD countries by 2004, a marked increase over the last 25 years, and by 86% of the MD countries, despite their public preference for higher fertility. The timing of fertility is also of growing concern to governments. The high level of adolescent fertility is a worry in more than 60% of the LLD countries by 2004, and in a third of the MD countries. A final aspect of population policy involves international migration, though it is not pursued in this chapter. From the UN survey it appears that a growing share of countries at all levels of development, constituting a quarter by 2004, favor policies that reduce emigration and immigration, and only 2% of MD countries favor policies to raise emigration and immigration.

Although a consensus has been reached among LLD countries in the last quarter century in favor of government support of contraceptives and efforts to reduce fertility, MD countries with low or negative rates of population growth have begun to favor policies to raise fertility. The UN survey makes no attempt to assess the willingness of the MD countries to underwrite pronatalist incentives to avoid population decline or mitigate the previously noted fiscal consequences of population aging. However, the inclination to favor higher fertility in MD countries has not reduced public support for voluntary contraception, or the preference for lower adolescent fertility, or increased support of immigration. Improved health and reduced mortality is valued universally, though the priority assigned to improved health appears to differ by levels of development. Fertility control in terms of contraception is widely valued as a public good. The social externalities of reducing fertility in low-income countries may be valued positively, whereas the positive social externalities of increasing fertility in MD countries are beginning to be appreciated. There is no consensus among economists on how to evaluate the magnitude of population externalities, or how much people are privately willing to pay to reduce mortality, or change fertility, or modify future rates of population growth through their provision of services and design of related tax-transfer policies. These should be active areas of research.

3. STYLIZED FACTS RELATED TO HEALTH, FERTILITY, AND DEVELOPMENT

When the registration of deaths by age and sex is relatively complete, and the underlying population is accurately enumerated by periodic censuses, demographers can assemble a “life table” to describe the pattern of mortality. This table is often summarized in terms of the average duration of life for a cohort born today that is assumed to experience this year’s age-specific death rates over their hypothetical lifetime. It is

based on the concept of a “synthetic” cohort for whom the current age-specific death rates persist indefinitely into the future, not the outcome experienced by any real cohort of births living out their lives. Gradually after the Second World War an increasing number of such life tables have been estimated for various populations and periods. At about the same time a consensus developed on how to best calculate national income. This led to the discovery that the current era of “modern economic growth” had sustained growth in *per capita* real output of 1–3% per year over a century or longer, which qualified as a unique historical period (Kuznets, 1966). Demographic evidence in more developed countries, as noted above, suggested that the expectation of life at birth was also increasing over time after the start of the nineteenth century by a quarter of a year per year (Oeppen & Vaupel, 2002). Both indicators of *per capita* national income and of life span are measured with error and may not represent the ideal concepts of economic well-being or health status. The long-run advancement in both income and health at the aggregate level of the nation has nonetheless become an accepted empirical regularity. Although income and consumption measures from household sample surveys provide an auxiliary microestimate of *per capita* income and consumption to compare to national income estimates, there is no analogous measure of health status at the individual, family, or even community level to compare with life expectancy calculated for a nation. Nor, as I argue later, is it clear that life expectancy at birth is conceptually a good summary measure of the health objective of society or a satisfactory indicator of the productive returns associated with secular improvements in health.

For many countries with incomplete death registrations, methods have been developed for estimating indirectly life tables. First, the age composition of the population from a census provides some information on historic trends in birth rates and age-specific death rates, and repeated censuses set limits on the likely range of vital rates consistent with intercensus population growth rates and age compositions (Coale, 1972; United Nations, 1967). Second, from the collection of life tables from higher-income countries, for which the population and mortality data are thought to be relatively accurate, similar life tables are clustered into regional model classifications, possibly influenced by distinctive diseases and resulting regimes of mortality prevalent in these geographically defined areas. Once a country is assigned to a regional class of life tables, the level of mortality within that class can be approximated in that country on the basis of estimated child mortality rates, derived generally from the responses of women to a census or representative survey question regarding the number of children they have born and number of their children surviving (Brass, 1975; Coale & Demeny, 1966). Imputations of life tables for less developed countries based on these indirect methods have been promulgated by the Population Division of the United Nations and the World Bank, and they are essentially extrapolations from countries with more reliable data sources.

National income accounts and household surveys periodically tell inconsistent tales regarding the occurrence of economic growth, but at least for income and the prevalence of poverty there are generally two empirical approaches to compare (Deaton, 2003). Health status, morbidity, and mortality are not yet subject to systematic comparisons based on macro- and microindicators, although height and weight are increasingly considered as providing critical information on nutrition, health status, and well-being at various levels of aggregation over time. Finally, to facilitate comparisons across countries, these national income estimates measured in terms of local currencies are revalued at common international prices. Early studies relied on official foreign exchange (FX) rates to convert local currency income into common dollar units of purchasing power, but this FX approach excluded cross-country variation in the prices for untraded goods and services and their share in income. Purchasing power parity (PPP) price indices were therefore constructed for a few countries at various levels of development where coordinated price surveys were conducted, and then these implied domestic price exchange rates for traded and untraded goods and services were extrapolated to the majority of countries where such comparable price surveys did not yet exist (Summers & Heston, 1991).

Figure 1 illustrates the problems in interpreting the relationship between health and development. The association between health status (H) and income per adult (Y) is positive. This may happen because increased income leads people to (1) demand directly better health as a consumer good adding utility per year lived and possibly

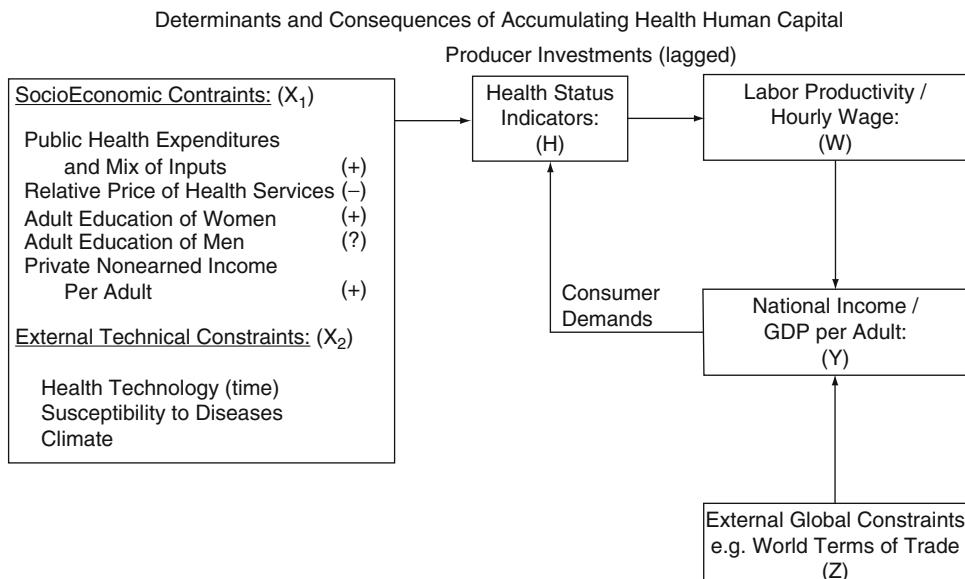


Figure 1 Determinants and consequences of accumulating health human capital.

increasing longevity, or (2) people's increased standard of living shifts consumption patterns which on balance contribute indirectly to health improvements, for example, suppose decreased malnutrition and reduced crowding of housing improved health more than increased smoking and drinking reduced health. Alternatively, improved health status can raise the productivity of people that adds to their market income, which sums into increased national income, unless (1) their market labor supply decreases by a larger percentage than (2) their wage opportunities increase over their lifetime.¹ These national estimates of income and mortality, summarized generally by variations across countries in terms of market income (GDP) and life expectancy, are analyzed in most macroeconomic studies discussed in Section 5. Household surveys provide the data for microeconomic studies on how, on the one hand, exogenous variation in income (due to changes in Z in Figure 1) increases inputs for the production of health (H), and how on the other hand, the exogenous improvements in health status (due to changes in X_2) enhance the economic output of workers (W) and probably boost market income and certainly increase "full" income. These empirical relationships may tell us how an increase in income from an exogenous improvement in a country's terms of trade (Z) contributes to the purchase of more food or better living conditions, to the adoption of healthier behaviors over the life cycle, or to the improved access to medical care and more effective health technologies, which in combination produce a longer life span. It may also tell us how global improvements in shared health technology, X_2 (Acemoglu & Johnson, 2006), or public health expenditures on HIV drugs provided by a philanthropic foundation reduce the consequences of disease and disability, on worker earnings, and aggregate income. Both possible causal relationships are thought to be positive, and may operate therefore to reinforce each other. The observed gross relationship between life span and income might overstate the causal effect operating in either direction. The choice of information that could identify the causal effects operating in one direction, or the other, is controversial. Without the assessment of these causal effects that describe specific pathways, however, it is difficult to evaluate whether health-related inputs foster development as well as health, or how population policies are designed to internalize social externalities associated with the production of health or the modification of fertility.

4. FRAMEWORK FOR STUDYING THE DETERMINANTS AND CONSEQUENCES OF HEALTH

One of the first descriptions of the empirical pattern between health and development is by Preston (1975), who represented health by the expectation of life at birth, and development by *per capita* national income in 1962 dollars (compared by foreign exchange rates). He reports estimates from life tables and national income for 10 countries in the 1900s, and for 38 countries in the 1930s, and for 57 countries in the

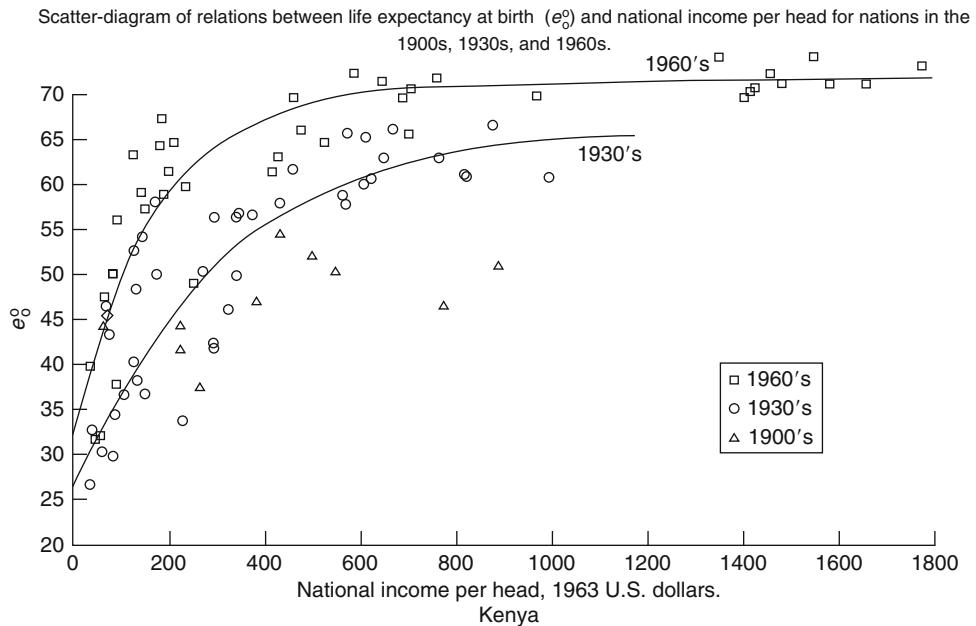


Figure 2 Scatter diagram of relations between life expectancy at birth and national income per head for nations in the 1900s, 1930s, and 1960s (source: Preston, 1975).

1960s.² His observations are plotted in Figure 2. If income is expressed in logarithmic (proportional) terms, the semi-log-linear fit or R^2 is relatively high, 0.885 in 1930s and 0.880 in 1960s. His paper explores the upward shift in the relationship over time in the twentieth century. Had the fitted function remained as it was in the 1930s, the gain in life expectancy would have been substantially smaller than was actually observed in 1960s, suggesting some other factors have also changed, whose inclusion would help to explain the increase in life expectancy, such as advances in the technical knowledge of how to improve health. The residual gains between 1930 and 1960 are disproportionately larger for lower-income countries. Countries with incomes between \$100 and \$500 “experienced” increases in life expectancy of 10–12 years between 1930 and 1960, whereas high-income countries achieved more modest increases, and the upper asymptote approached at very high-income levels increased by less than 5 years.³

Strauss and Thomas (2008) update these patterns by adding World Bank estimates from 2004 in Figure 3. The fitted lines are nonparametric flexible approximations with income now expressed in logarithmic terms.⁴ The increases in life expectancy controlling for income from 1930 to 1960 among countries with \$200 to \$500 dollars of income *per capita* (1995 dollars) is evident, but the gains from 1960 to 2004 are

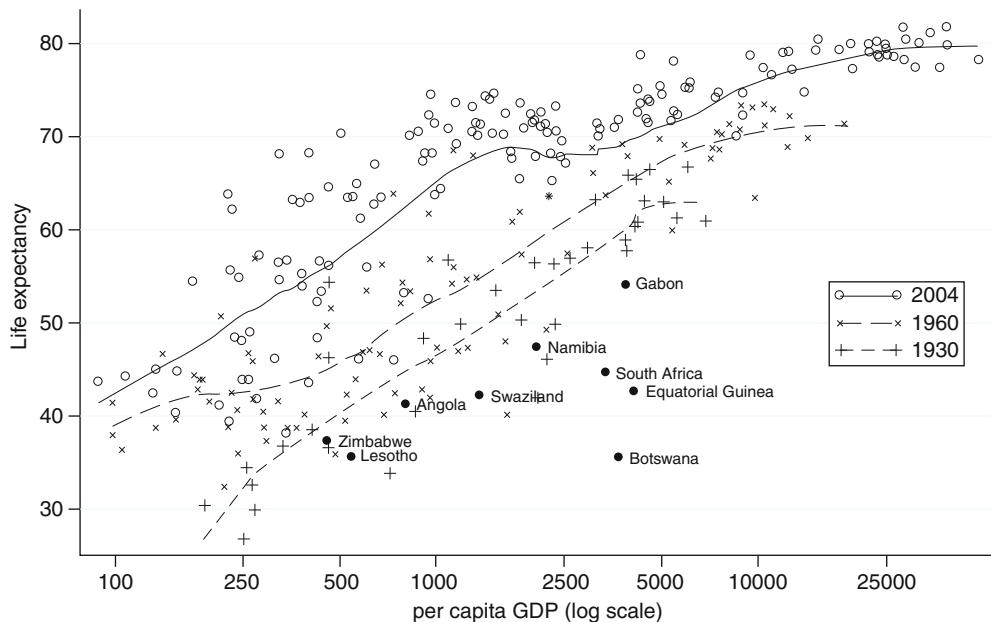


Figure 3 Life expectancy and GDP *per capita*, identifying outliers in 2004.

largest for the middle income countries, between \$500 and \$2000 dollars *per capita*. Even without forcing an upper asymptote to the life expectancy pattern as implied by Preston's exponential function, it is clear that most countries with incomes greater than \$1000 have life expectancy in 2004 in a relatively narrow range between 70 and 80 years. The outliers in 2004 are those with exceptionally low life expectancy, which are named in Figure 3 (Strauss & Thomas, 2008), and are those countries where HIV/AIDS has raised adult mortality to exceptional levels in sub-Saharan Africa, disrupting the past empirical pattern between health and development.⁵

4.1 Estimating without bias health production functions with endogenous inputs

What type of information would help to disentangle from the simple association between health and development the underlying causal links and specific pathways between these variables that could be relevant to the choice of policies that contributes effectively to improvements in health or economic development or both? To identify these causal effects, a variable must be known that affects either health (H) or income (Y) in Figure 1, but does not directly affect the other outcome. The causal effect of private behavior and public institutions and conditions on the production of health can only be evaluated without bias under special circumstances. Typical measures of the association between the use of health inputs (e.g., consumption of drugs) and health

outcomes (e.g., less disabled or living longer) have little information value to the policy maker, because the choice of using the health input may be associated with other unobserved health factors and other intercorrelated omitted personal choices whose health effects are likely to otherwise be attributed to the observed health inputs.

To establish a framework for discussing this problem, assume individuals value good health (H) and other consumption (C):

$$U = U(H, C). \quad (1)$$

An individual's current stock of health human capital is assumed to be produced by a function of the prior period's health endowment (e), current health-related private inputs (I), and exogenous environmental disease conditions (D). Over time D might be modified by cooperative or public health investments and infrastructure, such as water, sanitation, and community disease control programs, but is initially taken as exogenous by the individual or household:

$$H_{ijct} = H(e_{it-1}, I_{ijct}, D_{ct}), \quad (2)$$

where the subscripts i denote the individual, j the household or family, c the community, and t the time period. Because the health-related inputs are expected to respond to prior health endowments (e) that are difficult for researchers to measure comprehensively or introduce into the analysis as exogenous control variables, or local health conditions (D), identifying variables or exclusion restrictions must be observed that cause variation in the demand for health-related inputs that can be assumed unrelated to variation in e or D . The health input demand functions could be portrayed in the following form:

$$I_{ijct} = I(e_{it-1}, X_{ct}, Y_{jt}, D_{ct}; S_{jt}, T_t), \quad (3)$$

where X_{ct} denotes access to and the prices and quality of health inputs that vary by community, Y_{jt} the household's income or lifetime resources, S_{jt} the schooling of those in the household who manage the production of health, such as a mother for her child, and T_t the prevailing medical technology or knowledge of public health that could change with time and customs. These final four variables might be imagined to vary independently of the initial health endowment, e , and public health environment, D , and provide identifying exogenous variation in health input use, required for estimating the causal effect of health input use on the production of health (Eq. 2). For example, suppose a public health program subsidizes the price of a health input at time t in a random subset of the communities c^* , and the subsidy increases the local use of health inputs in these treated communities. Alternatively, a seemingly random exposure across farm communities to crop

pests in period t might reduce current household income, Y_{jt} , and thereby impact the use of health-related inputs, and not otherwise affect health conditions.

As discussed later in Section 5, parameters of the health production function might be recovered from panel observations on population aggregates, such as countries. Factors external to a country may presumably affect the value of a nation's resource endowments, which impacts the income of an average household, but does not otherwise affect the inducements or capacities to invest in health (Pritchett & Summers, 1996). Finally, shifts in locally effective medical and public health technology (T) may also occur from time to time due to research and development, affecting health differentially in different locations depending on the initial prevalence of the specific diseases controlled by the technological change, disseminated under a variety of arrangements involving trade, licensing, shared property rights, or international grant assistance (e.g., Acemoglu & Johnson, 2007).

It should be clear that these types of identifying exclusion restrictions to recover the health production function parameters are likely to be controversial. This is why randomized trials for medical interventions are often supported by the public, despite their complex ethical ramifications, to confirm the effectiveness of promising health inputs and associated technologies when health delivery systems and living conditions differ greatly (e.g., Thomas et al., 2006).

4.2 Estimating the marginal product of health human capital

To allocate efficiently health and social resources, a policy maker must understand a second health relationship: How do improvements in reproducible aspects of health influence the potential productivity of individuals and improve their well-being? One approach to evaluating how health outcomes alter the productive potential of society is to estimate an earnings function conditional on measures of human capital including population health. Although this may be undertaken at the aggregate level of a country or health administrative region, most work in this area builds on the empirical study of earnings function estimated at the individual level, following the research on schooling by Card (1999), Griliches (1977, 1979), and Mincer (1974), in which an approximation for health stocks is included with schooling as a determinant of wages or labor productivity (Schultz, 2003, 2005).

The logarithm of the productivity of labor or the wage rate ($\ln W$) is expected to be greater for a healthier worker (H), as well as for one with more schooling (S), and ability (a), and the supply of other productive factors, such as land and capital per worker (K) that may complement labor and increase its marginal product. Productive technology (T) may shift over time, and it has been hypothesized that recent global improvements in technology have disproportionately increased the productivity of better educated workers, and technological change could also differentially affect the productivity of workers according to their health status or disabilities:

$$\ln W = W(H, S, a; K, T), \quad (4)$$

where local prices of health inputs, X , or the community health environment, D , could possibly identify estimates of the effect of exogenous variation in H on the wage, just as the distance to and quality of local schools could serve as instruments (prices) for schooling, S . The positive correlation between ability (a) and schooling (S) is expected to bias up (positively) the estimated effect of schooling on wages, if ability is omitted from the estimated wage function, and both ability and schooling add to the productivity of workers (Card, 1999; Duflo, 2001; Griliches, 1977, 1979; Schultz, 1988a). Empirically, however, this source of bias does not appear to be a severe problem for policy inferences, which focus on the measurement of the private wage returns to years of schooling, such as estimates based on educational reform instruments for schooling or estimates based on between twin differences in schooling and earnings. This lack of systematic or apparently substantial omitted ability bias may be due to the measurement error associated with schooling, which would in the classical case of random errors introduce an offsetting downward bias in the ordinary least squares (OLS) estimate of the wage return to schooling in the estimation of the wage function (Eq. 4).

There is less empirical research on the returns to aspects of health in a wage function than to schooling, and this issue is complicated by the multifaceted form of health and the greater difficulty of measuring health than schooling for individuals in a general survey. Among the studies that have explored the issue of health consequences on productivity, health is first approximated by body mass index (BMI) and local food prices were used as instruments (Deolalikar, 1988; Strauss, 1986) followed by studies also using disability days, height, menarche (Savedoff & Schultz, 2001; Schultz, 2002, 2003; Schultz & Tansel, 1997; Strauss & Thomas, 2008).⁶ Among the studies I know that estimate wage returns to indicators of individual health, there are few which find the OLS return estimates are positively larger than the instrumental variable (IV) estimates that treat health as endogenous or measured with error, identified by health conditions and services at birthplace or in their family. Thus, the endogeneity of these adult health indicators do not appear to introduce a dominant upward bias in the OLS estimated wage returns to these health outcomes.

4.3 Parent investment in children in response to their children's initial endowments

If parents' investment in their children's human capital responds to the child's initial endowments at birth, either in the form of health, e , or ability, a , the OLS estimates of the child's wage returns to adult health or schooling would be biased by the unobserved investment behavior of parents. For example, assume that parents invest more in their less healthy and less able children, perhaps because they are motivated to equalize their children's life-cycle consumption opportunities or compensate those children

who initially exhibit low endowments (Griliches, 1979, Table 1, p. S61). Then, the omission of these early endowments in estimating the wage function (Eq. 4) leads OLS to underestimate the wage returns to the observed health inputs and indicators of schooling attainment. Griliches in his early review of the literature estimating wage returns to schooling within families and between twins concludes returns to schooling tend to be lower when estimated based on the variation between siblings, than when estimated across families. The coefficients on IQ, approximating ability, are also reduced when wage functions are estimated within families (i.e., with family fixed effects) rather than across families. Griliches interprets this empirical pattern to suggest that family allocation of schooling investments among offspring and twins is reducing inequality within families, consistent with compensatory allocation of human capital by parents across their children according to their initial endowments.

Becker and Tomes (1976) suggest that wage returns to schooling are greater for more able children. If parents then maximize the earnings of the sum of all dynasty members, they would concentrate their human capital investments on their most able children. If it were costless to then reallocate the dynasty's total earnings among members, parents could subsequently accomplish the distribution of consumption among family members they wanted, without sacrificing their initial objective of efficiently allocating their human capital investment. Given Becker-Tomes assumptions regarding parent reinforcing investments in the better endowed children, the OLS estimated wage function (Eq. 4) would tend to overstate the returns to human capital, S and H , unless the researcher could also control for the child's initial endowments. Estimating returns within families, by including family fixed effects, or by analyzing differences between twins. These within family return estimates would be purged of bias due to other confounding factors that do not vary within families or twins, respectively. The direction of the bias due to parent child investment response will thus depends on whether parents compensate or reinforce the exogenous initial endowments they observe in their offspring. In this regard, a natural initial endowment of the child to study is birth weight.

Health endowments at birth, such as birth weight and gestation, may be influenced by parents, however, by their prenatal behavior, such as maternal nutrition, the timing and form of prenatal care, smoking, and the number, timing and spacing of births (Mwabu, 2009; Rosenzweig & Schultz, 1983), all of which could be affected by parent heterogeneous preferences toward child health and might persistently influence later family resource allocations among children, making these measures of endowments at birth endogenous to the family's reproduction, life-cycle behavior, and subsequent child investments. Behrman, Rosenzweig, and Taubman (1994) find evidence that family allocation of schooling reinforces the preschool endowments of children, and add to the final inequality in their earnings, through their analysis of fraternal and

identical twins from the Minnesota Twin Registry. One way to identify exogenous variation in endowments birth, approximated by fetal growth rate, is to examine differences in birth weight between identical (homozygotic) twins that cannot be affected by parents (Behrman & Rosenzweig, 2004; cf. Figure 4). Identical female twins from the Minnesota Twin Registry ($n = 804$) are analyzed, and the effect of birth weight per week of gestation (i.e., fetal growth rate) is estimated on the child's later outcomes, such as adult height, completed years of schooling, and adult wages. This within twin estimate holds constant any effect of different genetics on ability (a), sex, gestation, maternal health, and family fixed effects from the time of birth. These within twin estimates of the effect of fetal growth rate on later child schooling are 70% larger than OLS estimates across twins, and within twin effects of fetal growth on wages are nearly five times as large as the OLS estimates, whereas the estimated effect of fetal growth rate on adult height remains essentially unchanged from the OLS estimate, and that on adult BMI ceases to be significant within twins. The increase in the effect of birth weight on schooling and wages within identical twins suggests parents are reinforcing the exogenous variation in fetal growth rate at the time of birth by their subsequent resource allocations that enhance schooling and adult productivity. The stability of the fetal growth effect on height within and across twins suggests that parents do not affect adult height through their child investment response to their child's initial endowment. Alternatively, adult height may be largely determined by genetics and conditions *in utero* and is not strongly affected by parent child investments between twins, possibly due to postnatal nutrition and exposure to early childhood diseases. Behrman and Rosenzweig adjust their estimates of birth endowments on schooling and productivity for the diminishing returns to fetal growth rate (nonlinear health-productivity relationship), incorporating the fact that the distribution of fetal growth rates for twins are lower than for singleton births. Consequently, the impact of raising the average birth weight for all births is likely to have a smaller effect than they estimated within the sample of twins who are on average low birth weight.

The objective of Rosenzweig and Zhang (2009) is to estimate from a sample of Chinese twins, the consequences of changing exogenously family size on child quality. To recreate the effect of increased family size for parents who have predominantly singleton births (99% of the Chinese births), they control for the fact that twins have on average a lower birth endowment (i.e., birth weight) than do singleton births, and the (closer) spacing of twins may also affect the relative cost to parents of investments in the human capital of their children, because the time parents have for child investments is limited, and parent credit constraints could raise the cost of borrowing to invest in both twins at the same time. They show that the birth order of the twins allows them to identify within bounds the tradeoff between increasing family size due to the exogenous occurrence of twins increasing family size, and the effect of twins reducing child quality compared with other family births. Their

analysis is based on a sample of twins and children of similar ages drawn from the 2000 Census of China from the Kunming district of Yunnan Province ($n = 1169$). They formulate a model for the biological and behavioral components involving parent response to twins that could bias direct estimates of the effect of child quantity on child quality associated with the occurrence of twins, assuming as Becker and Lewis (1974) do that child quantity and quality are substitutes for parents.⁷ Child quality might be represented by either schooling or health. Rosenzweig and Zhang (2009) consult several indicators of schooling—expected college enrollment, years completed, and standardized math and language exams—and health—subjective good health, weight, height, and BMI.

Evaluating the magnitude of this cross-effect of family size due to twinning on child quality is potentially important for population policies regarding fertility and health, because if family planning or reproductive health and education programs encourage woman to reduce voluntarily their fertility, the consequence of these program-induced declines in fertility would raise child quality and this could be viewed by a policy maker as a positive social externality of the program, potentially justifying public subsidies and the burden of taxes needed to finance publically such subsidies (Schultz, 2008a). First, twins are biologically lighter at birth, have shorter gestations, and lower APGAR scores. There is also evidence that twins experience lower survival rates, fewer years of school attainment, and at least for males, lower cognitive achievements and wages in later life (Almond, Chay, & Lee, 2005; Black et al., 2005; Black, Devereux, & Salvanes, 2007; Cesur & Rashad, 2008; Royer, 2009). But these subsequent measures of child “quality” could be due to both the lower endowments of twins at birth, and the effects of differential investments by parents after birth, which could reinforce (or compensate) for twin lower birth endowments, and thereby strengthen the negative quantity-quality negative effects of twins on child quality (or if compensating, reduce their magnitude).

Assume that parents reinforce the birth endowment effects between twins, as Behrman and Rosenzweig (2004) and Behrman et al. (1994) found in the US, and Rosenzweig and Zhang (2009) find in China. The effect of twinning on later child quality outcomes allows for twins to have lower endowments at birth than singleton births. If the twinning occurs at the second pregnancy the estimated effect of the twin on second birth children’s quality provides an upper bound (more negative) estimate of the average negative effect of increasing family size. If the twinning effect occurs on the first pregnancy, the effects on later child quality provides a lower bound (less negative) estimate of the effect of family size on child quality. When the child quality equation is also conditioned on birth weight, as an approximation for initial birth endowment that parents observe, the range of these upper and lower bound estimates of the quantity-quality tradeoff should be narrowed, with reinforcing parent behavior according to their model (Rosenzweig & Zhang, 2009).⁸

A panel study of births from the US and UK of child-specific postnatal human capital investments by parents suggests that observed child inputs, for example, breastfeeding, respond inversely to the child's initial endowment measured by birth weight, holding constant for fixed mother effects (i.e., within families), but recognizing the endogeneity of birth weight and identifying its independent variation by the effects on birth weight of the mother's prenatal smoking and working in the labor force until less than 2 months before the birth (Del Bono, Ermisch, & Francesconi, 2008). This finding is consistent with Griliches' (1979) hypothesis that parent exhibit compensating behavior that reduces sibling inequality. But the smoking and labor supply behavior are also endogenous, and the birth weight outcome of earlier births might affect subsequent maternal behavior and thereby impact the difference in prenatal behavior between births. Further empirical study of the direction and magnitude of parent allocation of child investments in response to initial child endowment is a priority area for further research on the family and population policy, especially in low-income and high child mortality settings (cf. Data, Gosh, & Sood, 2007). Decisive studies will probably require panel data where the endogeneity of prenatal maternal behavior on initial child endowments are reasonably identified, and important postnatal investments in children, such as breastfeeding and the timing of immunizations, are observed by gender.

Complementarities between health and education inputs are also potentially an important feature of the process of child development (Glewwe, 2005; Glewwe & Jacoby, 1995; Glewwe & Miguel, 2008). There may be different private as well as social externalities in the timing of these child investments that could be significant for setting priorities for social welfare and population policies. Diminishing returns to child investments in each time period seem plausible, and intertemporal specificity of the formation of different socioemotional skills and abilities suggests that social returns to compensatory interventions for some low endowments groups may be more effective at an early age (Cunha & Heckman, 2007). Economic returns to child investments may also differ significantly across socioeconomic strata, advantaged and disadvantaged classes, ethnic groups, geographic populations, and thus these programs could be strengthened and their overall effectiveness increased by appropriate targeting of the program to specific groups.

Disabilities among children are another endowment widely observed by parents at an early age that appear to occur at all economic levels in society, but are less often studied as a factor in parent investment behavior. In low-income countries, children identified with disabilities in household surveys are less likely to enroll in school, and complete fewer years of schooling when they do enroll. In 8 out of 12 recent DHS conducted in low-income countries, adults with disabilities are significantly more likely to reside in poorer households, with imputed poverty in the lowest two fifths of the population. But controlling for the lower educational attainment of the disabled member of the household, only 3 out of 12 surveys continue to find a

significant partial association between disability and this measure of poverty (Filmer, 2008). This study suggests that disabilities among children increase their poverty as adults, primarily through reducing their schooling. The economic burdens associated with disabilities might be reduced if opportunities to improve schooling opportunities for the disabled were pursued. There are limitations in available data from developing countries on disability, verbal and nonverbal ability, noncognitive and social skills, and geographic lifetime migration and how these characteristics of children are themselves influenced by parents, families, schools, and other social institutions and policy. Research on these topics could have implications for population policies, but it will require both improved data and better analyses of the economic returns to improving the health, schooling, cognitive ability, noncognitive skills of youth and their determinants.

4.4 Spillovers from health human capital beyond the individual and family

The health human capital stock of workers could also raise the returns on capital, by reducing absenteeism, which could reduce the need for firms to have a redundant work force, and could even reduce household size if hiring casual replacement labor for sick family workers is less productive than relying on workers with more specialized experience in the respective firms and families. These externalities of health in the functioning of the economy are hypothesized to be important in those countries in which the HIV/AIDS epidemic is severe, but estimates of their magnitude in the aggregate economy are elusive and even estimates of the microeconomic consequences of health on worker productivity are rare and generally treat health as an exogenous shock and not as an accumulated family and individual investment (cf. Thirumurthy, Zivin, & Goldstein, 2008).

As already noted, there are few forcing variables that theoretically provide identification of the unanticipated effect of income on health, or the effect of exogenous shocks to health on labor productivity and income. This line of research in health economics does not have a comparable natural experiment as with twins affecting exogenously fertility and thereby family size. It is attractive, therefore, to find settings where social experiments can be implemented in which access to specific promising health treatments or subsidies for health-related inputs can be randomly offered to separated populations who are then followed over time. Survey data on these treatment and control populations should be considered as a basis for assessing the health consequences of welfare expenditures, and the short-run and long-run productive consequences of these public programs on individuals, families, and communities. Social experiments assessing health interventions are ethically complex. But neglecting the task of evaluating health program effectiveness is also difficult to defend, particularly when returns to health appear to be promising, but the uncertainty of these returns

leads policy makers to discount the contribution of health to economic growth and treat health as merely a valuable consumer good (World Bank, 2008). Controlled field experiments have proven useful for testing medical procedures and drugs in order to maximize their benefits to society, while protecting the privacy of individuals and providing participants with an understandable basis on which they can decide whether they want to participate in the social experiment and grant their informed consent (Duflo, Glennerster, & Kremer, 2008).

Health status reflects the accumulation over a lifetime of three types of factors. First, there is a genetic health endowment, which is assumed fixed after conception and is generally not affected by policy. Taking genetic variation into account may increase the effectiveness of policy interventions and influence the personal distribution of benefits from a given intervention. Second, after conception the individual's stock of health is modified by investment choices. The health care system as well as the family and individual makes choices that affect exposure to environmental risks, modify health-related behavior, and provide inputs relevant for the production of health, and possibly influence the formation of preferences, such as individual discount rates. These choices are called here investments in health human capital or health inputs, although they may satisfy other possibly conflicting consumption goals, such as smoking, drinking, and risky behavior. The third health component includes unpredictable stochastic factors associated with incidence of disease, climate variability, economic conditions, and epigenetic biological developments that are outside of the control of the individual or family or community institutions. Whatever indicators of health are studied, it is likely that many inputs to health will be omitted from an empirical analysis of health, and omitted variable bias is a serious limitation in estimating health production functions, if the omitted variables are correlated with those health inputs that are being empirically assessed (Mwabu, 2009; Rosenzweig & Schultz, 1983).

Measures of health human capital are, moreover, likely to be heterogeneous, by which I mean some part of variation in health indicators is innate or produced by genetics at the time of conception and affected by subsequent unpredictable shocks. Some part of health is produced systematically by the private and social investment choices, and I call this second component *reproducible health human capital*. The uncontrolled and reproducible components of health may be indistinguishable in a survey, because surveys tend to rely on indirect proxies for latent health status and conditions. But variation in these proxies for health may have different effects on worker productivity and well-being, depending on the source of variation in the proxy. Therefore, to evaluate the effects of health status for policy purposes, the productive effects of changing the *reproducible* component of health is relevant to most social policy choices. For evaluations of health policies, we are primarily interested in the production of health and specifically the productive returns to the reproducible health human capital component of measured health that can be effectively increased by policies, and possibly targeted to disadvantaged individuals.

To estimate the impact of reproducible health human capital on worker productivity, the effect of such heterogeneous indicators of health is estimated as though they are measured with error and are potentially endogenous. Consequently, instruments are needed for the observed stock of health that identify the health effects of health changes due to environmental sources of variation as would be affected by household investments in nutrition and health care and community health investments.

4.5 Phases of the mortality transition and implications for health inequality

The decline in mortality in high-income countries until about 1950, and in low-income countries from about 1940 to 2000, are attributable mainly to the first phase of control of infectious, communicable diseases. The related medical and public health technologies which accomplished this mortality revolution were relatively cheap to disseminate and were available to most basic community health care systems in the low-income world. Reduced malnutrition, combined with these medical interventions, achieved a sharp decline in infant and child mortality. The second phase of medical interventions is designed to deal with degenerative, noncommunicable, and chronic diseases which claim the lives mainly of prime aged adults and the elderly. It remains unclear whether effective low cost therapies for chronic diseases of the heart, lungs, cancer, arthritis, diabetes, etc., can be implemented in low-income countries or even provided to the poor in high-income countries without increases in public and private expenditures on health, or how effective they will be, if the fetal and childhood health conditions were disadvantaged. Drug therapies often involve large research and development costs, but low marginal costs due to large scale manufacturing. Some of these drugs have reached low-income countries because, as they lost patent protection, their prices as generic drugs have declined substantially. Diseases which are not prevalent in high-income countries do not receive much attention from researchers, because there is no high-income community of potential users willing to pay for the initial research and development costs. The diffusion of medical technology to reduce mortality from degenerative diseases among adults and the elderly may not equalize life span or health status across countries in the future, as occurred dramatically in the first phase of the mortality transition when infectious diseases were controlled in childhood (Becker, Philipson, & Soares, 2005). Declines in adult mortality since 1955, increasingly due to control of chronic and degenerative diseases, such as heart, lung, cancers, or diabetes, have not noticeably reduced health inequality as measured by life expectancy at age 10, at least not among 21 high-income countries, or within these countries as can be measured between distinguishable race and ethnic groups (White, 2002).

Another approach to measure health inequality does not seek to distinguish differences between socioeconomic status groups, but relies on the unconditional frequency distribution of cohort survival, as reported in a national life table. The variance in

expected lifetimes for the synthetic cohort is calculated in addition to the mean or median (Edwards & Tuljapurkar, 2005). The variance in life span for the cohort of survivors to age 10 has not shown a tendency to decline within high-income countries since 1960, though it does when the variances are calculated from birth. With less reliable adult age-specific registrations of death by cause in low-income countries, there are few studies of the causes of adult mortality declines in the low-income world, or who in the population has benefited most from the actual declines. And when health objectives are medically defined in terms of the incidence of specific categories disease, such as Malaria, TB, or HIV/AIDS, it is not surprising that knowledge about the determinants of death by cause is uncertain. Many deaths are due to multiple causes, making it even more difficult to infer how deaths from specific causes respond to policy interventions.⁹ Presumably many developing countries are reaching the stage today when further mortality reductions will depend on preventing adults from dying of chronic and degenerative illnesses and a handful of drug-resistant infectious diseases.¹⁰

5. MACROECONOMICS EVIDENCE ON HEALTH DETERMINANTS AND ECONOMIC GROWTH

The primary causes of death before the mid-twentieth century in more developed countries, and in most less developed countries until the end of the twentieth century, were infectious and parasitic diseases. Today deaths due to these diseases are geographically concentrated in a few tropical regions of the world with poor health infrastructure, extreme poverty, and other conditions unfavorable to health. These geographic patterns of health and the prevalence of particular diseases, such as malaria, TB, and HIV/AIDS, have been used to account for economic development in a variety of ways. Poor health, short life span, and specific endemic diseases have been treated by economists as if they were exogenous determinants of the income levels and growth of income. Based on cross-country regressions, growth in income since 1960 is greater in countries for which life expectancy is greater. These partial correlations, however, may or may not remain significant statistically after controlling for other economic inputs to growth, such as the *per capita* supply of labor, human capital, and physical capital (e.g., Barro, 1997; Barro & Sala-I-Martin, 1995; Bloom, Canning, & Sevilla, 2004; Bloom & Sachs, 1998; Gallup & Sachs, 2001; WHO, 2001). But many of these economic inputs to growth are themselves endogenous to households and countries. If these intercountry comparisons are extended to interpret changes over time within countries (i.e., allowing for fixed country effects), the impact of invariant geographic and climatic indicators are, of course, then not identified. It is only reasonable to expect unobserved factors which are correlated with levels and changes in measured health conditions to also be related to income levels and income growth, for reasons not directly related to health, biasing upwards cross-country regression estimates of health

effects on income and economic growth (Acemoglu & Johnson, 2007; Weil, 2007). It is widely concluded, therefore, that the prevalence of disease and indicators of health status used in cross-country studies are themselves produced by a combination of environmental and institutional conditions, as well as social and private behavioral responses to the conditions in these countries. The social scientist must understand and explain these health determinants if they are to clarify how these general health outcomes causally contribute to the economic development of countries, as illustrated earlier in the stylized Figure 1. Several studies have used country-level data to identify what are interpreted as causal estimates linking health and development.

Acemoglu, Johnson, and Robinson (2001, 2003) propose a dynamic framework involving a two-stage explanation for the contemporary relationship between development and disease environments. In the first stage, the health-disease environment impacts how colonial governments settle areas which in turn facilitate or impede the development of social institutions that favor long-term economic growth. European colonialists were more likely to settle permanently in temperate climates where their native agricultural crops and production practices could be transferred with the least investment in local adaptation, and where local diseases were least lethal to the European immigrants, for example, North America, Argentina, Chile, and South Africa (Diamond, 1997). Agricultural settlers brought with them concepts of property rights and institutions to protect those rights and sought to replicate institutions of governance that corresponded roughly to those they had known in Europe. Conversely, where regimes confronted tropical agricultural climates and endemic diseases that were especially lethal for the European immigrants, the colonial regimes were more likely to resort to extractive industries, which exported precious metals, natural resources, or slaves, and depended frequently on restrictive governing arrangements to assure a reliable supply of labor in the export industries. These “extractive colonies” adopted institutions which did not promote property rights or evolve rules of governance which called forth investments that would increase total factor productivity and sustain modern economic growth. Acemoglu and his colleagues hypothesize that the distinctive institutions evolved by colonial regimes in response in part to their health conditions offer one explanation for the subsequent poor economic performance of colonial sub-Saharan Africa and some tropical areas of Latin America and Asia, if not the later collapse of Argentina. The contemporary efforts to control tropical diseases and improve health in areas such as Africa are not likely to achieve by themselves the economic benefits correlated today with better health outcomes (WHO, 2001), because the institutions, which are beneficial for sustained economic development, such as protection of private property and good governance, might be lacking for some time.

To quantify how health affects development, Acemoglu and Johnson (2007) assess how exogenous improvements in health directly affect growth in low-income countries since World War II. First, they approximate the time when new effective public health

technologies were introduced in the world against seven major infectious diseases or causes of death. Second, they consult League of Nations estimates of country-specific cause-of-death rates. The approximated discontinuities over time in these disease-specific health technologies are then interacted with the initial prevalence of each disease in each low-income country for which they have data. These disease-specific health technology changes weighted by country-specific initial prevalence of the disease become their instruments to account for life expectancy in the first stage of their analysis. The instruments account for a significant share of the increase in life expectancy (and in population growth) in their sample of countries from 1940 to 1980. But the increases in life expectancy predicted by these instruments do *not* account for the level or growth in *per capita* income in this period. They cast “doubt on claims that unfavorable health conditions are the root cause of the poverty.” As already noted, life expectancy is not a perfect indicator of the contemporaneous productive benefits of health, and although they introduce discrete lags for income growth of 5, 10, or 20 years, to allow the gains in life expectation due to reduced child mortality to affect income growth, this method may still not capture the dynamic life course biological process by which healthier surviving children become more productive adults and thereby enable their countries to grow more rapidly in the long run. African countries are also notably absent from their sample, because of the scarcity of reliable mortality data in these countries for this time period.

Lorentzen et al. (2008) consider growth regressions across countries from 1960 to 2000 and concludes adult (i.e., age 15–60) mortality is more significantly associated with growth than is infant mortality. They assume first that 12 environmentally fixed indicators of malaria ecology, climate, and geography are valid instruments for predicting infant and adult mortality rates at the country level, and second that these age-specific measures of mortality affect three inputs to growth, namely, the investment rates (physical capital), secondary school enrollment rates (proxy for human capital), and total fertility rates (proxy for cohort fertility). Their two- and three-stage structural estimates of growth depend on the exclusion restrictions they impose on the two intermediate levels of their model: (1) the 12 fixed environment instruments affect growth inputs and growth *only* through their correlations with the age-specific measures of mortality and (2) that mortality affects growth *only* through their correlations with the distinguished inputs to growth of investment, enrollment, and fertility. Both restrictions appear tenuous. For example, why should being a landlocked nation affect economic growth only by means of variation in age-specific mortality, and not by traditional economic routes, such as access to international trading opportunities. Why should fertility be specified as a negative input to growth, except conveniently because it happens in this period to be negatively correlated with income and growth (cf. Barro & Sala-i-Martin, 1995; Kuznets, 1967). Moreover, for some parents fertility is a choice, which appears to respond as might be expected in a simple economic model to the wages of men, women and children, wealth, production technology and technical

change, women's education, and access to birth control, among other factors (Schultz, 1997, 2008a; Section 8). It has also been hypothesized that exogenous variation in mortality could affect the demand of parents for births. Yet most empirical models of economic growth fit to cross-country data continue to assume health, mortality, and fertility are all determined outside of the model or exogenous, and this methodology conceals whatever role there may be for population policy to affect development.

Conditional on the assumptions underlying the model of Lorentzen et al. (2008), the estimated standardized association between adult mortality and growth is larger, as expected, than the standardized association between infant mortality and growth, if for no other reason than that children do not immediately contribute to national income, whereas adults often do. For example, a one standard deviation decrease in adult mortality is associated with an increase in annual economic growth of between 0.8% and 1.1% points, when the average growth in their sample is 1.8% per year. A standard deviation decline in infant mortality is associated with a smaller 0.27% point increase in growth, and their estimated association with infant mortality is not significantly different from zero (Table 10). Thus, the specification choice of Acemoglu and Johnson (2007) and most other studies of growth in national income to rely on life expectancy at birth as their measure of health assigns a larger weight to the variation in infant mortality in this phase of the demographic transition than is justified for understanding short-run variation in production. Lorentzen et al. (2008) also find their reduced-form association between adult mortality and growth is attributable in their recursive model to an association between lower adult mortality and increased capital investments and lower fertility. However, the association between adult mortality and secondary school enrollment rates, or enrollment and growth are both insignificant, which may signal a limitation of this specification, or a shortcoming of the Barro and Lee (2000) school attainment series used in many cross-country studies (Topel, 1999).¹¹ It is interesting, however, that changes in adult mortality and infant mortality are empirically distinguishable in this period, suggesting that future studies should at least decompose changes in life expectancy at birth into variation in adult and in child mortality as potentially distinct determinants of growth, even if the pathways by which mortality today affects growth inputs and outcomes in the future are not yet understood (cf. Weil, 2007).

5.1 Macroeconomic evidence of the impact of income on health

The second causal relationship to assess in the field of health and economic development is the effect of personal income and other features of economic development that may influence mortality, morbidity, and health status. To estimate the effect of income on health at the country level, Pritchett and Summers (1996) examine changes between 5-year intervals in income and health in low-income countries from 1965 to 1985. Estimating by ordinary least squares (OLS) the linear association between changes in

national income *per capita* and changes in the infant mortality rate, controlling for time-period effects and adult schooling levels, they find a 10% increase in income is associated with a 1.9% decrease in infant mortality. Because other factors could affect both income and mortality, they then propose several possible instrumental variables (IV) accounting for income changes that are assumed to affect health only through their impact on income, or in other words instruments correlated with income but uncorrelated with the errors in mortality or unexplained variation in health. Among their proposed instruments, the most plausible is a country's "terms of trade" (Z in Figure 1). Presumably an increase in the price of a country's exports relative to the price of its imports is an exogenous "shock" promoting domestic income growth, but because export prices are determined outside of the country in the global economy, this shock should not affect the health of the domestic population except through income. This IV estimate of income's effect on infant mortality is not significantly different from zero (i.e., $t = 1.28$). Among their other four proposed IV estimates, only the "investment to income ratio" yields a second-stage estimate which is statistically significantly different from zero, implying a 10% increase in income is associated with a 3.5% decline in the infant mortality rate over a 5-year period. However, the "investment ratio" is not a valid instrument, because it could be affected by many factors in the domestic economy such as weather, civil conflict, or development policies, all of which could arguably also affect health and thereby violate the assumption for IV methods. When Pritchett and Summers consider 5-year differences in life expectancy as their dependent health variable, the direct OLS association implies a 10% increase in income is associated with a smaller 0.15% increase in life expectancy, and this association is not significant. Moreover, none of the IV estimates for income's effect on life expectancy differ significantly from zero. Nonetheless, Pritchett and Summers conclude that "gains from rapid economic growth flow into health gains," whereas the actual evidence they report suggests the contrary that direct or IV linkages in the cross section of countries are not significant *within* low-income countries. When a credible instrument, such as terms of trade, is used to predict exogenous variation in national income, which is sufficiently powerful in the first-stage equation, there is no significant income effect on health according to IV methods, even when health is proxied by infant mortality, a relatively sensitive short run measure of health.

Pritchett and Summers (1996) do not consider a suitable instrument which might allow them to vary exogenously adult mortality without directly affecting national income, and thereby estimate the influence of health on the productivity of workers and assess empirically the strength of causation flowing in the opposite direction from population health to national income. These macroeconomic studies do not offer a convincing basis as yet for identifying causal relationships between health/mortality and development in either direction. The historical relationships Acemoglu et al. (2001, 2003) describe between disease conditions and colonial settlement strategies

and resulting institutional developments are plausible, but they are not relevant for assessing the contemporary consequences for income of health policies or measures designed to control specific diseases. Pritchett and Summers (1996) report no significant evidence of the impact of income on infant or life-span mortality. These cross-country studies do not succeed in disentangling satisfactorily the causal relationships that might account for the provocative association between health and development.

5.2 Alternative indicators of health status and their economic consequences

Contemporary epidemiological and historical studies have documented that bad early life health conditions experienced by a pregnant woman, her fetus, and young child are significantly associated with the child's reduced longevity after age 50 (e.g., Alters & Oris, 2006; Barker, 1994, 2001; Bengtsson & Brostrom, 2006; Bengtsson, 2009). The unstated reason for focusing on late life mortality is probably because too few deaths occur between age 15 and 50 to analyze in the moderate sized historical samples. A growing number of econometric studies are also finding that bad early life shocks to health are associated with decreased cognitive test scores, schooling attainment, occupational status and earnings, and with increased adult morbidity, chronic health problems, nonparticipation in the labor force, and disability even before the age of 50, and more notably thereafter (e.g., Almond, 2006; Bleakley, 2007; Costa, 1996; Glewwe, Jacoby, & King, 2000; Maccini & Yang, 2006; Maluccio et al., 2006; Miguel & Kremer, 2004). If these early bad health shocks increase morbidity and disability among working aged adults, one might then expect four to five decades after an improvement occurs in maternal and child health, adult health and labor productive potential per working aged population should increase, while the decline in late mortality would raise slightly the adult share of the population and add further to growth in income *per capita*. The challenge for research is to identify exogenous variations in fetal growth and early child health that can be convincingly linked to long-gestating adult health, worker productivity, and economic growth per adult. These lines of research are discussed further in Sections 6 and 7.¹²

Many studies analyze the birth weight of a child, which is highly positively correlated with child survival, and child's cognitive performance, school attainment, adult earnings, and late life survival and health status (Alderman & Behrman, 2006). Cross-sectional association between birth weight and subsequent outcomes, however, could be biased from causal relationships, because unobserved preferences and behaviors of parents, and their access to medical care and health-related inputs could not only influence the child birth weight, but also affect subsequent health inputs and behaviors, contributing to later welfare outcomes (Rosenzweig & Schultz, 1983). Evidence is more confidently interpreted when differences within twins are followed, for whom within twin differences in birth weight measure exogenous differences in rates of fetal growth, and can then be causally related to later parent health inputs and the child's development, holding constant for

such commonly omitted variables as gestation, mother's health, family fixed effects, and even genetic potential affecting ability or initial health endowments (Behrman & Rosenzweig, 2004; Black et al., 2007; Rosenzweig & Zhang, 2009).

This literature is reviewed later in the chapter and suggests that the productive benefits of the decline in child mortality, which is responsible for most of the increase in life expectancy in low-income countries in the twentieth century, might be expected to impact adult economic productivity after a lag of several decades. Moreover, declines in child mortality are frequently associated with declines in fertility within a decade or two. And voluntary reductions in fertility may facilitate economic growth, if they are associated with women increasing their participation in the market labor force as they reallocate time from child care and allied home production, and potentially increase life-cycle savings and investments for retirement, and lead to a substitution of more human capital per child for having fewer children (discussed in Section 8). If change in life expectancy at birth is a poorly designed indicator of change in the productive stock of health of the current working age population that might drive economic growth, how can the latent productive stock of health human capital be better measured at both the aggregate level and the individual level?

The present discounted value of the expected future productive lifetime of an individual is another approach to value the economic gains from reducing mortality for individuals at each age. This present value calculation would assign a larger productive value to savings a life of a 25-year-old adult compared with saving the life of an infant who would not work for more than a decade. Alternatively, the standard summary measure of life expectancy at birth implicitly weights the saving of an infant's life more heavily than that of an adult, because this synthetic cohort measure of mortality does not (1) accumulate the burden of investment costs of child rearing, (2) discount these costs and eventual adult earnings back to the time when an intervention might have avoided a premature death, and (3) allow an individual's future productivity to depend on the intervening accumulation of capital, technical change, and inputs that could complement labor. Life expectancy attaches the greater value to improvements in the survival of a cohort when the improvement occurs at the earlier age. If the valuation of life is based on a discounted value of the profile of consumption for a synthetic cohort, rather than on the profile of productive potential, the social weight assigned to savings the infant's life would increase relative to that of a working aged adult (Usher, 1973). On the other hand, national income statistics do not currently treat health or even schooling as an investment which can add to future productivity and growth. If national income statistics treated human capital as they do physical capital, then the present discounted value of lifetime total productivity (in market or nonmarket activities) could provide an analogous "production basis" for imputing a value to health improvements that would also account for future growth in national income. Valuing lives among the elderly may still require an extended human capital conceptual framework involving a social utility function, or the elderly

consumer's willingness to pay for extending their life. Public health services may in addition be viewed as producing a public good or a positive externality, if your welfare improves when your neighbor's health improves, even though the neighbor is too poor to pay for life-extending public health care (e.g., Murphy & Topel, 2006; Tolley, Kenkel, & Fabian, 1994; Usher, 1973). Observed wages may also be an increasingly unrepresentative indicator of the average potential productivity of the elderly cohort, because those who continue to work for a wage are influenced in their participation decision by their health status, public pensions, and by the tax treatment of earnings and savings (Gruber & Wise, 1995, 2004).

As emphasized earlier, the pattern of mortality change by age has evolved, possibly due to changes in health technologies, the economic benefits and health burdens of urbanization, and the diffusion of new diseases in the world. Consequently, a year increase in life expectancy at birth is associated at different times and in different countries with changes in different causes of mortality, affecting persons of different ages, with different implications for *per capita* productivity and desired fertility.

The rising costs of treatments for degenerative diseases is one factor underlying the growing share of national income expended on health-related goods and services in rich countries, and studies suggest the income elasticity of health expenditures exceeds unity, that is, consumers treat health expenditures as a luxury (see Table 1). Unless these health-related expenditures in high-income countries spill over and reduce substantially the opportunity cost of health therapies available to low-income countries, the convergence by income levels in health status across countries may not continue. However, in the case of anti-retroviral drug therapies for those living with HIV/AIDS the cost for low-income countries has decreased sharply in the last 5 years, suggesting it is possible for modern medical technological advances to spill over as a public good and benefit the world's poor. Nonetheless, to sustain the convergence in health across countries and across socioeconomic status groups within countries, new policies may be needed. Increased population assistance may be necessary, a social redistribution of the costs of effective preventive and even curative health care may be required, and intellectual property rights associated with new drugs and medical technologies may have to be evolved that benefit more widely poor countries or peoples (see Table 2).

6. MICROECONOMICS OF HEALTH AND DEVELOPMENT: INDIVIDUALS AND HOUSEHOLDS

There exist critical periods of early human development during which conditions can encourage or deter the healthy development of humans and influence not only their early survival and extended life span, but also affect in a complementary fashion their cognitive performance, schooling, and their productivity as adults, as measured by wages and earnings. Research on this topic is reported by epidemiologists, medical researchers,

demographers, anthropologists, economic historians, and finally most recently, by economists studying development in low-income countries. These disparate literatures are only beginning to be integrated, and the policy implications of the research are not always clear, though it is expected that more social attention will be directed in the future to care of pregnant women and to the health environment of children in the first few years of life in order to achieve long-term improvements in the health and economic potential of populations (Frankenberg, 1995; Frankenberg, Suriastini, & Thomas, 2005; Gluckman & Hanson, 2005; Koenig, Fauveau, & Woityniak, 1991). Only a few segments of this emergent field are summarized here to show how the measurement of health has evolved to rely on anthropometric indicators of health and physical development, how these indicators are used to link background conditions to economic behavior, the use of health inputs, and health outcomes, and motivate the use of two-stage estimation methods for quantifying key relationships and testing causal hypotheses.

McKeown (1976, 1979) concluded from an analysis of cause-specific death rates in the UK that the gradual increase in life span from 1750 to 1900 was due to rising standards of living, because the decline in death rates could not be explained by effective medical interventions or public health initiatives. Without finding any change in medical or public health technology or practices to explain adequately the actual reduction in deaths by cause, he hypothesized that the decline in mortality until the end of the nineteenth century was related to the improvement in standards of living and specifically nutrition, and perhaps also the reduced crowding of poor populations in urban housing.

Fogel (1986, 1994, 2004)) extends this line of thinking by drawing on the nutrition literature to show how the available supply of calories to a population is initially consumed to maintain basic metabolism and physical maintenance activities, as well as used to fight off infections and disease (Scrimshaw, Taylor, & Gordon, 1968). Only calories in excess to those required for sustaining life are then available to support productive “work” and leisure activities. This nutritional accounting scheme suggests that increases in the *per capita* supply of calories yields first a phase of increasing productive potential, followed by decreasing productive returns to calories as individuals approach a level of nutrition where the cost of additional calories exceeds the value of their marginal contribution to health and productive work. Increases in the consumption of calories may also be transformed into increases in weight, represented by BMI,¹³ which is associated with improved health and increased physical capacity for work, until BMI reaches a value of about 25, illustrated by the relative risk of mortality, based on Norwegian data from the 1970s (Waaler, 1984) and the Union Army Veteran sample from the 1880s (Fogel, 2004). Obesity occurs as BMI exceeds a threshold of about 30, above which adult risks of mortality increase and productivity is expected to fall. Fogel (2004) extends this form of calorie requirement accounting to his historical estimates of the personal distribution of calories available to the populations of England and France from 1700 to 2000. He thereby offers his explanation for France’s initially higher mortality and lower labor

productivity in terms of France's lower *per capita* supply of calories compared to England's in the eighteenth century. He thus accounts for much of the increase in European labor productivity since the industrial revolution in terms of increased adult height and improvements in the distribution of BMI (Komlos, 1994; Steckel, 1995, 2008).

In other words, increases in the current flow of nutrition in calories in excess of the requirements for work and illness should reduce the fraction of the population with very low BMI, and shift the lower tail of the distribution of BMI to the right. This shift is interpreted as an accumulation of the population's health human capital stock, which tends to be associated with both declines in mortality and increases in labor productivity. By approximating stocks of health human capital in these objective terms of the physical stature (i.e., height and weight) of populations, Fogel has oriented researchers to measure time series variations in these anthropometric indicators of well-being or health status within relatively closed populations. Researchers have also documented how adult chronic health conditions improve for those who have experienced less nutritional deprivation and less exposure to infectious disease causing inflammation or micronutrient deficiency, especially as a child (Crimmins & Finch, 2006a; Field, Robles, & Torero, 2007; Finch, 2007). The challenge is to quantify inflows of nutrition, various claims on these calories, such as diseases, that could together affect stature and imprint on the epigenetic process, and work through these hard to observe pathways from measurable indicators of health status to biological processes shaping physical and mental development, and economic productive capacity, as well as life expectancy.

Substitution and complementarity between nutrition, health inputs, preschool stimulation and training, as well as intertemporal tradeoffs, and diminishing returns to scale suggest child development should be modeled to include health, cognitive and non-cognitive capacities. Health status or the stock of health human capital can in addition be assessed in household representative surveys by the functional capacity of persons to perform activities of daily living (ADLs) which seem to discriminate health status among the elderly. Finally, the effect on productivity and labor supply of "days ill and unable to work" during a specified interval, such as the month before a survey, can also be treated as an endogenous indicator of health status and evaluated as a determinant of wages and labor supply using instruments related to the individual's birthplace and socioeconomic origins (Schultz & Tansel, 1997).

6.1 Height, weight, body mass index, and birth weight as indicators of health stocks

Anthropometric indicators of height and weight and birth weight are being increasingly used in comparative studies of the nutritional and health status of historical and contemporary populations. Height as an adult includes the long-run effect of fetal and childhood nutritional limitations and disease environment and is referred to as "stunting" when height is two standard deviations below the average in a reasonably well-fed reference population,

whereas weight for height (BMI) responds to the shorter-run nutritional balance among food, disease, and work and is referred to as “wasting” when BMI values are two standard deviations below average (Fogel, 1994, 2004; Waaler, 1984; WHO, 1995, 2006).

Adult height is thought to be largely determined by an individual’s very early development, and to be more or less fixed by the age of four, conditional on genetic height potential (Eveleth, 1986; Floud, Wachter, & Gregory, 1990; Martorell & Habicht, 1986). Elo and Preston (1992) conclude their review of the mortality literature succinctly: “Height is probably the single best indicator of nutritional conditions and disease environment of childhood. Like date and place of birth, it is a summary measure of many health-related circumstances and events, but it has the advantage of reflecting the experiences of an individual child.” Because adult height is more or less stable over the life course from age 25 to 55, the age “profile” of heights estimated from a cross-sectional survey or census of a population with negligible migration should portray time trends in childhood nutrition and exposure to infections (i.e., scarring). But these historic trends of early biological development is revealed among survivors, and may be altered by swings over time in past mortality that cull from the survivors of a cohort (i.e., selection) less healthy individuals at more vulnerable ages, who might be shorter, on average, as in the Chinese famine of 1959–1961 (Almond, Edlund, Li, & Zhang, 2007; Gorgens, Meng, & Vaithianathan, 2007; Schultz, 2004a). Fertility is also to some degree a choice variable, and some groups in society reduce their birth rate when economic and health conditions are uncertain or threatening, introducing the possibility that the socioeconomic composition of birth cohorts could change in anticipation of hard times that could modify the apparent effect of scarring and selection, and could affect as well the composition of births by season (Buckles & Hungerman, 2008; Loshin & Radyakin, 2009; Thomas, 2008). Composition, scarring, and selection could therefore all potentially leave their mark on the stature and health of a series of birth cohorts observed in surveys or censuses, though the scarring effect is thought to be the dominant factor in the last century. It is ultimately important to consult panel data for a large population in which individual stature is measured and subsequent mortality is recorded reliably. Waaler (1984) documented that the risk of mortality relative to others in an age and sex group declines for taller Norwegians over the age of 50 in the 1970s, and the regularity of declining relative mortality risk for those of greater height is approximately linear (Figure 4), and is especially salient for death due to cardiovascular disease, tuberculosis, and chronic obstructive lung disease.

In low-income countries the relationships between height and BMI, as indicators of health human capital, and health outcomes, such as mortality, morbidity, and labor productivity, are difficult to document among adults because there are few large panel surveys that record stature, health and economic outcomes and health-related inputs, and register death relatively completely by cause. However, height can be measured both for aggregate populations and for individuals from a single survey, although it

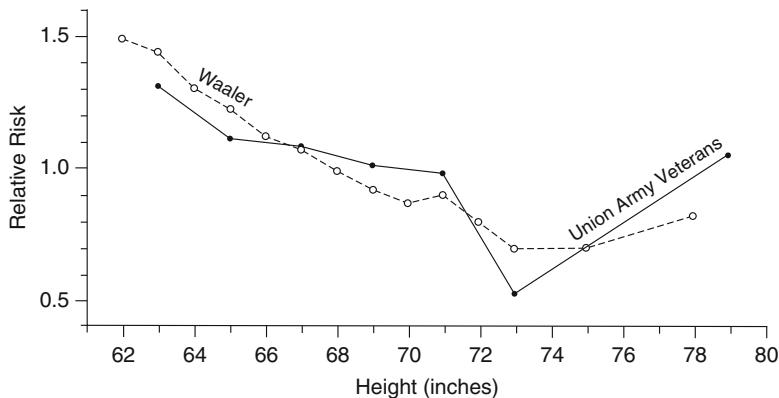


Figure 4 Relative mortality risk among Union Army Veterans and among Modern Norwegian males. Note: a relative risk of 1.0 means that the risk at that height was equal to the average risk of death in the entire population of males of the specified ages. Also note that data point for the tallest category, in both Norwegian and Union Army cases, is not statistically significantly different from 0.1.

should be emphasized that the information “signal” conveyed about prior nutrition, disease, and latent health by adult height is relatively small, unless averaged over a large population, because of the relative importance of genetic variability in this characteristic. However, height appears to be more objective than many self-assessed health indicators, and therefore may be less biased by subjective considerations or cultural conditioning, and may discriminate between the latent health status of younger adults, before chronic diseases and disabilities emerge in late middle age, after which ADLs may be a reliable survey indicator of health status.¹⁴ Panel data on the evolution of health outcomes for individuals over a lifetime and responses to potentially exogenous health, economic, and weather shocks promise to reveal how these shocks affect health stocks and welfare, holding individual initial health endowments constant. But even correlations between changes over time in health and productivity may still not capture causal effects, if the earlier outcomes are themselves endogenous or responding to prior unobserved constraints and preferences of the parents and children, and to their family’s available health care.¹⁵

Where death registration is incomplete and unreliable, data on adult height can still be collected in a representative survey to provide a historical window on the evolution of health human capital in the population, and even time trends may be estimated within closed ethnic, racial, or caste groups, who do not change their assignment by group over their lifetime (Schultz, 2003). Figures 5 and 6 report the pattern of average adult female height for rural/urban/total populations by 5-year moving averages of year of birth in Ghana (1987–1989), a country experiencing little economic growth

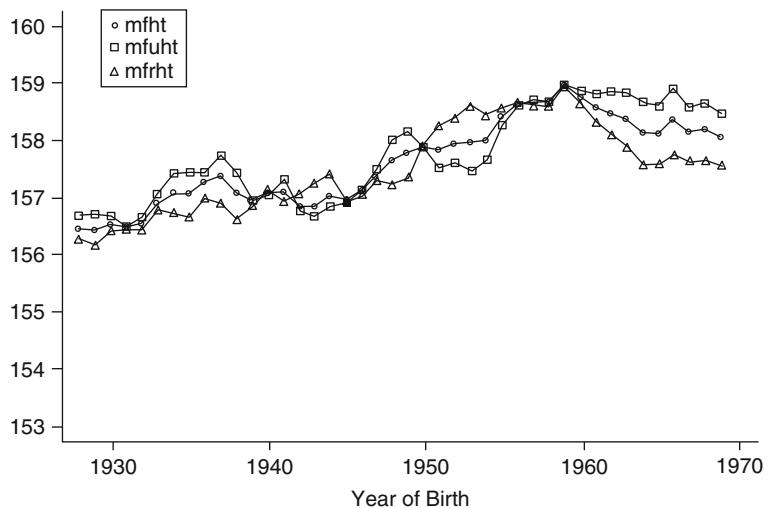


Figure 5 Height in centimeters of adult females in Ghana in 1987-1989 by year of birth, total current residential population (circle: ○), rural (triangle: Δ), and urban (square: □) regions.

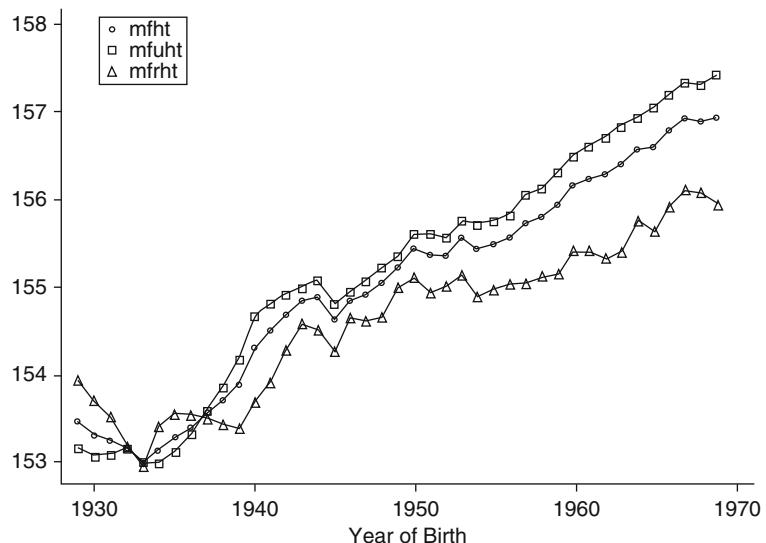


Figure 6 Height in centimeters of adult females in Brazil in 1989 by year of birth, total current residential population (circle: ○), rural (triangle: Δ), and urban (square: □) regions.

in the three decades after its independence in 1957, and Brazil (1989), a country growing relatively rapidly from 1935 to 1980. *Per capita* national income increased in this period more rapidly in Brazil than in Ghana, and this is one possible explanation for the more marked increase in height among younger women in Brazil compared with

those in Ghana, even though both countries should have had access to similar international improvements in medical technology (Schultz, 2005).

Attributing the secular increase in adult height to specific causes in high-income countries remains controversial, and estimating the longer-run cumulative and even short-run responses of height to increases in income or decreases in infectious diseases are debated, especially in less developed countries. Yet the magnitude of differences across birth cohorts cannot be readily attributed to genetic changes in the high-income populations in such a short period of time. The share of the variance in individual height attributed to environmental risks and socioeconomic conditions (in contrast to genetic factors) is thought to be on the order of 20% in high-income countries today, such as the United States (Stunkard, Foch, & Hrubec, 1986) and Finland (Silventoinen, Kaprio, Lahelma, & Koskenvuo, 2000) based on twin studies of trait heritability. But the environmental share of height variation is believed to be larger in low-income settings where malnutrition is much more common and disease imposes a heavy burden on available proteins and calories needed for satisfactory physical development (Silventoinen, 2003). Gene-environment interactions can also be important, and if these factors are correlated and contribute nonlinearly to height and socioeconomic outcomes, the task of decomposing the effects of genes and environment on adult height is still more complex. Genetic studies of five ethnically and geographically dispersed populations suggest that 39–56% of the variance in individual adult height is correlated with a few major gene effects (Ginsburg, Livshits, Yakovenko, & Kobylansky, 1998). Socio-economic characteristics of parents and birthplace also account for a significant amount of the variation in individual height, but for a substantially smaller fraction of the variance than is believed to be determined by a few genomic regions of chromosomes that have been scanned (Liu et al., 2004). Length at birth and height as a child and adult are, of course, persistently highly correlated over the life cycle of an individual, although the timing of the adolescent growth spurt occurs earlier as does puberty in the better nourished population. Both child and adult height are significantly associated with cognitive ability, schooling attainment and adult wages or productivity, in both high- and low-income countries (Schultz, 2003; Strauss & Thomas, 1998). The causal interpretation of these associations between height, and health, ability, and productivity remains open to debate, though some view height as simply one indicator of cognitive ability, and when these IQ measures are appropriately controlled, height loses much of its power to predict productivity in a recursive framework (Case & Paxson, 2008).

Data on adult height has only recently been collected from representative household surveys in low-income countries, such as for women of childbearing age in the DHS (downloadable at <http://www.measuredhs.com>). Because these surveys do not provide much information on economic or disease conditions at birth or even at the time of the survey, analysis of these height data examine time trends in averages for birth cohorts. Deaton (2007) has matched the height of surviving female birth cohorts age 25–55 to national

income ($\log \text{GDP per capita}$ in purchasing power parity from Penn World Tables) at the time of the women's birth, and the infant mortality rates (interpolated to the birth year from World Bank estimates every 10 years). In a pooled sample of low-income country survey cohorts from the DHS, he reports few stable relationships between height and income and infant mortality, interpreted as measures of nutrition and the burden of disease at the time of birth, respectively. However, there are distinct regional differences in height. On average, the 28 African surveys and that from Haiti (whose population is 95% of African descent) report women are on average taller, despite incomes being lower and infant mortality being higher than in the rest of the sample. Conversely, female heights are distinctly shorter in the South Asian countries (i.e., India, Bangladesh, and Nepal). These robust regional regularities in height suggest a genetic component may differ among these regions that accounts for some of the height differences, and Deaton concludes it is not possible to infer disease or development effects on height outside of the OECD developed countries, because of the likely bias introduced by omitted variables, such as genetic differences, which might be correlated as well as with income and infant mortality. Deaton and coauthors also propose that in low-income countries such as Africa (but not in South Asia?) the force of mortality selecting out shorter individuals dominates the force of scarring which reduces individuals achieving their genetic potential height (Bozzoli & Deaton, 2008). Anthropologists have long noted regional variation in height, but they tend to hypothesize that within a few generations nutrition improvements and decreases in disease lead to convergence in height, as documented in Japan and East Asia, and noted among recent immigrants to the UK from South Asia (e.g., Eveleth & Tanner, 1990; Floud et al., 1990; Komlos, 1994; Tanner, 1981; Tarozzi, 2009; Van Wieringen, 1986).

Akachi and Canning (2006) focus on the DHS female height data within sub-Saharan African countries born from 1965 to 1985. Average height by birth cohort for 19 African countries provides 438 observations and the simple linear time trends by birth date in height across women age 20–49 are mixed: positive in two countries, negative in two, and statistically insignificant in the remaining 15 countries. Then they add linear controls for GDP per capita and infant mortality, following Deaton (2007), as well as protein and calorie availability, as estimated by FAO.¹⁶ All of their explanatory variables are initially included not only in the year of birth, but also lagged 5, 10, and 15 years. The health and nutrition literature suggests that when a cohort is growing most rapidly, in the first 3 years and during the adolescent spurt, physical development is most vulnerable to nutritional stress and disease. The final regression specification they prefer is based on retaining only those variables that are statistically significant when country and 5 year of birth time dummies are also included. Infant mortality is significant only in the birth year, which is negatively related to the cohort's height; income is significant only when lagged 15 years and is positive; protein availability is positively associated with height in the year of birth and 15 years later; calorie availability is not significant jointly or independently at any lag. In the final regression, the

coefficients estimated from the birth cohort time dummies decline, especially for cohorts born from 1968 to 1977 (Figure 3). To account for this residual negative time trend in height, the authors speculate that infant mortality may not capture entirely the deteriorating effect of the disease environment on height, and possible increases in inequality in nutrition and disease might account for the declining trend in height and be important omitted variables in these aggregate regressions based only on population means. However, their reliance on “stepwise regression” to specify their preferred model is a shortcoming, though it does confirm Deaton’s observation that younger women in Africa do not appear to be taller than older ones. Nonetheless, other data sources than the DHS suggest positive time trends in height for male and female birth cohorts, as in the LSMS surveys (<http://www.worldbank.org/LSMS/>) from Cote d’Ivoire, Ghana, Viet Nam, a monitoring survey of Kenya, as well as health and nutrition surveys from Brazil, Colombia, and China (Savedoff & Schultz, 2001; Schultz, 2003, 2005). Deaton (2008) subsequently reports modest gains in height across male and female birth cohorts in India, based on the 2005/2006 NFHS (DHS), but the time trends in height are smaller in rural than in urban areas, especially for women. Most of these surveys, unfortunately, do not include health information on birthplace, or sufficient information on migration histories on which to impute health and nutritional conditions at the time and place of birth.

6.2 Heterogeneity of height as a measure of the stock of health human capital

Height as an indicator of an individual’s stock of health human capital is a proxy for latent health and can be thought of as having at least two components: one which is a form of “reproducible human capital” created by families, individuals and society by nutrition, control of disease, and early health care, and a second component which is not readily explained by observed socioeconomic variables, or greatly affected by measures of environmental risk. Only about a tenth of the variation in individual adult height across a population tends to be accounted for by parent socioeconomic characteristics (e.g., education, wealth, and socioeconomic status) and location of birth and childhood, which may be related to exposure to diseases and availability of health care. The remainder of the variation in height might be attributed to characteristics which may be genetic in origin or conditions affecting unpredictably biological development, and of course errors in survey measurement (Ginsburg et al., 1998; Gluckman & Hanson, 2006).

Schooling is another form of reproducible human capital created by investments of student time and educational resources. A fifth to a half of the variance in years of schooling attained across adults can be typically explained by family and student exogenous characteristics along with the family’s local access to schooling, and a smaller fraction of the variation in schooling is generally linked to intrinsic ability of the student, measured by such variables as IQ. The omission of ability as a determinant

of the wage rate in estimating a wage function is expected to bias upward the estimate of the private wage return to schooling, as discussed in Section 4.2, although this has not been widely confirmed perhaps because of errors in measurement of schooling. However, other studies find upward bias in OLS estimated schooling returns, in the sense that if they also condition wages on school quality, treating it as exogenous, the estimated coefficients on years of schooling is likely to fall because quality and quantity are commonly positively correlated (Behrman & Birdsall, 1983).

BMI could affect wage productivity as a physical attribute of workers or as a latent indicator of health, but the association could also capture reverse causation, because more productive workers have more income to spend on food and health care that would add to worker BMI in the short run. To deal with the potential reverse causality, BMI has been treated as endogenous in the wage function and local prices of food have been used as an instrument for BMI (Strauss, 1986; Strauss & Thomas, 1998). Generally, the IV estimates of the effect of BMI on wages are significant and positive, though subject to diminishing returns and even counterproductive if the individual is overweight.

Adult height, because it is thought to be primarily determined by an early age, is not expected to be subject to reverse causation with wages. But height could be heterogeneous, in the sense that the majority of variation in height potential is genetic and is not explained by socioeconomic conditions at an early age. Therefore the socio-economic and genetic variation in height need not capture equally the latent health human capital stock. The two components of height could enhance wage productivity by different amounts. Studies have used instruments from family background at time of birth, food availability, local health services, exposure to diseases, or weather shocks to predict adult height in estimating the determinants of the wage rate of individuals. They find an added centimeter in height is associated according to OLS with a significantly larger proportionate effect on wages in low-income countries than in high-income countries, such as the United States, presumably because of wage returns to height diminish at higher levels of nutrition, health care, and income (Schultz, 2002; Strauss & Thomas, 1998). However, IV estimates of the log wage returns to height tend to be several times larger than the direct (OLS) estimates of height on log wages, holding constant for schooling, postschooling potential experience, and sex (Maccini & Yang, 2006; Savedoff & Schultz, 2001; Schultz, 2002, 2003, 2005).

Because the IV estimates of the effect of adult height on adult productivity tend to be significantly larger than the OLS estimates, the Hausman specification test for the exogeneity of height in the wage function is generally rejected, implying that the IV estimates are preferred if the instruments are valid (Schultz, 2002). There are two possible explanations for the difference between the OLS and IV estimates of the wage returns to height: (1) the survey measurement error in height in household sample surveys is relatively large, leading to a downward bias when height is treated as measured without error in OLS estimates and (2) measured height is heterogeneous

and the health human capital component of height identified by variation in instruments such as parent socioeconomic characteristics and local health conditions at birthplace contributes to health human capital and increases health and wage productivity more than does the residual (presumably genetic) unexplained variation in height.¹⁷

First, measurement error in anthropometric indicators from surveys may not be negligible. Indeed, multiyear averages of measures of adult height from a panel survey in Cote d'Ivoire yield significant estimates of the proportion of measurement error in height measured in any single round of a survey, and this proportionate error appears to be larger for height than the proportionate measurement error in responses on years of schooling or year of birth. But measurement error in height is still not of a sufficient magnitude to explain a several fold increase in IV compared to OLS estimates for height's effect on log wages (Schultz, 2003). The second hypothesis as the more plausible: height is a latent indicator for health human capital and includes the modest effect of early biological development processes that also contributes to the individual's persisting health endowment, adult productive capacities, and longevity. The genetically induced variation in height, on the other hand, that accounts for the bulk of the observed variation in height across individuals is less closely related to productivity than is the reproducible component of height predicted on the basis of socioeconomic characteristics of birth environment and family endowments.

Investments in health human capital are thought to occur throughout life, but as with determinants of height, they appear to be critical in biological developments during specific periods in fetal development *in utero* and during very early childhood (Crimmins & Finch, 2006b; Finch, 2007; Gluckman & Hanson, 2005, 2006; Martorell & Habicht, 1986).¹⁸ Height is likely to be affected by household resources and the environment of the mother during her pregnancy and during her child's first few years of life, and her education may increase the effectiveness of these health inputs and facilitate her adoption of best health practices for her child (Barker, 1992, 2005; Rosenzweig & Schultz, 1983; Strauss & Thomas, 2008). Childhood height also forecasts reasonably well adult height, and it is generally correlated with lifetime health and life span, mental or cognitive capacity, learning of skills, and productive outcomes for the individual over the life cycle (e.g., Almond, 2006; Cunha & Heckman, 2007; Elo & Preston, 1992; Floud et al., 1990; Fogel, 2004; Heckman, 2008; Komlos, 1994; Maluccio et al., 2006; Martorell & Habicht, 1986; Schultz, 2003, 2005). Much remains to be understood regarding this dynamic process and the socially and privately optimal allocation of health and development inputs over the course of a lifetime.

Cross-tabulations of surveys and censuses since the 1970s document the significant positive correlation between child survival and the schooling of the mother, even after controlling for the schooling of the father and many other household and community exogenous factors (Caldwell, 1979; Cochrane, O'Hara, & Leslie, 1980; Mensch, Lentzner, & Preston, 1985; Schultz, 1980, 1997). One year more schooling is

associated with a 5–15% decline in early child mortality. It is an open question whether this empirical regularity reflects the knowledge of child health care obtained by better educated women, or women's empowerment that stems from their schooling relative to their partner, which allows them to allocate more household resources to food and child health in response to women's greater altruism toward their children (Schultz, 1990; Thomas, 1990). The impact of mother's schooling also generally impacts more positively than that of the father on other child investments, such as schooling, and this has led many to conclude that public transfers to women are more effective in raising the welfare of children than transfers to fathers. Hence, the structure of the influential Progresa program started in 1997 in Mexico mandated that the mother should receive the income transfer, conditional on her children attending school and receiving regular preventive health care (Gertler, 2004; Parker, Rubalcava, & Teruel, 2008). The schooling of the mother is also generally associated with the favorable physical development of the child, such as birth weight, height for age, and weight for height.

Weight for sex and age, as already noted, is generally transformed into a BMI, defined as the weight in kilograms divided by height in meters squared, and its impact on health is distinctly not monotonic, in contrast to that with height. Waaler (1984) in his study of the Norwegian population (1963–1979) demonstrated in Figure 7 that the relative risk of late mortality varied by BMI, controlling for age, sex, following a

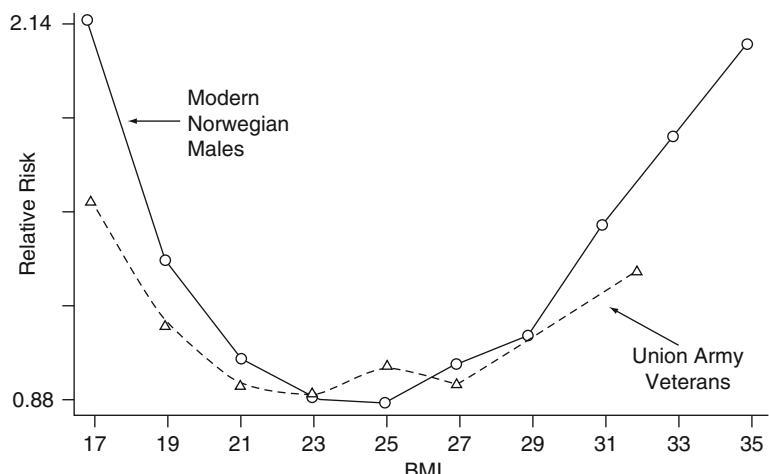


Figure 7 Comparison of relative mortality risk with BMI among men 50 years of age, Union Army Veterans around 1900 and Modern Norwegians. (source: Costa & Steckel, 1997) Note: in the Norwegian data, BMI for 79,084 men was measured at ages 45–49, and the period of risk was 7 years. BMI of 550 Union Army Veterans was measured at ages 45–64, and observation period was 25 years.

U-shaped pattern with respect to BMI for a variety of causes of death or diseases, with above average relative risks of death associated with BMIs less than 21 and over 29. Reporting these patterns for specific causes of mortality among persons over the age of 50, Waaler linked the effects of BMI on health to the functioning of particular organs and the resistance of individuals to specific degenerative diseases. The analogous calculation is performed for the Union Army Veteran's sample of Fogel (2004) from about 1865 to 1900, resulting in a similar pattern (Figure 7) (Costa & Steckel, 1997).

It is hypothesized that a malnourished fetal environment will imprint on genes a metabolic adjustment which will compensate to some degree for the expected lifetime shortages of nutrients facing the child, by converting more dietary calories into weight gains and stores of fat to sustain regular development in lean periods. Then, in the event the child's family experiences an increase in income or improved access to nutrition, the child as an adult will be more likely to become obese or exhibit an unhealthy elevated BMI. As obesity has become more common in high-income countries, its prevalence has increased even more rapidly in a number of relatively low-income countries, such as Mexico (Strauss & Thomas, 2008). It should be recalled, however, that individuals with very low BMI are still the most likely to die in many parts of the developing world, including South Asia and sub-Saharan Africa. Few other large panel studies of risk factors for mortality and morbidity are found in the literature to replicate Waaler's association between height and BMI and older adult mortality in low-income countries. Further replications of Waaler's findings would help to evaluate these anthropometric measures of health outcomes, and how variations in these measures are reproduced by the individual's origins and environment.¹⁹

Su (2005) provides such a comparison over time of the Waaler curve between BMI and mortality for US white males age 50-59. He relates the overall mortality risk to BMI based first on the Union Army Veteran Survival sample followed from about 1860 to 1900, and then for a representative US survival sample from the 1971-1974 NHANES-I to the epidemiological vital status follow up survey in 1993. Su reports, as did Waaler, a U-shaped variation in the relative risk of death with increasing BMI, measured over an 18-year period (Figure 8). Using a variety of methods to fit quintiles in BMI to relative mortality rates starting in about 1860 and 1972, Su shows that the healthier BMI values appear to have shifted to the right or to higher values of BMI, between these two observations of the US population separated by roughly a century. The optimal BMI, namely that with the lowest relative mortality risk, increases from about 21 in 1870s to nearly 26 in 1972, although it may be noted that the lowest mortality segments of the Waaler curve are quite flat in this middle range of BMI.²⁰ Many factors could have contributed to this shift over time in the Waaler mortality curve with respect to BMI in the US, and to the difference between Norway and the US in the 1970s. I do not know of other research accounting for changes in this association between BMI and mortality or morbidity and disability.²¹

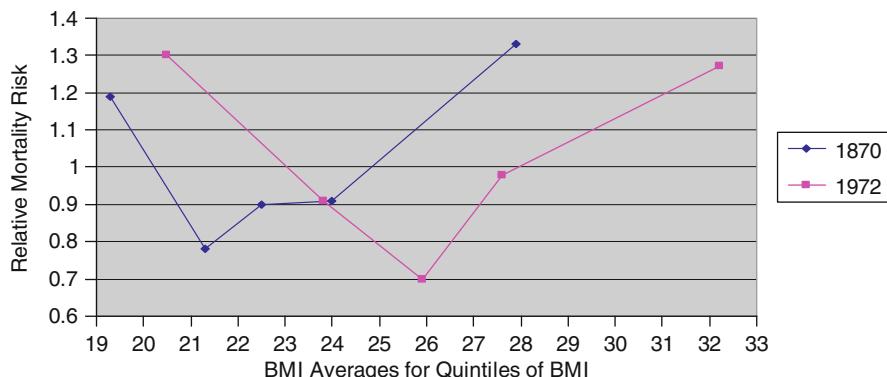


Figure 8 Relative risk of mortality as estimated from Union Army survival sample US white males age 50-59 about 1870, and NHANES-I, US white males age 50-59 about 1972 (source: Su, 2005, Tables 4-6 and Figure 4; model 1 reported based on all deaths).

In addition to height and BMI, birth weight or fetal growth rate or rate of uterine development are related to later life health status and economic performance. The relationship between birth weight and developmental consequences for child survival, health status, schooling attainment, and economic performance of the mature child is interpreted by some as the causal effects of birth weight (Case et al., 2005; see footnote 15). But this inference depends on the assumption that parents have no effect on their child's birth weight and on other prenatal and postnatal health and developmental inputs for their children that might be correlated with those observed, that is, birth weight. More realistically, parents do influence their children's birth weight by their prenatal care and most obviously by their smoking behavior and nutrition, spacing of births, and the timing of prenatal care (Rosenzweig & Schultz, 1983). It is, therefore, likely that birth weight will be correlated positively with other unobserved child health and developmental inputs, because all of these parent health behaviors may be affected in the same direction by unobserved parent preferences regarding investments in child health and development and intergenerational altruism. Other unobserved constraints affecting the family and its children may also impact both birth weight and subsequent health inputs for the children. To avoid this "heterogeneity bias" in estimating the determinants of birth weight, Rosenzweig and Schultz (1983) use arguably exogenous local prices of health inputs and the local availability of maternal health services as instruments to predict variation in parent provision of prenatal health inputs, smoking, age, and parity, and these predicted inputs are then used to explain birth weight in a consistently estimated child health production function using two-stage least squares or instrumental variable methods.

Another basis for identifying exogenous variations in birth weight that is not affected by parent preferences or unobserved constraints on the family involve comparisons between the birth weight of identical twins, as introduced in Section 4.3.

Behrman and Rosenzweig (2004) conclude that fetal growth and presumably uterine nutrition is a significant pathway for lifetime effects on health and economic outcomes of the adult female, but this research does not shed light on how health inputs or behavior of parents might be modified to improve birth weight, or how policies could increase birth weight.²² Black et al. (2007) examine all Norwegian births from 1967 to 1997 to estimate the effect of within twin birth weight differences on infant health indicators (e.g., APGAR scores), completing high school, and labor market success. They also can assess IQ for males who were required to take a military exam at age 18–20, and find within twin 10% differences in birth weight are associated with 1/20 stanine difference in IQ.²³ With the large magnitudes of causal effects of birth weight on human development and adult productivity, the economic case can be made for policies that reduce the incidence of low birth weight children in low-income countries (Alderman & Behrman, 2006). However, the precise policies that would achieve this objective of reducing the proportion of low birth weight children in the world have not been clarified or the cost of such policies estimated.

These indicators of prenatal physical growth and early health endowments may embody not only the effects of parent preferences in and between children, but also the response of parents to independent variation in child initial health status, including as well their gender, and thereby compensate or reinforce the children's prior health endowments. Consequently, indicators of the physical growth of children cannot be viewed as exogenous to their family, except for prenatal growth in the restricted sample of twins, where fixed effects are allowed for each pair of twins. In standard samples of all births, birth weight and subsequent development of the child are likely to reflect parental choices informed by parent understanding of the child health production technology, their preferences for different health and welfare outcomes for their children, to their child's health which can clearly also influence postnatal health investments in their different children.²⁴

Studies of differences between twins, reviewed earlier, also shed light on the parental tradeoff of quality and quantity of children. They suggest the adjustment in quality to an exogenous increase in fertility is generally negative, but it is absolutely smaller and more complex than association across all births or twins. Behrman and Rosenzweig (2004) show that the exogenous differences in fetal growth rate between identical female US twins are associated positively with twin differences in schooling and wage rates, confirming that this initial endowment of a child is reinforced by later investments in child quality. Rosenzweig and Zhang (2009) estimate from a sample of Chinese twins the consequences of exogenously increasing family size due to twins on child quality. In a high-income, low-fertility Norwegian population, Black et al. (2005) finds the occurrence of twins leads to increased family size, but does not significantly lower child quality as measured by schooling of the earlier born siblings.

The Norwegian parents appear to compensate for the lower birth weight of their twin children, as Griliches' (1979) concluded from his early survey of US sibling studies.

The task of estimating how height, BMI and birth weight affects the child's later adult's life requires that the researcher first understand what exogenous factors affect variation in these early health indicators, and more specifically the variables that affect these health indicators which can be influenced by policy. In other words, how is health human capital produced? For example, favorable rainfall in a location in the year before an individual is born into an agricultural society may be associated with an increase in height of this adult. In Indonesia local rainfall deviations predict the height of females born locally in that year, and also increase the probability that these women marry a taller husband, who has higher expected earnings, and she will enjoy higher *per capita* consumption in her household (Maccini & Yang, 2006). Estimates based on the actual (all of the variation in observed) height, and not just the component of height related to the rainfall instrumental variable, would mix together genetic factors permutating height in the population which are hypothesized to be less closely related to lifetime productivity than the local rainfall-induced component of height, which could be thought of as simulating a quasiexperimental effect of variation in reproducible health human capital. In several country studies that use as instruments for the offspring's adult height and BMI the socioeconomic background characteristics of the parents, the local health infrastructure, and disease environment at birthplace. These IV estimates of the effect of height and BMI on the offspring's log wage rates tend to be a larger positive value than if the association of height and BMI on log wages are estimated directly by OLS (Savedoff & Schultz, 2001; Schultz, 2003, 2005). This tendency for instrumental variable estimates of the apparent productive returns to anthropometric health human capital indicators to be larger in absolute values than the simple partial correlations is consistent with the previously advanced hypothesis that the reproducible health human capital component of height, identified by their association with socioeconomic instruments or time series deviations in weather operating on income and possibly disease vectors, has a relatively larger effect on the child than does the variation in height unexplained by these socioeconomic variables, which is more likely to be predominantly genetic variation.²⁵

Table 3 illustrates these empirical patterns for adult height and wages for three countries, Ghana in 1987-1989, Brazil, 1989, and the United States in 1989-1993 (Schultz, 2005). The coefficient on height in centimeters in a standard log wage equation, separately for men and women, age 25-54 who report hours and earnings.²⁶ The first column (a) reports the direct OLS coefficient which implies a male 1 cm taller receives a wage in Ghana that is 1.5% higher, and females 1.7% higher. In Brazil the association is 1.4% and 1.7% higher wages associated with being 1 cm taller for males and females, respectively. In the United States where nutrition is better and people

Table 3 Estimates of the partial association between height and log hourly wage in Ghana, Brazil, and the United States

Country, year, survey, age, and sex	Ordinary least squares	Instrumental variable			Sample size	Mean (standard deviation) Height (cm)			
		Regional and household characteristics	Regional characteristics and parent education	Ethnic or race groups					
<i>Ghana, 1987-1989</i>									
Living standard measurement survey age 25-54									
Males	0.0154 (4.78)	0.0943 (3.71)	0.0934 (4.27)	-0.1017 (3.66)	2876	169 (6.49)			
Females	0.0167 (4.04)	0.1091 (4.14)	0.1522 (6.56)	-0.0765 (2.37)	2754	158 (6.16)			
<i>Brazil, 1989</i>									
Health and nutrition survey age 25-54									
Males	0.0140 (10.3)	0.0775 (9.34)	NA	0.0701 (12.4)	7808	168 (7.20)			
Females	0.0166 (8.13)	0.1086 (8.12)	NA	0.0862 (8.11)	4047	156 (6.63)			
<i>USA, 1989-1993</i>									
National labor force survey of youth age 24-36									
Males	0.0056 (5.12)	0.0210 (1.75)	0.0353 (4.92)	0.0183 (3.08)	18,491	178 (7.65)			
Females	0.0043 (3.62)	0.0453 (1.93)	0.0473 (4.06)	-0.0023 (0.48)	16,695	163 (6.98)			

Source: Schultz (2002), Appendix Table.

Notes: beneath the estimated coefficient on height in the log wage equation are reported in parentheses the absolute value of the *t* ratio. In addition to height in the wage function, the standard specification includes a spline in years of schooling completed by levels, postschooling years of experience (*age-schooling-6*), experience squared, and rural residence.

are taller on average (10 cm taller than Brazil) a gain of 1 cm is associated with only 0.6% higher wage for males and 0.4% for females. Columns (b) reports the instrumental variable estimate for height predicted on the basis of regional and household characteristics at birthplace, and these IV estimated effects of height on log wage increase three- to fivefold. The variation in height systematically associated with socioeconomic background and regional birthplace is more steeply associated with wage variation than is the entire distribution of the variation in height including presumably genetic components. If years of schooling of the mother and father are added to the list of instruments in Ghana and the United States, the IV estimates in column (c) of the effect of height remains roughly the same but tends to becomes more statistically significant as the power of the first-stage regression increases and these overidentification restrictions are accepted by conventional tests (Schultz, 2002; Wooldridge, 2002). The IV estimates based on birthplace and family socioeconomic status are believed to identify the effect of variation in height primarily due to reproducible health human capital, and therefore down weights the genetic components of height. The IV estimates of height on wages may represent the causal effect of height changes that might be induced by nutrition and health interventions, whereas the OLS estimates portray to a greater extent the association between genetic and other sources of variation in height for which social welfare policy probably has little leverage.

Table 4 reports the joint estimation in Ghana (1987–1989) of the wage effects of four forms of human capital: years of education, migration from birthplace, adult current BMI, and adult height. First it is assumed that all four forms of human capital are homogeneous, exogenous, and measured without error. Under these working assumptions, OLS estimates as reported in columns (1)–(4) have desirable properties. Because the four forms of productive human capital tend to be positively associated across people, the conventional approach to estimating of wage functions which tends to omit several of these human capital factors would be inclined to attribute too large a productivity effect to the included form(s) of human capital, typically only schooling. The conventional returns to schooling that exclude these additional human capital variables are only moderately diminished by the inclusion of the other three forms of mobility and health human capital, declining 15% for males from 0.052 to 0.044. But if these forms of human capital are endogenous, heterogeneous, or measured with error, column (5) provides consistent IV estimates, based on the individual's birthplace health and schooling environment, and parent education and agricultural occupation. Although this IV approach to estimating the returns to the reproducible human capital component of these four forms of human capital does not change significantly the estimated wage returns to schooling or to migration, it often increases the coefficient on BMI significantly, and increases the coefficient on height several fold. Hausman tests of the exogeneity of these forms of human capital imply education and migration appear to be exogenous, whereas the health variables—height and BMI—are generally

Table 4 Coefficients on four indicators of human capital inputs in log wage equation in Ghana, 1987–1989

Gender and variable ^b	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV
<i>Males: sample size 3414</i>						
Education	0.0521 (11.7)	0.0475 (10.7)	0.0449 (10.1)	0.0437 (9.86)	0.0445 ^a (2.46)	0.0445 (9.95)
Migration		0.388 (7.48)	0.360 (6.97)	0.348 (6.75)	0.218 ^a (2.26)	0.295 (5.34)
BMI			0.0542 (6.93)	0.0530 (6.80)	0.0793 ^a (1.95)	0.0658 ^a (1.76)
Height				1.48 (5.02)	5.69 ^a (3.45)	5.56 ^a (3.58)
<i>Females: sample size 3400</i>						
Education	0.0481 (9.23)	0.0425 (8.22)	0.0395 (7.69)	0.0375 (7.26)	0.0356 ^a (2.69)	0.0346 (6.56)
Migration		0.617 (9.85)	0.537 (8.55)	0.531 (8.46)	0.361 ^a (2.98)	0.447 (6.51)
BMI			0.0425 (7.72)	0.0420 (7.63)	0.0981 ^a (4.11)	0.0881 ^a (4.32)
Height				1.29 (3.63)	7.48 ^a (3.44)	7.62 ^a (3.80)

Source: Schultz (2003), Table 4.

^aVariable is assumed endogenous and estimated by instrumental variables, which include parent education and occupation, local health infrastructure, and food prices.

^bOther control variables include region of birth, ethnic group, age and season of interview. Beneath regression coefficient is the absolute value of the *t* ratio in parentheses in columns (1)–(4) and asymptotic *t* ratio in columns (5) and (6).

endogenous (Schultz, 2003).²⁷ The interpretation proposed earlier is that these forms of health human capital are heterogeneous and the reproducible IV component of height and BMI are significantly more productive than the remaining variation in these anthropometric indicators of adult health associated with genetic variability in the physical characteristics of the worker.

Table 5 illustrates that adult height and years of schooling have increased rapidly in some countries, such as Brazil, Vietnam, and Cote d'Ivoire, by nearly a centimeter per decade for height and a year of schooling or more per decade, but not in stagnating

Table 5 Means and standard deviations in parentheses of height and schooling by country, for selected age groups, by sex, and change over 30 years

Country and age	Height in cm (standard deviation)		Schooling in years (standard deviation)	
	Female	Male	Female	Male
<i>Ghana, 1987-1989</i>				
Age 25-29	158.53 (6.25)	169.46 (6.63)	5.29 (4.97)	8.29 (5.09)
Age 55-59	156.93 (5.96)	169.00 (6.51)	2.12 (4.29)	5.68 (5.97)
Change	+1.60 (+0.29)	+0.46 (+0.12)	+3.17 (+0.68)	+2.61 (-0.88)
<i>Cote d'Ivoire, 1985-1987</i>				
Age 25-29	159.11 (5.67)	170.11 (6.70)	2.78 (3.99)	6.12 (5.07)
Age 55-59	157.57 (6.11)	168.48 (6.88)	0.23 (1.32)	2.30 (3.98)
Change	+1.54 (-0.44)	+1.63 (-0.18)	+2.55 (+2.67)	+3.82 (+1.09)
<i>Brazil, 1989</i>				
Age 25-29	156.27 (6.62)	168.90 (7.27)	6.36 (4.31)	5.66 (4.22)
Age 55-59	153.16 (6.59)	165.79 (7.47)	2.21 (2.86)	2.52 (3.16)
Change	+3.10 (+0.03)	+3.11 (-0.20)	+4.15 (+1.45)	+3.14 (+1.04)
<i>Vietnam, 1992-1993</i>				
Age 25-29	152.16 (5.39)	162.10 (5.39)	7.90 (3.21)	8.35 (3.38)
Age 55-59	148.73 (5.64)	159.19 (5.93)	3.74 (2.59)	6.48 (3.82)
Change	+3.43 (-0.25)	+2.91 (-0.54)	+4.16 (+0.62)	+1.87 (-0.44)

Source: Schultz (2005) Table 3.

Ghana. Wage functions are estimated for Brazil in 1989 that includes years of completed schooling and height, where height is estimated by instrumental variables as in Table 3. Had these proportional gains in Brazilian wages associated with schooling and height in 1989 been invariant over time from 1960 to 1989, the observed advances across cohorts born from 1940 to 1960 in height would account for a 4.1% increase per decade in male wages, and 5.8% increase per decade in female wages.²⁸ The advance in schooling in Brazil is associated with an even larger increase in wages of males of 16% per decade, while for females the schooling gained across birth cohorts would account for a 22% growth. Although the probable wage effects of secular growth in both forms of human capital are large, the impact of increased schooling on wage growth, according to these estimates for Brazil from adults born from 1940 to 1960, is able to account for two to four times more growth than that associated with the increase in height health human capital (Schultz, 2005).

Do instrumental variables for local public services, health input prices or subsidies, and parent socioeconomic status summarize variation in health human capital embodied in height or achieved through improving fetal growth? They may also capture secular improvements in fetal and early child development that widely parallels economic growth, improved nutrition, and increased access to preventive child health care. Research on the productive contribution of health needs to describe explicitly geographic variation in child health interventions, relative prices of food, and gains in household real income opportunities for female and male workers that are expected to affect adult height. If these local health policies, income and relative price instruments suggest pathways for policy to affect height, more focused IV estimates are now needed to better approximate local average treatment effects of actual policy interventions that could be scaled up by governments (Imbens & Angrist, 1994). Health policy measures will eventually need to be assessed by allocating randomly across localities the prescribed policy interventions, and then relate these policies to the regions where individuals were born and grew up in order to obtain more credible policy-relevant estimates of how public and private programs affect child and adult health and thus impact adult productivity.

6.3 Pathways from fetal and early child development to adult productive life span

Research primarily in epidemiology and demography is being extended in economics which links the early health environment of individuals to their longevity, health status, and economic performance, and relates the consequences specifically of health for welfare outcomes and productivity over an individual's lifetime. This is a brief and very selective summary of a few studies from this large multidisciplinary literature. More detail is found in Alderman, Hoddinott, and Kinsey (2006) and Strauss and Thomas (2008) from an economic perspective, and Finch (2007), Gluckman and Hanson (2006) and Finch and Crimmins (2004) from an epidemiological and medical perspective, though it should be noted that approaches to statistical modeling differ between fields making some comparisons tenuous.

The “fetal origins hypothesis” of Barker (1992, 1994, 2001, 2005) postulates that some chronic health conditions, such as cardiovascular and lung diseases, are caused by the environment *in utero* which affects fetal growth and development at various critical periods in gestation. The initial evidence for the hypothesis was the correlation between place of birth in the UK and risks of mortality as older adults, where it was reported that being born in a high mortality period and location was positively associated with the level of late adult mortality, disproportionately due to the failure of specific systems of organs, such as the heart and lungs (Barker, 1992). However, many other factors could be related to birthplace health conditions and to later adult chronic health problems, reducing the credibility of these early epidemiological studies as tests of causal hypotheses. More discriminating tests of these hypotheses have sought to specify a “narrower”

pathway for a “treatment” of the fetus, and differential effects on the treated, say by gender or economic status, and construction of matched control populations, followed often in a representative sample survey or census. Discontinuity designs are also being used to target the statistical tests to changes occurring at a particular time, given existing time trends. Because chronic degenerative health problems begin to cause substantial numbers of deaths only after middle age, these studies require matching birth and early childhood conditions with the cohort’s health and economic outcomes many years later. With such long lags between treatments and the consequences on health and productivity, the problem of sample selection bias arises, simply due to mortality removing from the birth cohort the more frail individuals, or due to other forms of sample selective attrition, such as due to selective migration or survey nonresponse.²⁹

Doblhammer (2004) examines mortality variation among the elderly by month of their birth to refine our understanding of the fetal origins hypothesis. She finds seasonal patterns among those in the Northern hemisphere, such as Denmark, where those who are born in the spring have below average life expectancy at later ages.³⁰ With the rate of growth of the fetus reaching a maximum in the third trimester, it is believed that if the mother experiences nutritional stress in this critical period as is more common with severe winters and late spring seasons, healthy development of the child’s heart and lung systems tend to be jeopardized (Barker, 2005). The seasonal pattern among natives in the Southern Hemisphere, such as Australia, is roughly reversed. Moreover, migrants born in the North and residing the South exhibit month of birth effects on their elderly life span similar to nonmigrant natives in the North. The regional effect of birth month on elderly longevity thus appears imprinted through fetal growth and early development, and cannot be attributed to the differences in the disease environment or health infrastructure that could affect mortality among the elderly differently in the two hemispheres. An active area of research explores what specific diseases, health conditions, and institutional arrangements are associated with the higher elderly mortality due to slower fetal growth rates or particular environmental risks. How are these limits to growth translated into higher mortality among the elderly, what chronic and acute health problems are observed with advancing age due to less adequate fetal development, how do these observable health conditions affect the economic productivity of adult workers and impact their health investments, and modify health inequalities by initial socioeconomic status?³¹ There are significant season-of-birth effects on height and weight for age Z scores of Indian children under age 3 in the three rounds of the National Family Health Survey collected in 1992, 1998, and 2005 (Lokshin & Radyakin, 2009). Births in the monsoon season exhibit lower height and weight, presumably due to greater malnutrition in this lean season and greater disease. These season-of-birth effects also interact with maternal education and a crude household wealth index, suggesting that socioeconomic behavior is a factor modifying the impact of season of birth on early child health and expresses itself in birth weight.

If seasonal variations in these conditions are important, then additional insults to fetal growth may be caused as well by unpredictable drought, floods, weather, pollution, epidemics, or radiation accidents. Narrowly focused analyses of birth cohort data may identify the physical and economic consequences of these “quasi-natural experiments” on aspects of the fetal development, impacts on the mortality of older populations, and on other life-cycle events such as schooling, the formation of cognitive skills, earnings capacities, migration and consumption opportunities. Do private and social insurance schemes, or food relief programs, buffer pregnant women and young children from these economic and biological shocks, or are these resources disseminated beyond vulnerable target groups perhaps due to corruption or political economy, and loose their effectiveness to raise child health?

A study of the 1918 flu pandemic by Almond (2006) illustrates how this literature combines a variety of data sources to describe connections between environmental shocks at fetal origin and delayed measures of outcomes among mature members of a birth cohort. The severity and timing of the flu infections varied by states in the United States, providing instruments to identify the likelihood a child born before, during, and after the pandemic had been affected *in utero* by the flu infection that her mother might have experienced. The cohorts born from January to September 1919, in the wake of the pandemic, exhibit “deficiencies” which can be documented in the US Population Censuses of 1960 and 1980 and can be attributed to the epidemic. For example, the male children of mothers who were likely to be infected by the flu were about 15% less likely to complete high school, and report wages which are 5–9% lower than those born earlier or later.³²

Historical studies have frequently found that periods of economic and health crises are associated with birth cohorts that experience greater late life mortality than do birth cohorts somewhat older and younger, and these extra deaths are often linked to cardiovascular and lung diseases and type 2 diabetes (Elo & Preston, 1992). Alters and Oris (2006) report in several preindustrial Belgium communities (1750–1830) that being born in years of relatively high mortality is a significant predictor of high mortality in middle and late ages for the survivors of these cohorts, even after controlling for a variety of individual socioeconomic characteristics and community effects. These periods of health crises are often associated with high food prices, and are more likely to be associated with higher late adult mortality among those originally born into poorer, more vulnerable, households in crises years. Bengtsson and Brostrom (2006) find in Southern Sweden from 1829 to 1894 the disease load that children are exposed to in the year of their birth is significantly related to their higher mortality after age 55, even after controlling for their landed/landless state (i.e., an indicator of economic status) at birth and at the end of their working career. Additional studies are collected by (Bengtsson 2009).

Preston, Hill, and Drenenstedt (1998) note evidence among African Americans that being born and raised as children in farm families is associated with lower mortality as

an adult in the early twentieth century United States. They hypothesize that this is due to the childhood disease conditions being more favorable at the start of the century in the rural South than in urban areas of the United States in either the North or the South. Bleakley (2007) finds that the eradication of hookworms after 1910 in the American South contributed in those areas where the infection was most common before the eradication campaigns to increases in school enrollment, attendance and literacy compared to neighboring areas which were subject initially to lower levels of hookworm infection. In the longer run, the eradication was associated with gains in personal income and increased private wage returns to schooling. He argues this one disease eradication campaign was responsible for closing one-half of the gap between the average years of schooling in the South and North in the United States, and to closing one-fifth of the income gap between these regions. With the evidence of an increase in returns to schooling, the hypothesis that child quality and quantity are substitutes is confirmed by the steep decline in fertility following the eradication of hookworm (Bleakley & Lange, 2009). Crimmins and Finch (2006b) suggest that the reduction in childhood infections with their burden of inflammation is responsible for secular trends in adult height and longevity of birth cohorts over the twentieth century in the United States, recognizing that height is an external anthropometric indicator (observable proxy) for the internal (unobserved) health human capital that impacts mortality (Fogel, 2004; Ginsburg et al., 1998).

Vitamin D deficiency is also biologically linked to bone development and height (e.g., rickets in children and osteomalacia in adults) as well as a variety of chronic diseases affected by fetal development, including cardiovascular, autoimmune, cancers and type 1 diabetes (Holick, 2004; Raiten & Picciano, 2004). Because vitamin D production is stimulated predominantly by sunlight (namely, UVB radiation) and is affected by skin pigmentation or race, variation in height has been studied as a function of latitude, location, occupation, and race (Carson, 2008), and its recent reemergence in the US population may be related to increased breastfeeding without vitamin D supplementation or the use of sunscreen which reduces UVB absorption.³³

Van den Berg, Lindenbroom, and Portrait (2006) examine the business cycle conditions in the year of birth in the Netherlands from 1812 to 1920s, and find significant effects of the business cycle, presumably operating through early life conditions, which are negatively related to end of life mortality rates. They conclude that especially among the poor, food, housing, and health care available at the time of birth affect later health and disability of the population. Although there is growing confirmation of the link from various sources of fetal stress to diminished schooling, cognitive capacity, and earnings for the resulting birth cohort, this literature has made less progress in describing how specific policies can effectively protect individuals during this vulnerable stage of fetal and early child development and mitigate the longer-run consequences of poverty and health shocks on lifetime health status, productivity, and longevity.³⁴

7. MACROECONOMICS OF FERTILITY AND DEVELOPMENT: DIMINISHING RETURNS TO LABOR, HUMAN CAPITAL AND SAVINGS, AND CHANGING AGE COMPOSITION

Several lines of macroeconomic reasoning are used to infer how declines in fertility during the demographic transition could affect economic growth and development. First, there was Malthus (1789), who relied on the classical economic idea of diminishing returns to labor when workers are employed with a fixed supply of complementary resources, such as agricultural land. The high fertility in low-income countries in the first two decades following the Second World War was viewed in a Malthusian framework as an impediment to economic development. Population growth increased in these poor countries from 0.5% per year in 1900, to 1.2% by 1940, and doubled again to 2.5% by 1960 (Kuznets, 1966; United Nations, 2003). Although due to reduced mortality rather than increased fertility, this “population explosion” appeared to overwhelm the capacity to accumulate capital and employ productively such rapidly growing populations. A demographic poverty trap could arise as Malthus had hypothesized, and slow economic development in these poor countries unless fertility declined quickly (Coale & Hoover, 1958).

However, the Malthusian link between rapid population growth and slower economic development was not evident to Kuznets (1967) from his analysis of historical data, nor did Malthus’ forecasts materialize in subsequent decades, as savings rates increased in many parts of the developing world, and human capital formation and technical change increased total factor productivity and achieved unprecedented growth in output per worker (Johnson, 1999; National Research Council, 1986). Technical change and possibly behavioral responses to the decline in mortality may have outweighed the diminishing returns to labor foreseen by Malthus, allowing growth in *per capita* income in Latin America and Asia. Even the crude associations across countries between population growth rates and growth rates in availability of calories *per capita* are insignificant until 1985, and only thereafter from 1985 to 1995 is the overall association negative. The growing scarcity of food supplies recently in countries with rapid population growth appears to be explained by declining net imports of food, whereas domestic crop production has continued to outpace population growth (Kravdal, 2001).³⁵

The second macroeconomic framework used by economists to assess the implications of the fertility decline focused on a life-cycle pattern of consumption and savings. It was expected that high fertility would depress aggregate rates of savings and thereby discourage economic growth, because the proportion of the population in their most productive ages would not increase until fertility began to fall secularly. In a classic formulation of the issue by Modigliani and Brumberg (1954), they assumed that the marginal utility from consumption diminishes as the level of consumption increases,

lifetime utility is time separable, and the age profile of adult productivity rises and falls over the life cycle. Adults in this setting would be motivated to accumulate savings in their most productive periods to sustain their consumption in old age when their productivity declines. These widely accepted assumptions led to the hypothesis that savings rates would rise and fall for a cohort over its life cycle. Holding factor productivity constant, the decline in fertility would contribute over time to an increase in the proportion of the population between the ages of 35 and 55, and thus to an increase in the average national savings rates, all else equal. Some studies have reported within East Asian countries a relationship over time between age composition changes and savings rates changes that could have contributed to the Asian miracle. But this relationship is fragile and evaporates when lagged savings is not included as an exogenous regressor or it is treated as endogenous, or country-specific time trends in savings rates are included in the estimated model (Higgins & Williamson, 1997; Schultz, 2004a). Counterexamples are also notable. Savings rates have stagnated in other regions, such as Latin America, even though the region experienced a relatively early demographic transition which generated the changes in age compositions as observed in East Asia. In a country such as India, in which the demographic transition has been more gradual and the changes in age composition more modest, savings rates have nonetheless increased (Schultz, 2004a). There are also inconsistencies between micro- and macroevidence. Microeconomic studies of household surveys do not find the pronounced life-cycle variation in savings rates in either high- or low-income countries as postulated by Modigliani (Deaton & Paxson, 1997). Although the life-cycle model of consumption and savings behavior remains a plausible conceptual framework for studying many macroeconomic issues, it does not provide a satisfactory explanation for savings and growth in terms of changes in the age composition of low-income countries.

Labor force participation rates for males also rise and fall with an individual's age in many populations, leading to the expectation that labor supply *per capita* would tend to rise as fertility falls, for at least three decades into the demographic transition. In the longer run, 40 or 50 years after fertility begins to decline, the share of the male population in the labor force is expected to decline, as the fraction of the population over age 50 increases more than the fraction of children falls, as illustrated in the rapid aging of contemporary Japan. This expectation is reinforced as the age-specific death rates among the elderly continue to decline.

However, most of the increase in labor supply per adult following the demographic transition in East Asia and Latin America is due to the increased market labor force participation of *women*, whose participation profiles by age differ across countries and changes over time. The female participation rates outside of the family in the wage labor force shows a general tendency to increase with women's education and the level of development in a country (Durand, 1975; Schultz, 1990). It is also likely that the women who enter the market labor force are also those who bear fewer children, and who are thus

able to engage more readily in economic activities outside of their household. The underlying factors changing fertility may thus be responsible for changing a variety of other family coordinated productive behaviors, including most centrally women's market labor supply. Therefore, changes in fertility and age composition cannot be treated as exogenous causes for change in labor inputs to the market. Both fertility and female labor supply are jointly determined in response to male and female market wage opportunities and nonearned income, among other factors (Mincer, 1963).

Schooling and vocational training are concentrated among youth, and thus if enrollment rates are held constant, these private and social investments in the human capital of youth will increase on a *per capita* basis for a decade or two after fertility starts to decline. Moreover, if the decline in family size encourages parents to invest in more years of schooling on average for their children or in higher quality education, this increase in child "quality" may be an indirect consequence of fertility decline that could favor long-run economic growth. In other words, if parents behave as if the schooling of their children is a *substitute* for the number of children they have, population policies which facilitate the decline in fertility by reducing the cost of birth control can contribute to the formation of more human capital as explored in more detail at the microlevel in the next section. A model is required to extrapolate from population policies that reduce fertility to assess how they modify other lifetime family resource allocations, notably the reallocation of the time of family members, and the investment in human capital and various forms of physical capital. The country-level correlations between age composition and economic growth do not clarify the causal pathways from the resources and opportunities of families, to their health outcomes and fertility, which drive changes in the age composition of the population. There is no reason to assume that age profiles of endogenous household behavior, such as savings or female market labor force participation, will remain unchanged when fertility declines during the demographic transition. Depending on whether the behavior complements or substitutes for fertility, the behavior may decrease or increase, respectively, and these evolving forms of life-cycle behavior may contribute to economic development or stagnation. The decline in fertility is also generally associated with increased schooling and in particular women's postschooling investment in career skills that will modify systematically the age profile of market earnings for women and men, and potentially spill over to affect savings and consumption behavior of their families.

An alternative research agenda to the study of patterns across and within countries seeks to understand how *individual* household behavior responds to arguably exogenous changes in the individual and family environment, such as changes in relative prices, wages of women, men, and children, returns to schooling and other human capital and other asset, and technical change, which may be biased toward better educated labor (Mincer, 1963; Schultz, 1981). For example, it might be hypothesized that the decline in fertility is caused by technical change, which raises the returns on human capital, and

thereby encourages parents to invest more in their children's human capital. Parents consequently have fewer children, and women increase the share of their adult lives working in activities other than child care (Foster & Rosenzweig, 2007; Galor & Weil, 1999, 2000; Rosenzweig, 1990). The intercorrelations between fertility and these family lifetime time allocations, investments, and outcomes do not describe the causal impact of fertility nor can these intercorrelations be treated as consequences of fertility and used as a reason to promote fertility reducing population policies (Behrman, 2001; Birdsall, Kelley, & Sinding, 2001; Bloom, Canning, & Sevilla, 2002). Much empirical evidence substantiates the view that fertility is subject to choice in many settings and is determined simultaneously with these other family behaviors (Rosenzweig & Wolpin, 1980a,b; Schultz, 1981, 1997). Cross-country regressions are notoriously difficult to interpret as representing causal relationships, because in the case of mortality and fertility they tend to be dominated empirically by poorly understood secular trends, and because mortality and fertility are jointly determined by unobservables at the family level. This suggests that empirical evidence be analyzed at the household level and organized to identify plausible structure in models by exploiting variation over time as well as across individuals (Moffitt, 2005; Schultz, 2004b).

8. MICROEVIDENCE OF DETERMINANTS OF FERTILITY AND CONSEQUENCES OF POLICY ON FERTILITY

There are few social experiments where individuals or communities are randomly assigned to population programs or policies, such as family planning programs that subsidize the adoption and use of birth control. The consequences of a population policy on fertility may generate long-term indirect or "cross-effects," in addition to their direct effect on lifetime fertility and the timing of births. These long-term cross-effects of fertility reduction may constitute important social "externalities" affecting the net social cost of these welfare programs (Schultz, 2008a). The concluding part of this section reviews the case of Matlab, Bangladesh, where a social experiment reducing the cost of birth control has been implemented, and the long-term consequences have been estimated. But to otherwise evaluate these consequences of policy-induced fertility change, an economic model of fertility as a constrained choice may be useful to help specify an instrument for exogenous variation in fertility as a determinant of family cross-effects. For an instrumental variable estimation strategy to be valid, the instrument must be correlated with fertility, but be arguably uncorrelated with the long-run family behavioral outcomes that are hypothesized to be affected by exogenous changes in fertility over the life cycle. The list of likely family consequences of fertility variation includes the health of women, their labor supply, women's investment in their own human capital relevant to their productivity in the labor market and home, for which the return may change as she gains more technical control over the timing

and number of her births. In other words, the simple association (OLS) between the number of births and related family outcomes cannot be interpreted as an estimate of a causal effect, whereas well-specified two-stage (IV) estimate of these long-run cross-effects of fertility may inform policy.

8.1 Income and price effects affecting the quantity and quality of children

Economists have considered a variety of factors potentially responsible for fertility change and in particular the secular decline in fertility since about 1870 in high-income countries, and since about 1965–1985 in a number of low-income countries. Malthus (1798–1830) hypothesized that increasing real wages would encourage youth to marry earlier, and this would lead them to have more children. In other words, as Adam Smith has postulated, marriage and children were normal goods whose demand would increase with income in the absence of price changes. Becker (1981) proposed the idea that because children were more “time-intensive” to produce by parents than other forms of their consumption, rising real wages would increase the opportunity cost of children and this price effect would offset the income-driven demand for more children.

The idea that parents substitute offspring “quality” or the resource intensity of each of their offspring for the number of offspring is consistent with the influential ideas of Darwin (1859), and is reflected in the work of various social sciences and evolutionary biology (e.g., Dawkins, 1989; Wilson, 1975). In the formulation of Becker and Lewis (1974), the quality and quantity of children are viewed as substitutes in producing a combination of “services” of children that parents value. But Becker and Lewis make the additional assumption that the income elasticity of demand for child quality is positive and larger in magnitude than the income elasticity of demand for quantity of children. Thus, their framework implies that increases in income may be sufficient by themselves to account for the shift of parents toward demanding higher quality children which could, other things equal, reduce their fertility. It offers an explanation for the demographic transition that does not require an exogenous increase in the relative price of children.

Galor and Weil (2000) conclude that increases in income levels alone are insufficient to explain the timing of the decline in fertility or the demographic transition across high-income countries. They also postulate that human capital complementing technical change occurred after 1870, and liberated European countries from Malthusian stagnation. But few empirical indicators are described that govern the timing or character of this “second industrial revolution,” and their theory does not help to forecast the timing of this technical change in low-income countries or how to foster its advancement. Lucas (2002) hypothesizes that social externalities of human capital were especially strong in urban areas of Europe, and urbanization could account for the increase in labor productivity as an increasing share of the population moved to the European cities during the

industrial revolution. However, the elevated mortality in urban areas until the end of the nineteenth century must have also dampened private and social returns to investment in human capital in these rapidly industrializing, but unhealthy, cities, especially in England and Scotland (Tanner, 1982). The unified growth theory that incorporates the outlines of the demographic transition and the human capital revolution remains a challenging field of economics (Galor, 2005). This integrated description of the industrial revolution and demographic transition requires, however, a stronger empirical foundation with more interdependent predictions for empirical study.³⁶

If the fertility decision is made by utility maximizing parents who are altruistic toward their children's consumption, Becker and Barro (1988) show in their intergenerational growth framework that fertility is an increasing function of the long-term real rate of interest and the probability that children survive. More rapid technical change (biased toward human capital) increases child rearing costs, and public transfers to the elderly of pensions and health care reduce the value of surviving children to parents as a means to support their consumption in old age. This dynamic family growth model draws attention to new constraints on fertility and the role of the welfare state, but these constraints may not all be exogenous from the point of view of families or societies.

What leads the state to provide transfers to the elderly? Why might the majority of citizens prefer to receive old-age support from the state, rather than to invest in rearing more of their own children? Is the increase in cost of child rearing for parents a response to the increasing returns to child human capital? What conditions then stimulate this increase in returns to human capital other than a black box of technical change, and why does technical change become a complement to educated labor in the twentieth century, when it appears to have been a complement to unskilled labor in the nineteenth century (Goldin & Katz, 1998, 2008)? How much of the decrease in child mortality is linked to differences in preferences for, or capacity to produce, child health, associated possibly with the increased schooling of mothers?

Many questions are skirted by Becker and Barro (1988) when they assume a priori technical change complements human capital, child mortality declines independently, and the state provide old-age pensions, while their aggregate model depends on importing all of these critical developments from outside of their intergenerational growth model. Mincer's (1963) observation that the gender gap in schooling and wages could be an important factor driving up the opportunity cost of children is lost in Becker-Barro's dynamic model, where there is no distinction between the schooling, productivity, or productive roles of men and women. The nonunitary bargaining model of household decision making as proposed by Chiappori (1992) may provide a more realistic framework to study the determinants of such behavior as fertility and family labor supply, and to assess their consequences.

More widely accepted is the specific notion that the returns to human capital, or more specifically schooling, began to increase at the end of the nineteenth century,

which induced parents to increase their demand for schooling of their children, and led them to substitute away from having a large family. Reliable empirical evidence of the returns to schooling is scarce before the 1940 US Census, with the exception of a census of Iowa conducted in 1915, which asked for information on both income and schooling (Goldin & Katz, 2000). The substantial wage returns to secondary school and higher education observed in Iowa in 1915 have gradually declined according to the subsequent census samples from Iowa for 1940, 1950, and 1960. Earlier insights into returns to skills or schooling relies on US wage data not by schooling but for blue and white collar workers (Goldin & Katz, 1998, 2008). It is reasonable that technological change in the early twentieth century complemented skills or schooling, and raised returns to schooling, especially for women, until the aggregate supply of education in the population caught up to the new aggregate derived demands for these skills. A second wave of increasing returns to higher education emerges in the 1970s, and may again be attributed to a skill bias in global technical change at the end of the twentieth century, adding to the constraints decreasing fertility (Acemoglu, 2002).

Mincer (1963) emphasized that children were produced primarily by women, and therefore any increase in women's wages relative to men's wages would raise the relative cost of child rearing, and encourage women's work outside of the home and discourage fertility, other things being equal. If women's education increased their productivity only in the market labor force, as assumed by Willis (1974), the decline in fertility and increase in women labor force participation could be attributed to an exogenous increase in women's schooling. Although women's schooling has been gradually catching up to men in many parts of the world (Schultz, 1981, 1990, 1995), there is little agreement on the underlying causes for this trend. Technological change may be attributed a role in increasing the share of services in national income, which favors derived demands for traditional female occupations, such as nursing, teaching, and clerical work, and thus to increase women's wage returns to schooling. In a growing number of countries female wage returns to schooling exceed male returns at the secondary and higher education levels (Schultz, 1995). As the overall gender gap in wages closes, and the schooling of women catches up to that of men, the private opportunity cost of childbearing should increase further, and the decline in fertility may continue, unless the private costs of child care are absorbed by the public sector of the welfare state (e.g., Sweden).

8.2 Family planning policies and fertility: Cross-country relationships

For couples who want fewer children than they would otherwise have, a reduction in the monetary and psychic cost of birth control due to a family planning program is expected to reduce fertility and yield private gains and possible social externalities (Rosenzweig & Schultz, 1985). One approach to measuring variation in fertility that

is attributable to policy initiatives that subsidize birth control is to examine information couples provide on surveys on their reproductive goals and behavior. This research strategy might distinguish between the behavioral demand for births and their biological supply through an analysis of what couples say they want and what they have. Pritchett (1994) provocatively explores survey responses regarding desired fertility (DF) and the number of wanted births (WB) to evaluate how responsive actual fertility is to population policies, such as family planning effort (FPE) as measured by Lapham and Mauldin (1984), where these policies are designed to reduce the cost of knowledge of birth control methods, their adoption, and their continued use. He shows that most of the variation across countries in their total fertility rate (TFR) can be accounted for by the DF ($R^2 = 0.91, n = 66$) or WB derived from some of the same representative surveys ($R^2 = 0.85, n = 47$). Conversely, unwanted fertility, or the fraction of births, which at the time of conception are reported as unwanted, does not explain much of the cross-country variation in TFR. About the same proportion of births are unwanted in high-income countries where contraceptive prevalence is high and family planning services are widely available, as in low-income countries, where in some cases these services are limited and their absence is expected to lead to more unwanted conceptions. These stylized facts lead Pritchett to conclude that family planning services and policies do not account for much of the variation in fertility, and if FPE were to increase in low-income countries from its indexed value of zero to the average level in his sample of countries (i.e., 31.4), holding desired fertility fixed, this major policy revolution would reduce TFR by only 0.22 to 0.37 births according to his model regression estimates.

Pritchett dismisses the idea that DF or WB could themselves be affected by family planning services, because the survey questions are phrased to focus on an “ideal” family size, as if there were no costs associated with future birth control. However, if some women viewed abortion as a likely means to control their unwanted fertility, DF and WB might be higher than in an environment where effective contraception and sterilization were widely available or understood as a reliable and safe remedy to avoid excess or ill timed childbearing. In addition, DF and WB are also undoubtedly measured with error in a survey as responses to a hypothetical situation. Therefore, WB and DF should be treated as endogenous choice variables that are potentially measured with error in a model of fertility determination. To estimate their effects on fertility without bias some exclusion restriction is required, or an instrument must be specified that is correlated with WB or DF, but is not correlated with the residual error in the equation predicting TFR. What might this instrument be?

Culture, customs, or religion could lead some women to report wanting more children than they would actually demand because it fit their norms or morals, but these variables would also likely impact actual fertility or TFR. Education of women might strengthen the willingness of women to report wanting fewer children than they had,

or admitting more of their conceptions were “unwanted.” But such a variable that is likely to influence DF or WB is expected to explain some of the error in the TFR equation as well. In other words, it is difficult to imagine an identifying variable that affects DF or WB, but does not also perturb TFR, because all three variables are closely interrelated at a subjective level of the couple and may be affected by complex economic constraints and psychological traits. Pritchett’s additional working assumption that DF and WB are unaffected by FPE could also be challenged by specifying an independent variable shocking FPE that does not affect DF and WB. When instruments of this form cannot be proposed to identify the causal channels between constraints and behavioral variables, social scientists might forego estimating structural relationships and estimate instead a reduced-form equation for fertility. How large an effect does the independent shock to FPE actually exert on TFR?

This is the objective of Schultz (1997) based on a cross section of low-income countries in the 1970s and 1980s in the period when FPE scores are periodically reported by the Population Council (Lapham & Mauldin, 1984). Family planning program subsidies by the state or nonprofit organizations tend to increase, however, when a substantial proportion of local women demand these program services to control their “excess” fertility, or where $(TFR-DF) > 0$ (Rosenzweig & Schultz, 1985). One way to begin to deal with the endogeneity of FPE in a country is to construct data in the form of a panel of cross sections, and regress differences over time in TFR on differences over time in FPE within countries and other conditions affecting reproductive demands over time, which sweeps out the effects of unobserved time-invariant characteristics of a country that might be associated with excess fertility and result in response in terms of population policy. This first-differenced specification of the fertility model also reduces the problem of measuring fertility in a cross section of age groups of women, summarized by TFR, rather than by measuring the more appropriate lifetime fertility in the form of children ever born for a birth cohort of women at the end of their childbearing years. When this first-differenced model is estimated from a panel including observations for 1972, 1982, and 1988, the estimated effect of change in FPE on change in TFR is not statistically significant (Schultz, 1994, 1997, Tables 8 and 9).

Alternatively, if FPE is treated as an endogenous policy variable, as it should be, the supply of international donor funding for domestic family planning in each low-income country (planned parenthood federation expenditures per woman of childbearing age) can be treated as an exogenous subsidy for FPE in the first-stage estimates, and the second-stage least squares estimate of this source of variation in FPE on TFR is also insignificant, though the donor subsidy is significantly related to FPE in the first-stage equation. In 1 year, 1988, the local price of contraceptive pills is available for 58 low-income countries as one objective indicator of FPE, and the estimated reduced-form price elasticity of TFR is 0.3, but it is not statistically

significant ($t = 1.51$). However, when the child mortality rate is also included as an endogenous determinant of TFR, identified by the available *per capita* supply of food calories within a country, the contraceptive price elasticity of fertility decreases to 0.05, but is now measured somewhat more precisely ($t = 1.70$). These estimates (Schultz, 1997, Table 7) are consistent with Pritchett's conclusion that the effects of FPE on fertility are probably small across all available countries. But using DF or WB, as if they were exogenous parent demand-driven variables in a cross-country regression does not recover the causal relationship between FPE and TFR as claimed by Pritchett. Nor does it provide a satisfactory basis for inferring the consequences of independent variation in population policies on fertility in an average country. Using outside donor assistance for family planning programs or the prices of contraception suggests that the impact of FPE on TFR is modest, if the policy variable, FPE, is treated appropriately as endogenous and identified by the international population assistance. Unfortunately, cross-country regressions are a blunt tool for deriving policy insights in this case, except under exceptional circumstances. More promising are analyses at the level of the family or community based on the increasing number of public use samples from censuses and surveys.

8.3 Estimating the cross-effects of fertility variation on family lifetime outcomes

At the individual level the two instruments used most commonly to represent exogenous fertility variation are twins and the sex composition of initial births (Schultz, 2008a). Twins are reasonably interpreted as an exogenous shock to fertility that may be more or less uncorrelated with many sources of economic and preference variations affecting other family outcomes. The main limitation of interpreting twins as simply one "unwanted" birth is that twins impose on their parents two additional burdens. First, there is no spacing of the two births which is not likely to be optimal for most couples and may thus impact negatively other family outcomes. Second, and twins tend to be of lower birth weight, health and cognitive performance compared to singleton births, or in other words, twins are on average lower quality births, at least initially. Subject to these limitations on the validity of twins as an instrument for "undemanded" singleton births, one finds that OLS estimates of the partial association of fertility on various family outcomes, for example, women's labor supply, her health, or the human capital of her other children, etc., tends to be an absolutely larger magnitude than the unbiased IV estimate of fertility's effect on the same family outcomes, where the IV "treatment" is a twin on first birth, or the ratio of twins per completed pregnancy. If the IV estimate is smaller in absolute magnitude than the OLS estimate, it suggests that a preference for quality tends to be inversely correlated with fertility, and omitting this family control variable from the OLS estimate of fertility's effect on family outcomes contributes to an omitted variable upward bias (in absolute magnitude) (Schultz,

2008a). Part of the difference between the IV and OLS estimate of the effect of fertility on family outcomes could also be due to the intrinsic differences between twins and one extra singleton birth.

As described earlier in Section 4.3, Rosenzweig and Zhang (2009) analyze twins and singleton births surveyed in China to estimate the effects of exogenous variation in family size due to a twin on their child quality, measured by various indicators of schooling and health. According to their model, they first adjust for the lower initial endowments of twins, given the estimated propensity of their Chinese parents to reinforce the initial endowments of their children. They then estimate the quantity-quality tradeoff by contrasting the twin instrumented family size effect, depending on whether the twin occurred on the first or second birth. The sacrificed child quality is larger, as suggested by their model, when the unwanted fertility arises because of a twin on the second birth providing the parents with fewer opportunities in the future to compensate for the additional child.

A less satisfactory instrumental variable for fertility is the sex composition of initial births as proposed first in the study of third- and higher-order births in the United States (Angrist & Evans, 1998). They report that couples with a boy and girl are significantly less likely to have a third birth, and using this information on sex composition of the first two births of parents in the United States as their IV for later fertility, they obtain IV estimated effects of fertility on family outcomes that are smaller in absolute magnitude than the OLS estimates of the partial association. Although this instrument is determined for the most part randomly, the fertility response to the IV varies by the strength of parent preferences for having both a boy and a girl. In some low-income countries parent preferences for the sex of their offspring may differ notably across cultures and could change over time with development. In some countries where daughters are provided a dowry to marry and sons are paid a dowry, family lifetime wealth is directly affected by the sex composition of births (Rose, 2000). The sex composition of births used as an instrument for exogenous variation in fertility is then no longer uncorrelated with other family lifetime constraints or preferences. These IV estimates based on sex composition of initial births are, therefore, invalid estimates of the effect of exogenous variation in fertility as would arise from the implementation of most population policies, such as sex education, family planning, or reproductive health programs. Additional studies using different instruments for fertility and its consequences on family outcomes are reviewed elsewhere (Schultz, 2008a).

Recent technological developments have also begun to undermine the validity of both twins and sex composition of initial births as instruments for exogenous variation in fertility or population policies, although analyses of birth and family histories from earlier periods are not affected. Drug therapies to assist women in controlling ovulation and conception have had the side effect of increasing the incidence of multiple births in such countries as the United States. Therefore, the occurrence of twins has become correlated

with parent demands for births and couple fecundity in complex ways, and may also be associated with family wealth since infertility treatments maybe relatively costly and not reimbursed under health insurance. Thus, twins born after about 1980 in high-income countries are a less satisfactory basis for estimating the effects of fertility.

Techniques to test for the sex of the fetus at an early stage in a pregnancy by means of ultrasound, amniocentesis, or chorionic villus sampling have allowed a growing share of couples who have sufficiently strong preferences for the gender of their child to abort a fetus of the unwanted sex. As this practice occurs more widely, the sex composition of children becomes correlated with the couples' reproductive preferences and other family choice outcomes, and the sex of initial births ceases to be a valid instrument for estimating the effects of exogenous variation in fertility. The increasing ratio of males to female births at each successively higher parity in countries such as China, Korea, and portions of India reflects this pattern (Schultz, 1997, 2008a).

Another way to distinguish exogenous variation in fertility is to recover a measure of biological heterogeneity in terms of couple fecundity, by assuming more structure for the reproductive process, and then predicting the choice of contraceptive practice during each reproductive cycle. The couple's latent fecundity is inferred by comparing the couples conception rate and the conception rate expected based on their predicted contraceptive practice and consistent estimates of the reproduction function (Rosenzweig & Schultz, 1987). This "residual" variation in fertility, which includes biological fecundity, is associated with having borne more (or less) children, leads to adopting more effective contraceptive methods presumably to compensate for higher fecundity, having lower birth weight children, and providing them with less schooling. These structurally estimated effects of exogenous residual fertility on child quality outcomes are again smaller in absolute value than if they are estimated from the partial association (OLS) effect of fertility on child quality.

8.4 A social experiment in family planning and reproductive health: Bangladesh

One family planning, maternal and child health program was designed as a social experiment in a remote rural area of Bangladesh, in the Matlab Thana. It was initiated in half of 141 villages for which there was already in place a reliable demographic surveillance system of the population, registering all births, deaths, marriages and population movements. The family planning program outreach effort was started in October 1977, which contacted in their homes all married women of childbearing age every 2 weeks, offering them various methods of birth control. The populations were periodically censussed and then randomly sampled in a comprehensive socioeconomic survey in 1996. A census in 1974 confirms that the program treatment and comparison villages did not differ significantly 3 years before the program started in terms of their surviving fertility, approximated by the village ratio of children age 0-4 to women age 15-49. A difference-in-difference

change between the program and comparison villages preprogram and postprogram indicates that by 1982 surviving fertility is 17% lower in the program areas, and remained 16% lower in the 1996 survey after the program was in operation for nearly two decades. Fertility is lower in the program areas only for women less than age 55, presumably because women over 55 were over 37 in 1977 when the program started, and these older women had essentially completed their childbearing at that time and hence their fertility did not respond to the program treatment (Joshi & Schultz, 2007).

With this social experiment, it is possible to estimate that women age 25–55 in 1996 had about one child less in the program villages compared with the comparison villages. These women in program villages were healthier measured by their BMI being 1.0–1.5 units higher than in the comparison villages, and their children experienced a death rate by their fifth birthday which was 25% lower in the program villages. The girls age 9–14 and 15–29 had obtained about one-third of a standard deviation more years of schooling for their age and sex in the program areas, whereas the boys had obtained about half a standard deviation more schooling. The estimated program effect on the boy's schooling was statistically significant at the 5% level, whereas this schooling effect was not significant for girls. On the other hand, girls age 1–14 were reported to have a significantly higher BMI in the program villages, normalized for age, whereas there was no significant difference in BMI for boys (Joshi & Schultz, 2007). Women age 25–54 in 1996 report log monthly earnings a third higher in the program villages compared to the other villages, and the households in which women reside have proportionately more financial, agricultural, nonagricultural, and housing assets, more consumer durables and jewelry and household tube wells in the program villages (Schultz, 2008b). The prediction of Malthusian diminishing returns to labor is not supported by the Matlab social experiment, in the sense that the wages of young men and women, age 15–24, are not higher in the program villages despite the tendency for there to be fewer children in these villages and they were more likely to attend school. Finally, the wage rates for adult males age 25–54 are no higher in program areas than in the higher fertility comparison villages, whereas in contrast the wages of adult women were at least one-third higher. Moreover these wage patterns are not affected when the selection of who participates among youth and adults into the wage labor force is jointly estimated with the wage function, allowing for the heterogeneity of workers. These findings suggest that women in the program villages who are likely to have had fewer unwanted births due to the efforts of the outreach program have also enjoyed disproportionate gains in their health, market earning opportunities, and household assets compared with comparable women in the comparison villages.

This single case study of the long-run consequences of policy-induced voluntary reductions in fertility suggest that in this poor rural South Asian region, a concerted outreach program achieved a significant decline in fertility and sustained lower levels of fertility for two decades, during which fertility has declined substantially in both

groups of villages. This policy-induced reduction in fertility is associated with women and children being in better health, sons receiving more schooling, and women earning proportionately more in the paid labor market, and living in households with proportionately greater assets. Whether these estimates of population policy effects would exist in other parts of the world remain an unexplored issue, because social experiments or even matching evaluation studies of family planning, reproductive health, and child health programs are rare. Multivariate regression studies of individual and community data suggest that where family planning programs have been expanded more rapidly, birth rates, especially among older women, have declined more rapidly. Evidence from such countries as Taiwan, Thailand, Indonesia, and Colombia are reviewed elsewhere (Frankenberg, Sikoki, Suriastini, & Thomas, 2003; Schultz, 2008a). There are, nonetheless, indications that the effectiveness of national programs to reduce fertility diminish as they expand to serve all parts of a country, and birth control technology becomes widely understood and birth control supplies are competitively available in the private and public sectors.

9. INTERNAL MIGRATION, DEMOGRAPHIC TRANSITION, AND EVALUATION OF POLICY

Internal migration is a major force redistributing the population during development as sectoral composition of the economy and the geographic distribution of employment change (Kuznets, 1966). This has a variety of implications for the evaluation of population policies. Assessing the impact of programs and policies are often based on inter-regional differences in program availability, generosity, or eligibility (in fact or in law) which may depend on residence. Implementing policy evaluations therefore frequently requires the joint analysis of migration and outcomes targeted by the program (Rosenzweig & Wolpin, 1986; Schultz, 1988b).

The regional allocation of programs, the intensity of those programs, and their effectiveness in achieving their objectives are likely to be nonrandom, due to many features of policy planning and political economy. Thus, if local program treatments are not administered as a social experiment such that they are allocated independently of the population's endowments and conditions, direct inferences from cross-sectional data of the effect of the program treatment on behavior and outcomes of the treated may mismeasure the local area treatment effect (LATE) on the average member of the population (Imbens & Angrist, 1994; Rosenzweig & Wolpin, 1986). Programs may be allocated to population groups which are thought to be most in need of the program's services, and therefore failure to control for these preprogram conditions could lead to an underestimate of the program's impact based only on postprogram variation across regions. Alternatively, the allocation of the program may favor those groups who are most likely to change their behavior due to the program, in order to

maximize the program's demand for services. This scheme of program allocation would tend to lead to an overestimate of the program's effect on the average population. When controls for preprogram conditions and behavior are included in the analysis of program impacts, or the analysis focuses on difference-in-differences between the changes in behavior between the program and control areas over time, these nonexperimental program evaluations are likely to be more reliable.

Interregional migration changes the composition of regional populations, adding another form of potential bias to conventional cross-sectional estimates of the response of a representative individual to regional variation in program treatment. There are many reasons in different environments for the regional access to population programs to differ. Most social welfare programs are better funded and more accessible in urban than in rural areas, and one reason for migration from rural to urban areas is to improve individual and family access to such social services, such as health, schooling, and family planning, and in some settings personal security and protection of private property. If populations are heterogeneous in their preferences for these population and social service programs, as well as other features of the local economy and its amenities, it is expected that those individuals who decide to migrate from rural to urban areas will on average attach greater value to these improved urban social services than would the average of potential migrants. This positive correlation between the probability of migration and the preferences for urban (destination) services will lead to a disproportionate number of rural-urban migrants in urban areas who demand the services provided in their destination areas. This correlation can bias up the regionally estimated "effect" of the program treatment on the demand for, and use of, program services. The problem is likely to be more serious where the average fraction of migrants in the population is larger, as in a Latin America after the 1970s, or where lifetime mobility is greater, or mobility is not impeded across regions by languages and ethnically diverse cultures. One diagnostic check on evaluation assessments is to estimate program effect on outcomes for the sample of only natives (i.e., nonmigrants living in their region of birth), and separately for in-migrants. The native estimates that exclude the migrants provides under certain assumptions a lower bound on the local average treatment effect (Duflo et al., 2008; Imbens & Angrist, 1994). Even when the design of regional variation in program treatments is randomized, comparisons over time for a panel sample may still be subject to selective migration, and also involve panel attrition in part due to migration and nonresponse (Fitzgerald, Gottschalk, & Moffitt, 1998).

Regional differences in the relative price of different types of food may be persistent, due to variation in local climates and agricultural potential, which engenders regionally heterogeneous patterns of food consumption. In this case, these patterns of taste for locally plentiful and lower cost foods would deter certain paths of interregional migration. Without recognizing these origin-determined tastes the apparent gains in

migrant real income would be overstated, at least until migrants adopted the tastes for food favored at their destination. Atkin (2009) estimates the magnitude of this regional taste effect on Indian migrant consumption and welfare.

Because education of individuals is widely observed to be positively correlated with the probability of migration between regions, or the likelihood of having moved from one's birthplace (Schultz, 1982, 2003; Schwartz, 1976), a family's investment in the schooling of children increases the likelihood that the child will migrate, other things being equal. A dynamic structural model is required to disentangle the consequences of education and migration. It is difficult to assess how migrant and native differences arise, though they tend to be notable in terms of adult health, labor market productivity, income, fertility, and even investments in the health and schooling of the children of immigrants and natives at destination.

The composition of internal migration in different parts of the world differs in terms of the sex composition. In North America men were initially the first to migrate toward the Western frontier territories and women followed. In the twentieth century with the frontier closed in the United States, women were more likely to migrate from the rural to urban areas and find employment outside of agriculture. Similar gender patterns in migration are later evident in Latin America, with the exception of indigenous populations where women's migration is restricted by their lack of education, limited knowledge of Spanish, and possibly other cultural constraints. For example, in Colombia and Brazil men dominated in the stream of migration toward rural frontiers, such as the Amazon basin, whereas the probability for women to migrate from the agricultural sector to towns and cities was larger than for men.³⁷ The reversal in the gender ratio of migration from rural to urban areas between the Americas and South Asia and Africa may be explained in large part by the greater levels of schooling of women relative to men in the Americas compared with South Asia, Africa (except for Southern Africa), the Middle East, and West Asia. The culture and family systems that provided girls and boys with similar levels of education in the Americas also facilitated the movement of women to the cities where they could earn more income and provide their children access to better educational opportunities than in the rural areas (Schultz, 1988b). Evaluation of the effects of social and population programs, especially those targeted to the needs of women and children, require a model of migration which can simultaneously account for the contribution of the programs to female and male productivity, fertility and family health, as well as the migration.

In many South Asian settings where women's education and mobility are more limited than men's, women marry at a relatively young age, and they can migrate to urban areas only in the company of their husbands. The effect of migration or its limitations on women's economic mobility has profound effects on gender differences in economic and demographic behavior and their persistence in different regions. In Latin

America migration of women and men from rural areas to towns and on to cities occurs when their schooling is completed, and marriage often occurred after migration, if at all, with many women establishing themselves first in the urban labor force, working in some instances as a domestic servant. Although it is not well understood how policies influence these sources of gender inequality and migration, a few examples are cited in the literature. Bangladesh implemented a secondary school scholarship program for girls, which appears to have contributed to the closure of the initially large gender gap in primary and secondary schooling (Arends-Kuenning & Amin, 2004). This promotion of educational opportunities for girls, especially in rural areas of Bangladesh, has been more recently reinforced by the expansion of urban employment of young women in industries exporting textiles and apparel. Basic education of women prepared them for migrating from their rural villages to work in urban factories, living in dormitories, and remitting much of their earnings to their parents until they marry, at a much later age than did their mothers.

In South East and East Asia, women also migrate to the cities to work in export-oriented industries, but the gender gap in education was smaller than in most South Asian countries and the overall level of schooling was higher. In Taiwan, for example, older married women in rural areas found work in rural-based industries. They thus contributed to their family's income, without having to migrate and incur the higher costs of living in urban areas (Brinton, Lee, & Parish, 1995). Circular migration in other areas of SE Asia, such as Malaysia, allowed men in rural areas to migrate to find temporary urban employment, while returning to help with family farm work during the peak agricultural seasons of planting and harvesting. Their families thus avoid the high urban costs of housing and moving their entire family to coreside with their urban jobs. Many development policies, restrictions on international trade, and factor subsidies combine to create the incentives for different migration patterns which affect the reallocation of the labor of family members from rural household production activities to urban jobs and off farm employments. In the Middle Eastern region women have made progress in obtaining more schooling, reducing their child mortality and fertility, but not finding employment in the formal labor market. The public sector in some settings provides women with more favorable employment opportunities (Panizza, 2003). Yet as globalization has increased privatization and reduced the restrictions on international trade, growth of the public sector has slowed. Reciprocal expansion of private exporting firms has provided in some countries more jobs for women, as is documented in Turkey (Ozler, 2000; Wood, 1995). But even in the period since 2000 when employment has grown rapidly in the Middle East and North Africa, unemployment rates have remained relatively high for women compared with that for men, and this "underutilized" supply of women's labor is greater for better educated women (Nabli, Silva Fauregui, & De Silva, 2007).

10. CONCLUSIONS AND POLICY DIRECTIONS

Population policies seek to reduce mortality, morbidity, to produce a longer disability free life, to reduce the inequality in health, and to provide couples with cost-effective techniques for birth control. Internal migration within countries is viewed here as another form of individual behavior which is coordinated with fertility, health status, and the balance of geographically distributed resources and population. Migration complicates the task of evaluating the effects of population policies on family behavior and lifetime outcomes, because migration is another form of human capital investment that responds to similar observed and unobserved determinants of health and control of fertility, and probably complements both.

Many lines of research indicate that environmental conditions existing at the time and location of the conception and birth of a child are related to the child's length of gestation, height and weight at birth, and these initial endowments of the infant help to explain early child health, and remain significantly associated with the child's health status as an adult, adult height, IQ, achievements in school, labor market productivity, reduced late age mortality, and delayed onset of chronic health limitations, which permit labor force participation at later ages, and improve well-being. These often delayed consequences of the latent health capability arising from ample fetal development and birth weight for gestational age (i.e., fetal growth rate) suggest the difficulties researchers face if they are to understand precisely how health human capital is produced efficiently and is accumulated equitably over a generation.

Then the question arises how health is potentially transmitted through initial health endowments, both genetic and reproducible, and the direct transfer of economic resources and care to the next generation? Understanding how parents respond to the initial endowments of their children and health shocks they experience over their life is only beginning to be explored by comparisons of siblings and twins. Are parents compensating and thus reducing the final inequality in health or productive outcomes among their offspring, or do they reinforce these initial allocations? What are appropriate fiscal and institutional roles of the state to encourage efficiently these early forms of investment in child health and development? How does this health human capital generate productive and consumption returns to the individual, and where are the most valuable spillovers which improve the lives of other family members and society more generally? In sum, how does policy-induced variation in health human capital and fertility control enhance total factor productivity? This opens an extensive field for research where the existing stock of survey and census data can support some explorations, while data collection efforts and social experiments may be designed to shed light on unresolved questions.

Cross-country macroeconomic comparisons do not seem well designed to add to our understanding of the specific pathways through which technological change, medical inputs, consumption patterns, and health-related behaviors might be channeled

by health insurance, or suitably subsidized by governments and NGOs to improve health outcomes, and enhance in a cost-effective manner the productive potential and well-being of populations. Such changes in economic potential related to health improvements parallel closely economic development, but because of the long lags between fetal growth and child development and adult productivity, the connections are only beginning to be persuasively documented, and then only at the microeconomic level of the individual and intergenerational family. One potential reason to pursue aggregate comparisons is that they could in the future be better focused to shed light on the productive spillovers of health status and externalities of fertility declines that are not captured in individual output and wages. These externalities might otherwise be overlooked in conventional microeconomic studies of households, yet could be captured in aggregate data for communities, health administrative areas, service areas for clinics and hospitals, or subnational states. Some preventive health measures and efforts to control infectious diseases may constitute an important public good with significant social externalities, but empirical measurements of these externalities are rare for either health or fertility. Policy-induced declines in fertility attributable to subsidies for birth control and reproductive health services may also be linked at the individual level to improvements in women's lifetime productivity and intergenerational gains in the health, nutrition, schooling and migration of their children. Convincing empirical studies estimating the magnitude of these spillovers of population policies on health and fertility would be valuable for setting priorities among programs in different localities, and assessing the effectiveness of alternative program designs.

A tentative review of the empirical evidence suggests that higher priority should be assigned to prenatal and reproductive health care for women, and to programs that prevent childhood and adult infectious diseases and control their spread. Some of these health programs can be relatively effective for their cost and yet are underutilized by the uneducated, poor, rural segments of the world's population, where related health problems are most severe. Childhood immunizations and those provided to pregnant women, such as tetanus toxoid, are often neglected. For example, in rural India the prevalence of childhood immunizations has been roughly stable for the last 7 years, a period when incomes have grown rapidly and public and private expenditures on curative health care have increased markedly.³⁸ BMIs of children age 0-3 are low in rural India and have not improved as rapidly for girls as they have for boys, or in rural areas as in urban areas (Tarozzi & Mahajan, 2007). As noted earlier, according to the 2005/2006 National Family Health Survey (NFHS), adult female heights in rural India, an indicator of early childhood health and nutrition, have advanced at a third the rate as those of adult males born from the 1960s to the 1980s (Deaton, 2008). These indicators of initial health conditions in India do not indicate a recent convergence in basic health outcomes. Modest progress is observed in a region such as Latin America, which is otherwise known for its entrenched economic inequality (Schultz, 2005).

The convergence over the last 50 years in life expectation at birth across countries has been interpreted as welfare equalizing (Becker et al., 2005), but the trend has slowed if not reversed since 1990. This is largely due to the AIDS epidemic raising adult mortality in many low-income countries. Is this a special exception to the prevailing pattern of technological diffusion of medical knowledge? Will increasing access to anti retroviral drug therapy reestablish convergence in life expectancy? The cost of curative care to treat noninfectious and degenerative diseases is increasing as the epidemiological transition progresses from high- to low-income countries. Even within high-income countries it is notable that the variance in life expectancy after age ten, by which time most infectious childhood illnesses have been confronted, has not decreased since 1960 in the OECD, indicating this measure of adult health inequality has not diminished within or between countries (Edwards & Tuljapurkar, 2005). The failure to reduce this measure of the variance in adult mortality may signal that further improvements in health and reduced mortality requires a larger redistribution of income to extend health gains to all socioeconomic groups. Providing low-income countries with cost-effective *curative* care for noninfectious and degenerative diseases may thus be a long-run fiscal and administrative challenge. This fiscal challenge could divert attention from delivering the core preventive child and maternal health measures which may be more cost effective, though apparently not adequately understood or used by a substantial share of the world's poorest populations.

Prevention involves more than the provision of public health services at the community level. It also involves the education of women who manage the household's production of health and nutrition. Women also appear to assign a higher priority to the medical care and feeding of children than do men, given the same household budget constraint (Schultz, 2001). Raising the level of women's schooling may be an efficient means to increase the fetal growth and development of their children, reduce family deaths, and mitigate childhood infections. Are there better ways to improve hygiene, properly use of water, sanitation, and select preventive and curative medical care, and monitor the growth and health status of children (Christiaensen & Alderman, 2004)? The potential links from a mother's schooling to her health and that of her family are not thoroughly assessed and used to coordinate public health strategies. This may be because the schooling of females does not fall strictly in the bailiwick of health ministries, medical systems, programs of public health, water and sanitation engineering, or the disease-specific interventions around which international agencies, such as WHO or the World Bank, organize their programs and marshal their resources.³⁹

The production of health in the household and the diffusion of knowledge to decision makers in the household that could improve the effectiveness of preventive and curative health care are rarely studied. This may be because household behavior relevant to producing health is not integrated into the planning of public medical systems.

National plans tend to prescribe the allocation of public expenditures across tiers of institutions in the public sector, from hospitals to clinics, down to local "health posts." But these medical plans often neglect the compensating behavior of the private medical care sector, and the role of privately determined demands of households for health-related inputs, and for modification in household behavior, including unhealthy behavior such as smoking, alcohol, and substance abuse.

There are many promising avenues for research on the connections between health and development, and fertility and development that would improve our understanding of economic and demographic behavior and the design of policy institutions. This research is conducted through analyses of cross-sectional and panel surveys and censuses for individuals, households and communities, on a country by country basis. These investigations increasingly include the collection of biological markers and health information and migration histories, to better define the causal pathways from the policy instruments to fetal growth to the lifetime adult outcomes of interest to economists and health scientists. Pragmatic and replicable policy and program interventions in the public and private health sectors may thus be documented in the regions where individuals are born, matured, and currently reside. As new programs are implemented experimentally, or at least in a well-documented scaling up of pilot programs by service areas, it should become possible to evaluate the consequences over time of such programs. These assessments will be difficult to perform. Only when the records from the underlying demographic surveillance systems are open to public evaluation by independent researchers, will confidence build as to the strengths and limitations of the data and the validity of evaluation studies they support.

Where the levels of health are lowest in the world, infectious diseases are still a burdensome problem, and reproductive and child health interventions may be most cost effective. Yet these are also the areas where today panel survey data focused on health and economics development are least adequate to the task, and what information is collected is not likely to be open to the research community, outside of the governmental agency who have a weakly defined interest in evaluating the success of their own health programs. In addition, the evaluation and design of health programs requires merging accurate administrative data on program operations, variation in performance incentives, and independent monitoring of local program staff. Finally, a political commitment is required to engage in such organizational variations in the delivery of public health and educational services, the long-run-independent collection of high-quality household sample surveys, and the provision of these merged administrative records and survey data to independent researchers to undertake the difficult task of program evaluation. These conditions are not common, but may be increasing. Evaluation of social programs within development assistance agencies appears to be receiving a higher priority today, and health and population programs would be a good place to start.

End Notes

- * The comments of various audiences at several institutions are appreciated, especially those of Germano Mwabu, Mark Rosenzweig, David Sahn, and John Strauss. The research has been supported by the MacArthur Foundation especially that summarized in Section 8.4.
- 1. Ben-Porath (1967) is the first to link formally the increase in expected life span to increases in human capital investments and growth in income, but he assumes that a fixed lifetime is allocated between human capital investment and market earnings, ignoring changes in leisure or nonmarket production or labor supply, and the endogeneity of life span or health. Implicit in Ben-Porath's framework is the idea that with a longer lifetime to collect the returns on human capital, individuals are encouraged to allocate longer periods to (schooling) human capital investment at the start of the life cycle. Hazen (2007) shows that lifetime hours in the market labor force for male workers in the United States has decreased since the cohort born in 1840, measured in synthetic period rates or more appropriately in cohort rates, and he concludes that this reduced lifetime labor supply when survival and human capital are increasing is inconsistent with his revised Ben-Porath model. But he neglects to account for nonmarket production or full income, and the offsetting increase in female market labor supply over time, and institutional changes in pensions, which impose additional implicit taxes on the earnings of workers after they become eligible for social security pensions, and may be responsible for the observed reduction in lifetime labor supply (Gruber & Wise, 1995, 2004).
- 2. Preston subsequently excludes the seven Soviet-bloc countries in the 1960s from his cross-country analysis, perhaps because Soviet national income estimates are based on different accounting practices, and international exchange rates into dollars for the Soviet-bloc currencies are not determined in a "free market."
- 3. Preston, however, prefers a more flexible exponential functional form, because the simpler semi-log fitted function systematically overestimated life expectancy at low-income levels, a sign of misspecification. This exponential fitted model is plotted in Figure 2, and reveals that in each of the 3 years, countries with higher *per capita* income tended to have lower mortality and a longer life span. Preston (1980) subsequently extended his analysis of the association of life expectancy and income using data for 1940 and 1970 by adding adult literacy and *per capita* calorie supply as determinants of life expectancy. Literacy is significant positively associated with life expectancy suggesting a 10% point increase in literacy is related to a 2-year gain in life expectancy, whereas a 10% increase in *per capita* income is related to about a half year increase in life expectancy. The coefficient on income is reduced by 40% when literacy is included, whereas the calories variable is not significantly related to life expectancy. Others have found child mortality in the first 5 years of life is significantly related to FAO calorie supplies across countries and over time within countries, but in a nonlinear functional form such as the quadratic form, suggesting that the effect of calories on child mortality is subject to diminishing returns (Schultz, 1997). The strongest covariate with child mortality in Schultz's study is the years of *female* education. Preston does not distinguish between literacy of men and women separately, although the empirical tendency in numerous studies at various levels of development and aggregation is that female education is associated with a large reduction in mortality, whereas male education is generally not statistically significant once female education is controlled. This pattern is especially pronounced when the analysis focuses on infant or child mortality, although it is evident in studies of crude death rates and life expectancy as well.
- 4. The fitted curves are locally weighted smoothed scatter plots, using a tricube weighting function with a bandwidth of 0.35 years for 1960 and 2004, and 0.70 for 1930. Following Preston, Strauss and Thomas rely on foreign exchange rates to convert incomes into 1995 dollars. If incomes are expressed in units of PPP, the PPP incomes of the poorest countries would increase or be shifted to the right in Figure 3, and the slope of the health-income curve would be increased across countries.

5. HIV/AIDS may reduce life expectancy by 10-20 years in a decade as in Botswana, but may affect the productivity of infected individuals only in the last year of life with the emergence of full blown AIDS. In the past, those dying between ages 15 and 50 were more likely to have chronic illnesses for many years and consequently experience diminished productivity for much of their lifetime. HIV/AIDS has altered the relationship between life expectancy and productivity.
6. The increase in the wage due to improved health of the worker may also affect the market labor supply of the worker, increasing hours worked if the substitution effect (positive) of the wage on labor supply outweighs the respective income effect (negative), or decreasing hours worked if the income effect exceeds the substitution effect (Schultz & Tansel, 1997).
7. Age of mother affects slightly the probability of having a twin, and could influence the timing costs of a twin. Rosenzweig and Zhang (2009) therefore control for age of mother at birth, which is also a choice variable of the family that could be associated with other child investments and family behavior. Conditioning on maternal age at birth purges the first-stage prediction of the probability of having a twin from its possible dependence on age. Age of mother is, nonetheless, endogenous to child investments, such as birth weight and length, and cannot be treated as independent of the error in the later child, mother, or family outcome variables.
8. Because the children in the Kunming China sample are still of school age, it is not possible to follow them into the labor force and estimate wage functions in order to evaluate the exogenous family size effects on adult productivity.
9. With such imprecise estimates of adult mortality and cause-specific deaths, there is much uncertainty, as recently reflected in the downward adjustments in global estimates of mortality due to HIV/AIDS and more recently malaria.
10. To account for the effects of health on economic growth, Weil (2007) analyzes the survival rate for adults from age 15 to 60, which is a reasonable refinement compared to crude death rates or life expectancy at birth, because the goal is to describe the productive health status of the population in the prime working ages, not early childhood or the elderly. (Lorentzen, McMillan, and Wacziarg 2008) also treat both infant mortality and adult mortality (age 15-60) as determinants of growth through their effect on investment, schooling, and fertility. But this step toward a more reasonable measure of adult mortality does not improve the empirical foundations for his adult health indicator. It remains essentially a residual imputation of deaths to adults which are not attributed to children dying, and may be prone to large and possibly systematic errors in low-income countries, an error which is difficult to assess without better death registration systems.
11. IIASA (Lutz & Goujon, 2001) has developed and projected estimates of the educational attainment distribution of the populations of countries by age and sex, using cohort methods that rely on additional data sources than the UNESCO published tabulations adapted by Barro and Lee (2000).
12. To assess this lagged effect of childhood health on economic productivity and growth, one might have asked researchers such as Acemoglu and Johnson (2006) to lag their increases in child survival 50 years and thereby to account for national income growth. The lack of sufficiently long time series for mortality and subsequent income in either low or high-income countries appears to limit this attractiveness of this line of inquiry.
13. BMI is typically measured by dividing the individual's weight in kilograms by her height in meters squared, and it is thus weight approximately normalized for the individual's height. Variation in both height and BMI is typically expressed as Z scores to be in comparable units of standard deviations. The high intercorrelation between height and weight would otherwise make it difficult to estimate jointly outcomes as a function of both dimensions of stature.
14. Activities of daily living are used to access functional disability among the elderly in high-income countries, although their value as measures of health status among young or middle aged individuals is not clear, nor is their validity in the context of low-income countries or across countries (Schultz,

2005). Banerjee and Duflo (2007) show for a poor tribal area of India and Indonesia that ADLs of poor and relatively rich individuals only diverge after about age 55. Since they conclude that mortality and bad health are more common among the poor than the rich, their results suggest ADLs are not a discriminating measure of reproducible health differences until individuals are over the age of 60 in low-income countries. Other measures of health stocks are needed for persons under age 60, such as possibly height and BMI.

15. Case, Fertig, and Paxson (2005) report persisting associations between economic circumstances at birth and childhood illness, self assessed health, and socioeconomic status as an adult. They conclude from multiple regressions fit to the UK 1958 National Child Development Study that birth weight, chronic health conditions in childhood from age 7 to 16, and height at age 16 and maternal smoking are related to adult outcomes, even after controlling for socioeconomic characteristics of the origin family. They conclude that the “gradient” or gap between the health of children by parent socioeconomic status (SES) widens (or accumulates) as the child ages. In other words, socioeconomic inequality in health increases over the child’s life cycle. Currie and Stabile (2003) estimate similar tendencies for the SES gradient between the health of children to widen with age in Canada, while Currie and Hyson (1999) examine the impact of low birth weight on schooling and labor market outcomes, as the children experience repeated health shocks. For these estimates to be interpreted as causal effects, the initial features of the child’s health must be assumed to be exogenous as a right-hand side determinant of subsequent child outcomes. Even birth weight embodies the endogenous effects of parent behavioral choices and differences in environmental conditions (Rosenzweig & Schultz, 1983) introducing heterogeneity bias into causal interpretation of regressions of lifetime ordered health indicators. These issues are discussed later in this paper.
16. Cross-country study of child mortality found a significant nonlinear relationship between the availability of calories *per capita* and child mortality, with diminishing effects with increased calories (Schultz, 1997).
17. Little is known about the gene control of human height, weight, and BMI, but standard variance component analysis or correlations between relatives show that up to 60% of these anthropometric indicators of stature can be attributed to genetic factors (Ginsburg et al., 1998). A growing number of genetic dissections have been performed that suggest major genes are involved in genetic control of the traits and are reflected in widely dispersed population around the world (Liu et al., 2004). Assortative mating is operating on the traits as well and the joint distribution of the genes of mates will ultimately be part of the dynamic process determining height in the population.
18. The adolescent spurt in physical growth may be another period in the life cycle when the individual is especially vulnerable to nutritional deprivation and health care, and though the timing of this growth spurt can be delayed by malnutrition, it is unclear whether adolescent insults affect the height attained by the mature adult (Floud et al., 1990; Tanner, 1982).
19. Notable regional variation in average adult height is evident in the world, although many studies since the mid-nineteenth century suggest shorter peoples have increased their height in periods when their nutrition and health conditions improved (Tanner, 1982). Nonetheless, outliers, such as the greater height of many contemporary African populations and those who migrated from Africa, suggest genetic traits of these populations express themselves in greater height (Deaton, 2008). However, there is little evidence that these variations in height associated with ethnic identification, controlling for socioeconomic characteristics such as parent education and urbanization, are necessarily associated with differences in individual economic productivity (Schultz, 2002). In the Ghana LSMS, the tribal/language ethnic groups explain significant differences in height, and one might imagine they are genetically affected, but as instruments they do not predict variation in height that is positively related to wages (Schultz, 2002; see Table 3).

20. Linearly interpolating values between the quintiles that Su (2005) reports, I find individuals who have a BMI less than 19 exhibit a mortality rate that is 20% above the average in 1870, whereas this elevated risk of mortality (i.e., 1.2) is associated in 1972 with a BMI of less than about 21, an increase of about two units of BMI. The relative mortality risks in 1870 suggest a 20% higher than average mortality occurs for Union Veterans whose BMI is about 27 or more, whereas in 1972 individuals with BMI of 31 or more are observed to experience mortality at least 20% higher than average. This mortality threshold shifted to the right by about four units of BMI in this century. If these small samples of 800–600 individuals are sufficient to draw tentative conclusions, in this century the mortality minimizing level of BMI has increased, and of course the proportion of the US population at higher values of BMI has also increased, shifting the population distribution substantially to the right (see also Sunder, 2005; Sunder and Woitek, 2005).
21. If health policy priorities target reducing “obesity” or the frequency of BMI in the population above some critical threshold value, these values should be assessed with confidence for a variety of populations, and if possible over time. The distribution of diseases and health conditions causing death, public and private health institutions, and health-related behaviors may differ, occupations and life styles may have changed. For example, the decline in cardiovascular deaths may have been due to the use of drugs reducing blood pressure, which could have disproportionately extended the lives of those who have relatively high BMI in the 1980s in the United States.
22. They consider the intergenerational effect of mother's fetal growth on the twin's own child's birth weight, by estimating the intergenerational relationship for all of the twins who became mothers prior to the resurvey. The intergenerational correlation is significantly positive, indicating a persistence in birth weight health across generations as typically reported. But estimated within twin mothers, which thereby removes the effect of shared nutritional habits and preferences between mothers who were twins, the effect of mother's fetal growth becomes statistically insignificant as a determinant of the differences between the twin's children's birth weight. They conclude the intergenerational correlation of birth weight is due to the transmission of preferences, habits, and matching a mate in the marriage market with favorable endowments, and not due to the biological transmission of health endowment from one generation to the next in the form of birth weight.
23. Males with 10% greater birth weight are 1.2% more likely to be full time workers (Table 3). Estimates of the birth weight effect on indicators of child development tend to decrease in absolute value when estimated within twins compared to for all singleton or twin births, especially for 1-year mortality, APGAR score, and full time work. A 10% greater birth weight within twins is associated with only a 1% difference in log full time earnings, a much smaller effect than estimated for Minnesota female identical twins by Behrman and Rosenzweig (2004). Among female twins who both have children in the sample, the within twin estimated effect of log birth weight on the log birth weight of their first child is 0.15, substantially larger than estimated by Behrman and Rosenzweig (2004).
24. Data et al. (2007) argue that Indian rural villages with higher mortality risks evidence larger differences in sibling health inputs (e.g., breastfeeding and immunization) between small-at-birth and large-at-birth siblings. They interpret their findings to suggest that in a higher mortality risk environment, Indian parents increase their concentration of child health inputs on their better endowed at birth of their children.
25. This pattern of larger estimated effects of child health on adult productivity using instrumental variables for economic origins and health conditions at birthplace is also consistent with there being substantial measurement errors in anthropometric health and fetal growth variables, and that the observed variation in individual height may prudently be filtered through the choice of an appropriate instrument to approximate the effect of economic policies which might increase the productive capacity of an adult through its effect on increasing the rate of fetal growth.

26. There is no particular reason to prefer this semi-log functional form except Mincer used it to analyze schooling and earnings. An analysis of data for men and women from the German Socio-Economic Panel suggests that the fit of the log wage function by OLS is improved when the logarithmic transformation of height is the conditioning variable (Hubler, 2006; Table 5). That paper finds evidence that males receive more wages with greater height until they exceed two standard deviations above the sample mean height, whereas the wage gains from height for women reaches its maximum at the mean height for German women. There are many possible interpretations to these nonlinear patterns that have not been extensively explored.
27. The effect of education on health in the United States is appraised by Lleras-Muney (2005) but using OLS and IV estimates identified by changes in state of birth compulsory school attendance legislation. The lack of significant change between the OLS and IV estimates suggests that schooling is exogenous with respect to the determination of health in the US in recent times.
28. If the selection effect of mortality by height were important in low-income countries, as conjectured by Deaton (2007), then the older cohorts would actually have been born shorter on average than the surviving cohorts, and my imputation of a gain in wages by younger cohorts due to their greater height is understated. In other words, the selection effect, to the extent that it occurs in low-income countries, would add to the wage returns to height as reported here due to the increased life span associated with height. The analogous argument applies to cross cohort estimates of returns to schooling, where most evidence suggests survival tends to be greater among the better educated.
29. Famines have also been studied as an instrument affecting health of those born in a specific “window” and compared to birth cohorts before and after. Development economists have frequently used droughts, rainfall, or floods to randomly affect the income constraint in poor populations, providing the instrument to identify the effects of income on health and behavior of family members. An early example is Rosenzweig and Wolpin (1985).
30. Infant mortality is also higher among spring births, and following colder winters, after bad harvests in Sweden (e.g., Eckstein et al., 1984). This raises the possibility that birth cohorts from less healthy months would be culled of the less healthy members, leaving the more fit to survive and live relatively longer lives, other things equal. This countervailing selection effect would conceal the pattern implied by the fetal origins hypothesis analyzed by Doblhammer (2004). Few studies have sought to assess how the selectivity of reproductive behavior modifies the composition of cohorts born in crises periods.
31. To the extent that the stressing conditions are somewhat predictable, the socioeconomic composition of those having children may vary according to their preferences and resources to have healthy or high-quality children. Buckles and Hungerman (2008) show that women who bear children in the winter months in the United States are disproportionately unmarried, teen age, not a high school graduate, and nonwhite. Thus, season of birth may not only reflect the weather during a child's early development, but capture the consequences of being born to a socioeconomic class who for other reasons experiences less adequate health care and educational preparation. Compare with the interpretation of the estimates of compulsory schooling derived from quarter of birth in Angrist and Krueger (1991) or estimates of late age mortality from season of birth in (Doblhammer and Vaupel 2001).
32. In another study, Almond, Edlund, and Palme (2007) analyze data on the radioactive fallout from the Chernobyl accident of 1986, as geographically distributed by rain and wind across Sweden, and find its deposition is associated with lower levels of school completion and IQ of the cohort exposed from the 5th to 25th week of gestation, compared with birth cohorts born before and after this accident, allowing for region fixed effects and time trends.
33. The vitamin D receptor (VDR) affects human stature and mediates calcium and phosphate homeostasis and has been linked to specific gene polymorphisms with wide ramifications (Holick, 2004; Xiong et al., 2005). Since skin pigmentation (i.e., melanin) blocks UVB radiation, lighter skinned races as well as those who are exposed to more intense sunlight (i.e., lower latitudes) appear to have greater

- levels of vitamin D and tend to be taller, although variations in height is more generally attributed to nutritious diet and reduced exposure to inflammatory diseases of early childhood, within biological limits (Finch, 2007; Fogel, 2004; Ginsburg et al., 1998). Vitamin D deficiencies in the US population in the last decade, especially among African Americans, despite the vitamin D fortification of milk started in the 1930s, has led public health experts (e.g., NICHD and CDC) to search for better biomarkers of adequate vitamin D levels, and a better understanding of their consequences on not only bone health, but also associated with other chronic diseases. The determinants of vitamin D deficiency and their consequences on height and health are becoming a public health priority, and economists could analyze this pathway to assess the economic consequences of different sources of variation in height, based on different instruments for height, including geography and diet.
34. Contrary to expectation, some studies find contemporaneous procyclical movement in mortality with the business cycle in today's high-income countries. In the twentieth century with rising levels of income in Netherlands, *current* adult mortality is positively related to the business cycle (Van den Berg et al., 2006). Ruhm (2000) has also shown that adult mortality varies procyclically in the United States given exiting social safety net programs and household assets. He explores pathways such as a possibly healthy reallocation of time from work to families, and from work to leisure activities in periods of recession, despite the concurrent decline in income and increase in unemployment. The changes in mortality rates from age 30 to 85 and real GDP growth rates appears to be positively correlated in six OECD countries until about 1990, as Ruhm (2000) noted in the US and Tapia Granados (2008) in Japan. But after 1990, the correlation becomes more negative (Hanewald, 2009). Unemployment rates also confirm the macroeconomic timing of changes in elderly mortality evolves from procyclical to countercyclical, possibly due to changes in the causes of mortality with the mortality transition (Tapia Granados, 2008). Studies at the macroeconomic level cast doubt on the robustness of the relationship between income growth and gains in life expectation, or even in improvements in infant mortality (Deaton, 2003). This does not dispel the idea that income growth allows for improvements in consumption which are likely to reduce in the long-run mortality, though clearly institutional arrangements and the diffusion of knowledge about the process producing health are important as well.
35. These cross-country regression studies do not control for the inputs to agriculture or the shift of population out of agriculture, such as investments in rural infrastructure, research and development of modern agricultural varieties and inputs, extension services, mechanization, irrigation, or schooling of the labor force. It is difficult to attribute the stagnation in *per capita* availability of calories in Africa to the region's population growth compared to any number of alternative factors. Analogous cross-country regressions of growth in *per capita* income and savings rates on population growth rates also finds that until the 1980s, the simple correlations are positive, in contrast to Malthusian predictions of diminishing returns to population. Only after 1980 does a negative association begin to emerge (Kelley & Schmidt, 1995). After 1980 population growth rates are subsiding in much of Asia and Latin America, and rapid population growth continues mainly in sub-Saharan Africa which has been subject to economic stagnation for many other reasons, including civil conflict, corruption, and government policies. Removing the many small African countries from these regressions that are not weighted by population size weakens the negative associations between population growth or age composition changes and growth in income, savings, and food availability.
36. Schultz (1985) proposes the sharp decline in the world price of cereal grains compared with butter and meat products after 1850, which triggered in Sweden a rise in women's wages in agriculture compared with men's wages across provinces and over time from 1860 to 1910. Because of the specialization of women and men in the production of animal products and grains, respectively, the global change in their relative prices after the opening of the US Middle West to European markets, may have contributed to the coincidence of fertility declines in many agricultural regions of Northern Europe.

37. In Colombia in 1974 women of a specific age and education who had migrated from the rural to urban areas revealed preferences for urban opportunities, which in this setting include better employment prospects for women in the wage labor market, better health services which are particularly effective in preventing child mortality, and better schooling services for youth. Consequently, the rural-urban migrant woman matched by education and age has on average fewer children, and their children are less likely to have died, and more likely to have completed schooling than those women who did not migrate and stayed in the rural areas. More surprising, the rural-urban migrant woman will also have no more children than the urban native women at her destination of the same age and education, and again the migrant's children appear to have no less schooling and about the same child mortality. The migrant woman will again be more likely to be in the labor force than the match of her urban native counterpart. If the heterogeneity in preferences between migrants and natives helps to explain these behavioral differences, it is also clear that the urban-rural differences in health and schooling services will in the cross section be correlated with larger differences in the beneficial outcomes associated with these services than can be reasonably attributed to the program treatments alone (Schultz, 1988b).
38. For children age 12–23 months old, the proportion having received all seven basic recommended vaccinations increased from 42% to 44% between the NFHS 2 collected in 1998/1999 and the NFHS 3 collected in 2004–2006. The prevalence of anemia for children age 6–59 months old increased to 70%, while it increased for all women to 55%, and rising also for pregnant women (National Family Health Survey 3, 2007). Tarozzi and Mahajan (2007) study the first two rounds of this NFHS from 1998/1999 and 1992/1993.
39. Disability-adjusted life years (DALYs) are proposed as a summary measure of the objectives of the health system, refining life expectation at birth to include reductions in disability associated with the control of diseases. But even DALYs focus attention on the immediate decline in mortality and universal estimates of how specific diseases and chronic conditions among adults cause disability. They generally rely on “expert opinion” that is not subject to validation by estimation from primary data. Survey estimates of the prevalence of disabilities tend to be higher in high-income countries suggesting they are culturally conditioned. These current health outcomes do not allow for the cohort linkages emphasized here between fetal and childhood health conditions and their consequences on lifetime economic functioning of adults many years later.

References

- Acemoglu, D. (2002). Technical change, inequality and the labor market. *Journal of Economic Literature*, 40 (1), 7–72.
- Acemoglu, D., & Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy*, 115(6), 925–985.
- Acemoglu, D., Johnson, S., & Robinson, J. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2003). Disease and development in historical perspective. *Journal of the European Economic Association*, 1(2–3), 397–405.
- Akachi, Y., & Canning, D. (2006). *The height of women in Sub Saharan Africa: The role of health, nutrition and Income in childhood*. PGDA Working Paper No. 2207. Boston, MA: Harvard School of Public Health.
- Alderman, H., & Behrman, J. R. (2006). Reducing the incidence of low birth weight in low income countries has substantial economic benefits. *World Bank Research Observer*, 21(1), 25–48.
- Alderman, H., Hoddinott, J., & Kinsey, W. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58(3), 450–474.
- Almond, D. (2006). Is the 1918 Influenza pandemic over? Long term effects of *in utero* exposure in the post-1940 US population. *Journal of Political Economy*, 114(4), 672–712.

- Almond, D., Chay, K., & Lee, D. S. (2005). The cost of low birth weight. *Quarterly Journal of Economics*, 120, 1031–1084.
- Almond, D., Edlund, L., Li, H., & Zhang, J. (2007). *Long-term effects of the 1959–61 Chinese famine: Mainland China and Hong Kong*. Working Paper No. 13384. Cambridge, MA: National Bureau of Economic Research.
- Almond, D., Edlund, L., & Palme, M. (2007). *Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden*. Working Paper No. 13347. Cambridge, MA: National Bureau of Economic Research.
- Alters, G., & Oris, M. (2006). Environment before and after birth and mortality in old age. In *Annual meetings of the population association of America*, Los Angeles, 30 March–1 April, 2006 (also in *Social Biology*, 52(3–4), 178–191).
- Angrist, J., & Evans, W. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3), 450–477.
- Angrist, J., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106, 979–1014.
- Arends-Kuenning, M., & Amin, S. (2004). School incentive programs and children's activities: The case of Bangladesh. *Comparative Education Review*, 48(3), 295–317.
- Atkin, D. (2009). *Trade, tastes, and nutrition in India*. Unpublished Ph.D. chapter, Princeton, NJ: Princeton University.
- Banerjee, A. B., & Duflo, E. (2007). *Aging and death under a dollar a day*. Working Paper No. 13683. Cambridge, MA: National Bureau of Economic Research.
- Barker, D. J. P. (Ed.). (1992). *Fetal and infant origins of adult disease*. London: British Medical Journal.
- Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. London: BMJ Publishing Group.
- Barker, D. J. P. (Ed.). (2001). *Fetal origins of cardiovascular and lung disease*. New York, NY: Marcel Dekker.
- Barker, D. J. P. (2005). Infant growth and income 50 years later. *Archives of Disease in Childhood*, 90, 272–273.
- Barro, R. J. (1997). *Determinants of economic growth: A cross-country empirical study*. Cambridge, MA: MIT Press.
- Barro, R. J., & Lee, J. W. (2000). *International data on educational attainment: Updates and implications*. Working Paper No. 42. Cambridge, MA: Harvard University Press.
- Barro, R. J., & Sala-i-Martin, X. (1995). *Economic growth*. New York: McGraw-Hill.
- Becker, G. S. (1981). *A treatise on the family*. Cambridge, MA: Harvard University Press.
- Becker, G. S., & Barro, R. J. (1988). A reformulation of the economic theory of fertility. *Quarterly Journal of Economics*, 103(1), 1–25.
- Becker, G. S., & Lewis, H. G. (1974). Interaction between quantity and quality of children. In T. W. Schultz (Ed.), *The economics of the family*. Chicago: University of Chicago Press.
- Becker, G. S., Philipson, T. J., & Soares, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *American Economic Review*, 95(1), 277–291.
- Becker, G. S., & Tomes, N. (1976). Child endowment and the quality and quantity of children. *Journal of Political Economy*, 84(4, part 2), 159–179.
- Behrman, J. R. (2001). Why micro-matters. In N. Birdsall, A. C. Kelley & S. W. Sinding (Eds.), *Population matters*. New York: Oxford University Press.
- Behrman, J. R., & Birdsall, N. (1983). The quality of schooling. *American Economic Review*, 73, 928–946.
- Behrman, J. R., & Rosenzweig, M. R. (2004). The returns to birthweight. *Review of Economics and Statistics*, 86(2), 586–601.
- Behrman, J. R., Rosenzweig, M. R., & Taubman, P. (1994). Endowments and the allocation of schooling in the family and in the marriage market: The twin experiment. *Journal of Political Economy*, 102(6), 1131–1174.
- Bengtsson, T. (Ed.), (2009). Early Life effects on socioeconomic performance and mortality in later life. *Social Science & Medicine*, 68, 1561–1564, (entire issue on topics).
- Bengtsson, T., & Brostrom, G. (2006). Old age mortality in a life course perspective: Southern Sweden, 1829–1894. In: *Annual meetings of the population association of America*, Los Angeles, 30 March–1 April, 2006.
- Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4, Part 1), 352–365.

- Birdsall, N., Kelley, A. C., & Sinding, S. W. (2001). *Population matters: Demographic change, economic growth and poverty in the developing world*. New York: Oxford University Press.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The more the merrier? The effect of family composition on children's education. *Quarterly Journal of Economics*, 120(2), 669–700.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From cradle to labor market? The effect of birth weight on adult outcomes. *Quarterly Journal of Economics*, 122(1), 409–439.
- Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American South. *Quarterly Journal of Economics*, 122(1), 72–117.
- Bleakley, H., & Lange, F. (2009). Chronic disease burden and the interaction of education, fertility and growth. *Review of Economics and Statistics*, 91, 52–65.
- Bloom, D. E., Canning, D., & Sevilla, J. (2002). *The demographic dividend: A new perspective on the economic consequences of population change*. Santa Monica, CA: The Rand Corporation (MR-12274).
- Bloom, D. E., Canning, D., & Sevilla, J. (2004). The effects of health on economic growth: A production function approach. *World Development*, 32(1), 1–13.
- Bloom, D. E., & Sachs, J. D. (1998). Geography, demography, and economic growth in Africa. *Brookings Papers in Economic Activity*, 2, 207–295.
- Bozzoli, C., & Deaton, A. (2008). *Adult height and childhood disease*. Princeton, NJ: Center for Health and Wellbeing, Princeton University.
- Brass, W. (1975). *Methods for estimating fertility and mortality from limited and defective data*. Chapel Hill, NC: Carolina Population Center.
- Brinton, M. C., Lee, Y. J., & Parish, W. L. (1995). Married women, employment in rapidly industrializing societies. *American Journal of Sociology*, 100(5), 1099–1130.
- Buckles, K., & Hungerman, D. M. (2008). *Season of birth and later outcomes: Old questions, new answers*. Working Paper No. 14573. Cambridge, MA: National Bureau of Economic Research.
- Caldwell, J. C. (1979). Education as a factor in mortality decline. *Population Studies*, 33, 395–413.
- Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1801–1963). Amsterdam: Elsevier.
- Carson, S. A. (2008). The effect of geography and vitamin D on African American stature in the 19th century: Evidence from prison records. *Journal of Economic History*, 68(3), 812–831.
- Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24, 365–389.
- Case, A., & Paxson, C. (2008). Height, health, and cognitive function at older ages. *American Economic Review*, 98(2), 463–467.
- Cesur, R., & Rashad, I. (2008). *High birth weight and cognitive outcomes*. Working Paper No. 14524. Cambridge, MA: National Bureau of Economic Research.
- Chiappori, P. A. (1992). Collective labor supply and welfare. *Journal of Political Economy*, 100(3), 437–467.
- Christiaensen, L., & Alderman, H. (2004). Child malnutrition in Ethiopia: Can maternal knowledge augment the role of income. *Economic Development and Cultural Change*, 52(2), 287–312.
- Coale, A. J. (1972). *The growth and structure of human populations: A mathematical investigation*. Princeton, NJ: Princeton University Press.
- Coale, A. J., & Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton, NJ: Princeton University Press.
- Coale, A., & Hoover, E. (1958). *Population growth and economic development in low income countries*. Princeton, NJ: Princeton University Press.
- Cochrane, S. H., O'Hara, D. J., & Leslie, J. (1980). *The effects of education on health*. World Bank Staff Working Paper No. 405. Washington, DC: World Bank.
- Costa, D. (1996). Health and labor force participation of older men, 1900–1991. *Journal of Economic History*, 56(1), 62–89.
- Costa, D. L., & Steckel, R. H. (1997). Long-term trends in health, welfare, and economic growth in the United States. In R. H. Steckel & R. Floud (Eds.), *Health and welfare during industrialization*. Chicago: University of Chicago Press.
- Crimmins, E., & Finch, C. (2006a). Infection, inflammation, height and longevity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 498–503.

- Crimmins, E. M., & Finch, C. E. (2006b). Commentary: Do older men and women gain equally from improving childhood conditions? *International Journal of Epidemiology*, 35, 1270–1271.
- Cunha, F., & Heckman, J. J. (2007). *The technology of skill formation*. IZA Discussion Paper No. 2350. Bonn, Germany: Institute for the study of labor (IZA).
- Currie, J., & Hyson, R. (1999). Is the impact of health shocks cushioned by socioeconomic status? *American Economic Review*, 89(2), 245–250.
- Currie, J., & Stabile, M. (2003). Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review*, 93(5), 1813–1823.
- Darwin, C. R. (1859). *The origin of species*. London: John Murray.
- Data, A., Gosh, A., & Sood, N. (2007). *Mortality risk, health endowments, and parental investments in infancy: Evidence from rural India*. Working Paper No. 13647. Cambridge, MA: National Bureau of Economic Research.
- Dawkins, R. (1989). *The selfish gene*. Oxford: Oxford University Press.
- Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature*, 49(1), 113–158.
- Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33), 13232–13237.
- Deaton, A. (2008). Height, health and inequality: Distribution of adult heights in India. *American Economic Review*, 98(2), 468–474.
- Deaton, A., & Paxson, C. H. (1997). The effects of economic and population growth on national savings and inequality. *Demography*, 34(1), 97–114.
- Del Bono, E., Ermisch, J., & Francesconi, M. (2008). *Intrafamily resource allocations: A dynamic model of birth weight*. Unpublished paper, University of Essex, UK.
- Deolalikar, A. B. (1988). Nutrition and labor productivity in agriculture: Estimates for rural South India. *Review of Economics and Statistics*, 70(3), 406–413.
- Diamond, J. (1997). *Guns, germs and steel: The fates of human societies*. New York: W.W. Norton & Co.
- Doblhammer, G. (2004). *The late life legacy of very early life*. Berlin: Springer-Verlag.
- Doblhammer, G., & Vaupel, J. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5), 2934.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia. *American Economic Review*, 91(3), 795–813.
- Duflo, E., Glennerster, R., & Kremer, M. (2008). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: Elsevier.
- Durand, J. D. (1975). *The labor force in economic development*. Princeton, NJ: Princeton University Press.
- Eckstein, Z., Schultz, T. P., & Wolpin, K. I. (1984). Short run fluctuations in fertility and mortality in preindustrial Sweden. *European Economic Review*, 26(3), 295–317.
- Edwards, R. D., & Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, 31(4), 645–674.
- Elo, I. T., & Preston, S. H. (1992). Effects of early life conditions on adult mortality: A review. *Population Index*, 58(2), 186–212.
- Eveleth, P. B. (1986). Population differences in growth, environmental and genetic factors. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (2nd ed., Vol. 3). New York: Plenum Press.
- Eveleth, P. B., & Tanner, J. M. (1990). *World wide variation in human growth* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Field, E., Robles, O., & Torero, M. (2007). *Iodine deficiency and schooling attainment in Tanzania*. Cambridge, MA: Harvard University Press.
- Filmer, D. (2008). Disability, poverty and schooling in developing countries: Results from 14 household surveys. *World Bank Economic Review*, 22(1), 141–163.
- Finch, C. E. (2007). *The biology of human longevity: Inflammation, nutrition and aging in the evolution of lifespan*. Amsterdam: Elsevier.

- Finch, C. E., & Crimmins, E. (2004). Inflammatory exposure and historical changes in human life-spans. *Science*, 305, 1736–1739.
- Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). Attrition in longitudinal surveys. *Journal of Human Resources*, 33(2), 251–299.
- Floud, R., Wachter, K., & Gregory, A. (1990). *Height, health and history: Nutritional status in the United Kingdom, 1750–1980*. New York: Cambridge University Press.
- Fogel, R. W. (1986). Physical growth as a measure of the economic well-being of populations: The 18th and 19th centuries. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (2nd ed., Vol. 3). New York: Plenum Press.
- Fogel, R. W. (1994). Economic growth, population theory and physiology: The bearing of long-term processes on the making of economic policy. *American Economic Review*, 84(3), 369–395.
- Fogel, R. W. (2004). *The escape from hunger and premature death, 1700–2100*. Cambridge, UK: Cambridge University Press.
- Foster, A., & Rosenzweig, M. R. (2007). Does economic growth reduce fertility: Rural India, 1971–1999. In S. Bery & A. Panagariya (Eds.), *India policy forum 2006–07*. Delhi, India: Brookings and NCAER.
- Frankenberg, E. (1995). The effects of access to health care on infant mortality in Indonesia. *Health Transition Review*, 5(2), 143–163.
- Frankenberg, E., Sikoki, B., Suriastini, W., & Thomas, D. (2003). Contraceptive use in a changing service environment: Evidence from Indonesia during the economic crisis. *Studies in Family Planning*, 34(2), 102–116.
- Frankenberg, E., Suriastini, W., & Thomas, D. (2005). Can expanding access to basic health care improve children's outcomes? *Population Studies*, 59(1), 5–19.
- Freedman, R. (1975). *The sociology of human fertility*. New York: John Wiley.
- Gallup, J. L., & Sachs, J. D. (2001). The economic burden of malaria. *The American Journal of Tropical Medicine and Hygiene*, 64(1), 85–96.
- Galor, O. (2005). From stagnation to growth: Unified growth theory. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1, pp. 171–293). Amsterdam: Elsevier.
- Galor, O., & Weil, D. N. (1999). Malthusian stagnation to modern growth. *American Economic Review*, 89 (2), 150–154.
- Galor, O., & Weil, D. N. (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review*, 90(4), 806–828.
- Gertler, P. (2004). Do conditional cash transfers improve child health? Evidence from Progresa's control randomized experiment. *American Economic Review*, 94(2), 336–341.
- Ginsburg, E., Livshits, G., Yakovenko, K., & Kobylansky, E. (1998). Major gene control of human body height, weight, and BMI in five ethnically different populations. *Annals of Human Genetics*, 62, 307–322.
- Glewwe, P. (2005). The impact of child health and nutrition on education in developing countries. *Food and Nutrition Bulletin*, 26(2), S235–S250.
- Glewwe, P., & Jacoby, H. (1995). An economic analysis of delayed primary school enrollment in a low income countries: Role of early childhood nutrition. *Review of Economics and Statistics*, 77(1), 156–169.
- Glewwe, P., Jacoby, H., & King, E. (2000). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics*, 81(3), 345–368.
- Glewwe, P., & Miguel, E. A. (2008). The impact of child health and nutrition on education in less developed countries. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: Elsevier.
- Gluckman, P., & Hanson, M. (2005). *The fetal matrix: Evolution, development and disease*. Cambridge, UK: Cambridge University Press.
- Gluckman, P., & Hanson, M. (Eds.). (2006). *Developmental origins of health and disease*. Cambridge, UK: Cambridge University Press.
- Goldin, C., & Katz, L. F. (1998). The origins of technology-skill complementarity. *Quarterly Journal of Economics*, 113(3), 693–723.

- Goldin, C., & Katz, L. F. (2000). Education and income in the early twentieth century: Evidence from the Prairies. *Journal of Economic History*, 60(September), 782–818.
- Goldin, C., & Katz, L. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.
- Gorgens, T., Meng, X., & Vaithianathan, R. (2007). *Stunting and selection effects of famine: A case study of the great Chinese Famine*. IZA Discussion Paper No. 2543. Bonn, Germany: Institute for the study of labor (IZA).
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica*, 45(1), 1–45.
- Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*, 87(5, Part 2), S37–S64.
- Gruber, J., & Wise, D. (1995). *Social security programs and retirement around the world*. Working Paper No. 6134. Cambridge, MA: National Bureau of Economic Research.
- Gruber, J., & Wise, D. (2004). *Social security programs and retirement around the world: Micro-estimation*. Chicago: University of Chicago Press.
- Hanewald, K. (2009). *Mortality modeling: Lee-Carter and the macroeconomy*. SFB Discussion Paper No. 2009–008. Berlin: Humboldt-Universität zu Berlin.
- Hazen, M. (2007). *Longevity and lifetime labor input: Data and implications*. Working Paper. Jerusalem: Hebrew University.
- Heckman, J. J. (2008). *Schools, skills and synapses*. ISA Discussion Paper No. 3515. Bonn, Germany: Institute for the study of labor (IZA).iza@iza.org.
- Higgins, M., & Williamson, J. G. (1997). Age structure dynamics in Asia and dependence on foreign capital. *Population and Development Review*, 23(2), 261–294.
- Holick, M. F. (2004). Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *American Journal of Clinical Nutrition*, 80(Suppl.), 1678S–1688S.
- Hubler, O. (2006). October *The nonlinear link between height and wages: An empirical investigation*. IZA Discussion Paper 2394. iza.org/publications/dps/.
- Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467–476.
- Johnson, D. G. (1999). Population and development. *China Economic Review*, 10(1), 1–16.
- Joshi, S., & Schultz, T. P. (2007). *Family planning as an investment in development: Evaluation of a program's consequences in Matlab, Bangladesh*. Economic Growth Center Discussion Paper No. 951. New Haven, CT: Yale University Press.
- Kelley, A. C., & Schmidt, R. M. (1995). Aggregate population and economic growth correlations: The role of the components of demographic change. *Demography*, 32(4), 543–555.
- Koenig, M. A., Fauveau, V., & Woityniak, B. (1991). Mortality reduction from health interventions: The case of immunization in Bangladesh. *Population and Development Review*, 17(1), 87–104.
- Komlos, J. (Ed.), (1994). *Stature, living standards, and economic development*. Chicago: University of Chicago Press.
- Kravdal, O. (2001). Has population growth restricted improvements in food availability per head, 1970–1995. *Population Studies*, 55, 105–117.
- Kuznets, S. (1966). *Modern economic growth: Rate, structure and spread*. New Haven, CT: Yale University Press.
- Kuznets, S. (1967). Population and economic growth. *Proceedings of the American Philosophical Society*, 3(3), 170–193.
- Lapham, R. J., & Mauldin, W. P. (1984). Family planning effort and birthrate decline in developing countries. *International Family Planning Perspectives*, 10(4), 109–118.
- Liu, Y. Z., Xu, F.-H., Shen, H., Liu, Y. J., Zhao, L. J., Long, J. R., et al. (2004). Genetic dissection of human stature in a large sample of multiplex pedigrees. *Annals of Human Genetics*, 68(5), 472–488.
- Lleras-Muney, A. (2005). The relationship between education and adult health in the United States. *Review of Economic Studies*, 72, 189–221.
- Lokshin, M., & Radyakin, S. (2009). *Month of birth and children's health in India*. Policy Research Working Paper No. 4813. Washington, DC: World Bank.

- Lorentzen, P., McMillan, J., & Wacziarg, R. (2008). Death and development. *Journal of Economic Growth*, 13, 81–124.
- Lucas R. E., Jr., (2002). *Lectures on economic growth*. Cambridge, MA: Harvard University Press.
- Lutz, W., & Goujon, A. (2001). The world's changing human capital stock: Multistage population projections by educational attainment. *Population and Development Review*, 27, 323–339 (see projection database on ASSA website).
- Maccini, S., & Yang, D. (2006). *Under the weather: Health, schooling, and socioeconomic consequences of early-life rainfall*. Ann Arbor, MI: University of Michigan.
- Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A., & Stein, A. D. (2006). *The impact of an experimental nutritional intervention in childhood on education among Guatemalan adults*. Working Paper No. 0614. Middlebury, VT: Middlebury College.
- Martorell, R., & Habicht, J. P. (1986). Growth in early childhood in developing countries. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (2nd ed., Vol. 3). New York: Plenum Press.
- McKeown, T. (1976). *The modern rise of population*. New York: Academic Press.
- McKeown, T. (1979). *The role of medicine: Dream, mirage, or nemesis*. Princeton, NJ: Princeton University Press.
- Mensch, B., Lentzner, H., & Preston, S. (1985). *Socioeconomic differentials in child mortality in developing countries*. New York: United Nations.
- Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159–217.
- Mincer, J. (1963). Market prices, opportunity costs and income effects. In C. Christ (Ed.), *Measurement in economics*. Stanford, CA: Stanford University Press.
- Mincer, J. (1974). Schooling, Experience and Earnings, New York: Colombia University Press for National Bureau of Economic Research.
- Modigliani, F., & Brumberg, R. (1954). Utility analysis and the consumption function. In K. K. Kurihara (Ed.), *Post-Keynesian economics*. (Chapter 15). New Brunswick, NJ: Rutgers University Press.
- Moffitt, R. (2005). Remarks on the analysis of causal relationships in population research. *Demography*, 42(1), 91–108.
- Murphy, K. M., & Topel, R. H. (2006). The value of health and longevity. *Journal of Political Economy*, 114(5), 871–904.
- Mwabu, G. (2008). Health economics in low income countries. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: Elsevier.
- Mwabu, G. (2009). The production of child health in Kenya: A structural model of birth weight. *Journal of African Economics*, 18, 212–260.
- Nabli, M. K., Silva Fauregui, C., & De Silva, S. J. (2007). *Job creation in a high growth environment: The MENA region*. Middle East and North Africa. Working Paper No. 49. Washington, DC: World Bank.
- National Family Health Survey 3, Indian Government. (2007, October 15). Summary Final Report. alerts@measuredhs.com.
- National Research Council. (1986). *Population growth and economic development: Policy questions*. Washington, DC: National Academy Press.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296, 1029–1031.
- Ozler, S. (2000). Export orientation and female share of employment: Evidence from Turkey. *World Development*, 28(7), 1239–1248.
- Panizza, U. (2003). *The public sector premium and the gender gap in Latin America: Evidence for the 1980s and 1990s*. Washington, DC: Inter-American Development Bank (SSRN, id = 238700).
- Parker, S. W., Rubalcava, L., & Teruel, G. (2008). Evaluating conditional schooling and health programs. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: North Holland Press/Elsevier.
- Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, 29(2), 231–248.

- Preston, S. H. (1980). Causes and consequences of mortality declines in less developed countries during the twentieth century. In R. A. Easterlin (Ed.), *Population and economic change in developing countries*. Chicago: University of Chicago Press.
- Preston, S. H., Hill, M. E., & Drenenstedt, G. L. (1998). Childhood conditions that predict survival to advanced ages among African Americans. *Social Science & Medicine*, 47(9), 1231–1246.
- Pritchett, L. H. (1994). Desired fertility and the impact of population policies. *Population and Development Review*, 20(1), 1–55.
- Pritchett, L., & Summers, L. H. (1996). Wealthier is healthier. *Journal of Human Resources*, 31(4), 841–868.
- Raiten, D. J., & Picciano, M. F. (2004). Vitamin D and health in the 21st century: Bone and beyond. Executive summary. *American Journal of Clinical Nutrition*, 80(Suppl.), S1673–S1677.
- Rose, E. (2000). Gender bias, credit constraints and time allocation in rural India. *Economic Journal*, 110(465), 738–758.
- Rosenzweig, M. R. (1990). Population growth and human capital investments: Theory and evidence. *Journal of Political Economy*, 98(5), S38–S70.
- Rosenzweig, M. R., & Schultz, T. P. (1983). Estimating a household production function: Heterogeneity, the demand for health inputs, and the effect on birthweight. *Journal of Political Economy*, 91(5), 723–746.
- Rosenzweig, M. R., & Schultz, T. P. (1985). The supply and demand of births and their life cycle consequences. *American Economic Review*, 75(5), 922–1015.
- Rosenzweig, M. R., & Schultz, T. P. (1987). Fertility and investment in human capital in Malaysia. *Journal of Econometrics*, 36, 163–184.
- Rosenzweig, M. R., & Wolpin, K. I. (1980a). Testing the quantity-quality fertility model. *Econometrica*, 48(1), 227–240.
- Rosenzweig, M. R., & Wolpin, K. I. (1980b). Life cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 88(2), 328–348.
- Rosenzweig, M. R., & Wolpin, K. I. (1985). Specific experience, household structure and intergenerational transfers: Farm family land and labor arrangements in developing countries. *Quarterly Journal of Economics*, 100(Suppl.), 461–487.
- Rosenzweig, M. R., & Wolpin, K. I. (1986). Evaluating the effects of optimally distributed public programs: Child health and family planning interventions. *American Economic Review*, 76(3), 470–482.
- Rosenzweig, M. R., & Zhang, J. (2009). Do population control policies induce more human capital investment? Twins, birthweight, and China's one child policy. *Review of Economic Studies*, 76(3), 1149–1174.
- Royer, H. (2009). Separated at birth: U.S. twin estimates of the effect of birth weight. *American Economic Journal: Applied Economics*, 1(1), 49–85.
- Ruhm, C. J. (2000). Are recessions good for your health? *Quarterly Journal of Economics*, 115(2), 617–650.
- Savedoff, W., & Schultz, T. P. (Eds.). (2001). *Wealth from health*. Washington, DC: Inter-American Development Bank.
- Schultz, T. P. (1980). Interpretation of relations among mortality, economic of the household, and the health environment. In *Socioeconomic determinants and consequences of mortality differentials*. Geneva: World Health Organization.
- Schultz, T. P. (1981). *Economics of population*. Reading, MA: Addison-Wesley Publisher.
- Schultz, T. P. (1982). Notes on the estimation of migration decision functions. In R. Sabot (Ed.), *Migration and the labor market in developing countries*. Boulder, CO: Westview Press.
- Schultz, T. P. (1985). Changing world prices, women's wages and the fertility transition: Sweden 1860–1910. *Journal of Political Economy*, 93(6), 1126–1154.
- Schultz, T. P. (1988a). Education investments and returns. In H. Chenery & T. N. Srinivasan (Eds.), *Handbook of development economics* (Vol. 1, Chapter 13, pp. 543–630). Amsterdam: North Holland Publishing Company.
- Schultz, T. P. (1988b). Heterogeneous preferences and migration: Self-selection, regional prices and programs and the behavior of migrants in Colombia. In *Research in population economics* (Vol. 6). Greenwich, CT: JAI Press.
- Schultz, T. P. (1990). Women's changing participation in the labor force. *Economic Development and Cultural Change*, 38(3), 451–488.

- Schultz, T. P. (1994). Human capital, family planning and their effects on population growth. *American Economic Review*, 83(2), 255–260.
- Schultz, T. P. (Ed.), (1995). *Investment in women's human capital*. Chicago: University of Chicago Press.
- Schultz, T. P. (1997). The demand for children in low income countries. In M. R. Rosenzweig & O. Stark (Eds.), *Handbook of population and family economics* (Vol. 1A). Amsterdam: North Holland Publishing Company.
- Schultz, T. P. (2001). Women's roles in the agricultural household: Bargaining and human capital investments. In B. Gardner & G. Rausser (Eds.), *Agricultural and resource economics handbook*. Amsterdam: Elsevier.
- Schultz, T. P. (2002). Wage gains associated with height as a form of health human capital. *American Economic Review*, 92(2), 349–353 (Economic Growth Center Discussion Paper No. 841, Yale University Press for appendix tables).
- Schultz, T. P. (2003). Wage rentals for reproducible human capital: Evidence from two West African countries. *Economics & Human Biology*, 1(3), 331–366.
- Schultz, T. P. (2004a). Human resources in China: The birth quota, returns to education and migration. *Pacific Economic Review*, 93(3), 245–267.
- Schultz, T. P. (2004b). *Demographic determinants of savings: Estimating and interpreting aggregate associations in Asia*. Discussion Paper No. 1479 Beijing: China Center for Economic Research. <http://www.iza.org>.
- Schultz, T. P. (2005). Productive benefits of health: Evidence from low income countries. In G. Lopez-Casasnovas, B. Rivera & L. Currais (Eds.), *Health and economic growth: Findings and policy implications*. Cambridge, MA: MIT Press.
- Schultz, T. P. (2008a). Population policies, fertility, women's human capital, and child quality. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4, Chapter 52). Amsterdam: North Holland Press/Elsevier (Economic Growth Center Discussion Paper No. 954, Yale University Press).
- Schultz, T. P. (2008b). Can family planning programs promote development beyond fertility reduction? Evidence from a long-term social experiment in Bangladesh—1977–1996. presented at a *Gates foundation/center for international development conference*, on *Health Improvements for Economic Growth*.
- Schultz, T. P., & Tansel, A. (1997). Wage and labor supply effects of illness in Cote d'Ivoire and Ghana: Instrumental variable estimates for days disabled. *Journal of Development Economics*, 53(2), 251–286.
- Schwartz, A. (1976). Migration, age and education. *Journal of Political Economy*, 84(4, Part 1), 701–720.
- Scrimshaw, N. S., Taylor, C. E., & Gordon, J. E. (1968). *Interactions of nutrition and infection*. World Health Monograph Series No. 57. Geneva: World Health Organization.
- Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, 35, 263–285.
- Silventoinen, K., Kaprio, J., Lahelma, E., & Koskenvuo, M. (2000). Relative effect of genetic and environmental factors on body height: Differences across birth cohorts among Finnish men and women. *American Journal of Public Health*, 90, 627–630.
- Steckel, R. H. (1995). Stature and the standard of living. *Journal of Economic Literature*, 33, 1903–1940.
- Steckel, R. H. (2008). Height and human welfare: Recent developments and new directions. Working Paper No. 14536. Cambridge, MA: National Bureau of Economic Research.
- Strauss, J. (1986). Does better nutrition raise farm productivity? *Journal of Political Economy*, 94(2), 297–320.
- Strauss, J., & Thomas, D. (1998). Health, nutrition, and economic development. *Journal of Economic Literature*, 36(2), 766–817.
- Strauss, J., & Thomas, D. (2008). Health over the life course. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4). Amsterdam: Elsevier.
- Stunkard, A. J., Foch, T. T., & Hrubec, Z. (1986). A twin study of human obesity. *Journal of American Medical Association*, 256, 51–54.
- Su, D. (2005). Body mass index and old age survival. *American Journal of Human Biology*, 17(3), 341–354 and presented at the *Annual meetings of the population association of America*, Los Angeles, 31 March–1 April, 2006 Chicago: Center for Population Economics, University of Chicago.
- Summers, R., & Heston, A. (1991). The Penn World Table (Mark 5). *Quarterly Journal of Economics*, 106 (2), 327–368.

- Sunder, M. (2005). Toward generation XL: Anthropometrics of longevity in late 20th century United States. *Economics of Human Biology*, 3(3), 271–295.
- Sunder, M., & Woitek, U. (2005). Boom, bust, and the human body: Further evidence on the relationship between height and business cycle. *Economics & Human Biology*, 3(3), 450–466.
- Tanner, J. M. (1981). *A history of the study of human growth*. Cambridge: Cambridge University Press.
- Tanner, J. M. (1982). The potential of auxological data for monitoring economic and social well-being. *Social Science History*, 6(4), 571–581.
- Tapia Granados, J. A. (2008). Macroeconomic fluctuations and mortality in postwar Japan. *Demography*, 45, 323–343.
- Tarozzi, A. (2009). Growth reference charts and the nutritional status of Indian children. *Economics & Human Biology*, 6(3), 455–468.
- Tarozzi, A., & Mahajan, A. (2007). Child nutrition in India in the nineties. *Economic Development and Cultural Change*, 54(3), 441–486.
- Thirumurthy, H., Zivin, J. G., & Goldstein, M. (2008). The economic impact of AIDS treatment: Labor supply in Western Kenya. *Journal of Human Resources*, 43(3), 511–552.
- Thomas, D. (1990). Intrahousehold resource allocation: An inferential approach. *Journal of Human Resources*, 25(4), 635–664.
- Thomas, D. (2008). Health and socioeconomic status: Isolating causal pathways. In: *ABCDE World Bank conference*, Cape Town, South Africa, 9–11 June 2008.
- Thomas, D., Frankenberg, E., Friedman, J., Habicht, J.-P., Jones, N., McKelvey, C., et al. (2006). *Causal effect of health on labor market outcomes: Experimental evidence*. Los Angeles, CA: UCLA.
- Tolley, G., Kenkel, D., & Fabian, R. (Eds.). (1994). *Valuing health for policy: An economic approach*. Chicago: University of Chicago Press.
- Topel, R. (1999). Labor markets and economic growth. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3c). Amsterdam: Elsevier.
- United Nations. (1967). *Methods of estimating basic demographic measures from incomplete data, Manual IV*. Population Studies No. 42. New York: Department of Economic and Social Affairs, United Nations.
- United Nations. (2003). *World population prospects: The 2002 revision* (Vol. 1). New York: Department of Economic and Social Affairs, Population Division, United Nations.
- United Nations. (2004). *World population policies*. New York: United Nations.
- United Nations Development Programme. (2000). *Human development report 2003*. New York: Oxford University Press.
- United Nations Fund for Population Activities (UNFPA/NIDA) (2009). Resource Flows Project Report (Tables 1–6; 3 March 2009). <http://www.resourceflows.org/index.php/articles/288>.
- Usher, D. (1973). An imputation to the measure of economic growth for changes in life-expectancy. In M. Moss (Ed.), *The measurement of economic and social performance*. New York: Columbia University Press.
- Van den Berg, G., Lindenbroom, M., & Portrait, F. (2006). Economic conditions early in life on individual mortality. *American Economic Review*, 96(1), 290–302.
- Van Wieringen, J. C. (1986). Secular growth changes. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (2nd ed., Vol. 3). New York: Plenum Press.
- Waaler, H. T. (1984). Height, weight, and mortality: The Norwegian experience. *Acta Medica Scandinavica*, 77(Suppl. 679), 1–56.
- Weil, D. N. (2007). Accounting for the effects of health on economic growth. *Quarterly Journal of Economics*, 112(3), 1265–1306.
- White, K. M. (2002). Longevity advances in high-income countries, 1955–96. *Population and Development Review*, 28(1), 59–76.
- Willis, R. J. (1974). Economic theory of fertility behavior. In T. W. Schultz (Ed.), *Economics of the family*. Chicago: University of Chicago Press.
- Wilson, E. O. (1975). *Sociobiology*. Cambridge, MA: Harvard University Press/Belknap Press.
- Wood, A. (1995). North-South trade and female labour in manufacturing: An asymmetry. *Journal of Development Studies*, 27(2), 168–189.
- Wooldridge, J. M. (2002). *Econometric analysis of cross sectional and panel data*. Cambridge, MA: MIT Press.
- World Bank. (1993). *World development report 1993*. New York: Oxford University Press.

- World Bank. (2008). *Growth report*. Washington, DC: World Bank.
- World Health Organization (WHO). (1995). *Physical status: The use and interpretation of anthropometry. Technical Report Series No. 854*. Geneva: WHO.
- World Health Organization (2001). *Macroeconomics and health: Investing in health for economic development*. Report of the commission on macroeconomics and health. Geneva: WHO. <http://www.who.int/shosis/cmh>.
- World Health Organization (2003). *Guide to producing national health accounts with special applications for low-income and middle-income countries*. Geneva: World Health Organization.
- World Health Organization (2006). *Child growth standards: Length/height for age, weight for age, weight for length, weight for height, and body mass index for age: Methods and development*. Geneva: WHO.
- Xiong, D. H., Xu, F. H., Liu, P. Y., Shen, H., Long, J. R., Elze, L., et al. (2005). Vitamin D receptor gene polymorphisms are linked to and associated with adult height. *Journal of Medical Genetics*, 42, 228–234.

This page intentionally left blank

Investment in Education—Inputs and Incentives*

Jere R. Behrman¹

Department of Economics, University of Pennsylvania, McNeil 160, Philadelphia, Pennsylvania, USA

Contents

1. Introduction	4884
2. Framework for Analysis and What We Know and Do Not Know	4886
2.1 Demand side—Models, estimates, and policy implications	4886
3. Supply Side—Models, Estimates, and Policy Implications	4914
3.1 Supply-side modeling	4915
3.2 Estimates and policy implications	4920
4. Markets for Products and Inputs—Models, Estimates, and Policy Implications	4943
4.1 Sorting in educational markets	4943
4.2 Supplies and sectoral choices of childcare workers, teachers, trainers, and other educational service sector staff	4944
4.3 Educational expenditures and age distributions	4945
5. Conclusions	4946
Appendix: Estimation Issues, Possible Methodological Resolutions, and Data	4947
A.1 Econometric Approaches to Dealing with Estimation Problems	4947
A.2 Strengths and Limitations of Analysis of Various Types of Data and Related Methods	4952
End Notes	4959
References	4965

Abstract

From the point of view of economic development, education is the acquisition of knowledge and skills through experiences from conception onwards over the life cycle that increase productivity broadly defined. Education can occur through, but is not limited to, formal educational activities such as preschool programs, schools, and formal training programs. The proximate determinants of education are experiences or inputs into knowledge and skills production functions. Within a dynamic forward-looking model of human capital investments, these experiences are determined sequentially by a series of family or individual decisions given past, current and expected future resources, markets, policies, and other institutions. The context in which these microinvestment demands are made, in turn, reflects decisions of suppliers of services that are explicitly related to education as well as of options that may be importantly related to education through other experiences, such as in labor markets. To understand the nature of inputs and incentives related to education in developing countries, attention must be paid to both the demand and the supply sides for investments in education, both of which are conditioned

significantly by policy choices. Therefore, there are numerous important policy questions related to educational inputs and incentives. What are critical inputs into different educational processes? How important are various incentives for improving these inputs? How effective are various demand-side policies versus supply-side policies? How important are policies that have direct impact on input decisions versus policies that alter longer-run incentives to invest in current education through altering expected longer-run returns from such investments? What are the benefits relative to the resource costs of alternative policies for improving educational inputs? This chapter assesses the current state of empirical knowledge, and gaps in that knowledge, on educational incentives and inputs in developing countries as related to such questions.

JEL classification: O100, Economic Development: General; O150, Economic Development: Human Resources; Human Development; Income Distribution; Migration, I000, Health, Education, and Welfare: General; I210, Analysis of Education, 1280, Education: Government Policy, I290, Education: Other

Keywords

education,
knowledge,
skills
human capital
test scores
incentives for education
economic development
schooling
preschool programs
early childhood development
training programs
learning by doing
program evaluation
experimental trials
demand-side educational policies
supply-side educational policies
school vouchers
conditional cash transfers
teacher incentives
student incentives
educational product markets
teacher markets
labor markets

1. INTRODUCTION

Education is broadly defined to be the acquisition of knowledge and skills through experiences from birth, or perhaps conception, onwards over the life cycle.² From the point of view of economic development, education is the acquisition of knowledge

and skills that increase productivity broadly defined. Education is thought by many to be an important input into economic development, if not the development process itself (as in the “capabilities” emphasis of Sen and others). Education can occur through, but is not limited to, formal educational activities such as preschool programs, schools, and formal training programs.

The proximate determinants of education are the experiences or inputs into the production function that determines knowledge and skills that increase productivity. Within a dynamic forward-looking model of human capital investment demand behaviors, these experiences are determined sequentially by a series of family or individual decisions given past, current and expected future resources, markets, policies, and other institutions. The market, policy, and institutional context in which these microinvestment demands are made, in turn, reflects the decisions of suppliers of services that are explicitly related to education such as preschool, school, and training programs as well as of options that may be importantly related to education through other experiences, such as those in labor markets.

This chapter focuses on empirical studies of inputs and incentives for investments in education in developing countries.³ To understand the nature of inputs and incentives related to education in developing countries, attention must be paid to both the demand and the supply sides for investments in education. And both the demand and the supply sides generally are conditioned significantly by policy choices. Therefore, there are a number of important policy questions related to educational inputs and incentives. Some examples include the following.

What are the critical inputs into different educational processes? How important are various incentives for improving these inputs? How effective are various demand-side policies (e.g., conditional income transfers, vouchers, tax incentives for education, attendance requirements) versus various supply-side policies (e.g., teacher incentives, expanded and improved capital stock, decentralization)? How important are such policies that have direct impact on input decisions versus policies that alter the longer-run incentives to invest in current education through altering the expected longer-run returns from such investments? What are the benefits relative to the resource costs of alternative policies for improving educational inputs?⁴ How important are peer effects and sorting of educational staff across heterogeneous educational service providers? What are the costs and the gains from policies that discriminate among various types of providers of educational services? What accounts for differences in educational outcomes for various demographic groups identified by characteristics such as gender, ethnicity, caste, tribe, and race? How central is formal schooling for the educational process? To what extent are pre- and postschool forms of education substitutes or complements for schooling? To what extent is there evidence of differential private and social impacts that would provide a justification for the use of public resources to subsidize (or tax) education to increase efficiency? How does the relatively poor

performance of developing countries on internationally comparable tests (e.g., Glewwe & Kremer, 2006) relate to the nature of educational inputs and incentives related to those inputs?

This chapter assesses the current state of empirical knowledge, and gaps in that knowledge, on educational incentives and inputs in developing countries as related to such questions.

2. FRAMEWORK FOR ANALYSIS AND WHAT WE KNOW AND DO NOT KNOW

Part II of this chapter lays the foundation for interpretation of empirical studies related to education and development in general and inputs and incentives in particular and then turns to selected empirical studies and policy implications. The three sections in Part II first present general economic frameworks of behavior for modeling the demand-side (Section 2), the supply-side (Section 3), and more aggregate market or broader behaviors related to education and development and incentives and inputs (Section 4). To proceed beyond “black box” impact assessments, such frameworks are essential for interpreting empirical studies. Then in each of these three sections, selected empirical studies are summarized. The emphasis is on empirical studies of inputs and incentives related to education in developing countries, mostly fairly recent studies, but with some reference to related studies in developed economies. To estimate critical parameters suggested by these frameworks, however, requires good data and estimation approaches that can account for some of the implications of such frameworks, including behavioral choices, selected samples, and measurement problems. Because these concerns cut across empirical studies in all of the sections of Part II, to avoid duplication they are summarized in one location in Appendix.

2.1 Demand side—Models, estimates, and policy implications

Decisions regarding investments in education and responses to incentives that may affect those decisions occur in important part on the demand side. Consider the following sketch of components of a stylized life-cycle model with multiple stages in which decisions are made sequentially about important inputs that determine an individual’s education in response to various incentives that are faced. For some specific purposes, it would be desirable to have in such a model many relative short periods, such as annual periods that coincide with preschool and school (and possibly postschool) enrollment cycles. For the purpose of organizing the discussion in the present chapter, however, focus on three life-cycle stages suffices. Because of the concentration on schooling as an input into the production of education (indeed often equated with education, though this chapter argues that that is misleading), it is useful here to consider

one stage to be the school-age years and the other two to be the preschool years and the postschool years:

Life-cycle stage 1: preschool (from conception through about age 5 or 6)—preschool education (child cognitive and noncognitive development) as well as physical development determined by factors such as nutrition, health, infectious diseases and stimulation as provided in the home, neighborhood, and through human capital service providers such as health clinics and preschool programs.

Life-cycle stage 2: school ages⁵ (from age 6 or 7)—education determined in part by formal schooling and characteristics of that schooling but also by out-of-school experiences ranging from homework to labor market work, all conditional on stage 1 outcomes as well as on individual, family, market, and institutional characteristics.

Life-cycle stage 3: postschool—education determined in part by formal programs such as training programs but primarily by experiences in labor markets, household production, and other activities.

As is discussed below in this section and in Appendix, household demand-side decisions can be modeled as if they are maximizing an objective function subject to budget constraints and production functions, which leads to a set of reduced-form demand relations. Estimates of the “structural” objective functions and production functions can give insight into the underlying technologies, typically under strong assumptions about functional forms. But these relations in themselves do not give estimates of *total* policy impacts because, for example, in response to a policy change directed toward one input in a production function households in general may alter other inputs into the production function. However, such structural relations may give some insight into the pathways through which policies have an impact. Further, they can be combined into “structural models” that can be used to simulate counterfactual policies, again subject to the assumptions necessary to estimate the structural models. Estimation of reduced-form demand relations, on the other hand, tend to require less strong assumptions and yield estimates of impacts of actual policies, but without insight into what are the channels through which policies are acting and with less potential for exploring counterfactual policies.

This section discusses in sequence the components of underlying behavioral models determining household responses to incentives regarding inputs in education—objective functions, budget constraints, and production functions—and then considers reduced-form demand relations that may be consistent with such models. While demand-side models (and supply-side models as in Section 3) help organize thoughts about what might affect different types of education and how empirically one might explore such possibilities, empirical estimates are necessary to understand the extent to which the various possible determinants of education that are suggested by these

models in fact are important within particular contexts. Summaries of selected empirical studies and possible policy implications are given for the production function and reduced-form estimates in Sections 2.1.3 and 2.1.4.

2.1.1 Objective function(s)

Decisions about investments in human capital in early life, many of which may have implications for knowledge and skills over individuals' life cycles, typically are assumed to be made primarily by parents or other family caregivers. The knowledge and skills of their children or the resources that the children are expected to have later in their lives based in part on their knowledge and skills may either enter directly into the parents' preference functions (because the parents are altruistic, perhaps simply to perpetuate their gene pool) and/or indirectly affect parental utility (because at least part of the returns from the human capital investments in their children enters into budget constraints relevant for the parents' own consumption at some time in the parents' life, possibly including while the children are young and/or while the parents are old). Expectations regarding future impacts of investments in children's knowledge and skills also may enter directly into the parental preference functions if the parents are interested altruistically in the children's expected future control over resources that is generated in part by their knowledge and skills.

Parental preference functions may differ between fathers and mothers regarding the impact on preferences of outcomes affected by child knowledge and skills even if there is considerable assortative mating between individuals and/or their parental families. A common perception is that mothers tend to weigh in their preferences more heavily child outcomes than do fathers. This perception has affected some major programs. For instance, the well-known Mexican antipoverty and human resource investment program *PROGRESA/Oportunidades* (and a number of programs elsewhere that were designed in part on the basis of this program) gives transfers to mothers explicitly because of the interpretation that previous social science evidence showed that more resources under the control of mothers as opposed to fathers tends to increase human resource investments in children (Levy, 2006; Levy & Rodriguez, 2004). Martinelli and Parker (2003) show that the conditionality of such programs (e.g., the transfers are conditional on, e.g., children attending school) can be welfare-improving for mothers and children at the cost of fathers. If there are such differences in preferences, then the joint parental objective function will reflect the relative bargaining positions of the individual parents and perhaps of other relatives, which in many common explanations depends on the relative resources actually or potentially under the command of each of the individuals involved in the bargaining.

While in life-cycle stage 1 the objective function of the child's parents is likely to dominate, in life-cycle stage 3 the objective function of the child as an adult is likely to dominate, perhaps reflecting not only the adult child's but also the adult child's

partner's preferences and relative resource position if there is intrafamilial bargaining over the allocation of resources in that life-cycle stage. While the considerations underlying this objective function are likely to be similar to those for the child's parents, there is a generational shift in the relevant objective function for most investments in the knowledge and skills of the adult child. Between early childhood and early adulthood, moreover, there is likely to be a steady shift in the relative strength of the bargaining positions of the children who are becoming adults and their parents, perhaps because the children have increasing direct resource-generating capacities through increased knowledge and skills as they complete formal schooling and training programs and gain greater work experience in labor markets and in household production. Even when the child is a mature adult, nevertheless, there may be some intergenerational bargaining over resources, as in models of bequests in response to attention paid by adult children to their aging parents (e.g., Bernheim, Shleifer, & Summers, 1985) or bargaining among adult siblings (and siblings-in-law) about amounts and forms of care for aging parents and parents in law.

In summary, the locus of decision-making relevant to investment in an individual's education is likely to shift from parents to the child over the child's life cycle, with bargaining initially between parents and increasingly between parents and children and then the adult child and his/her partners and siblings. But expected or actual adult knowledge and skills of the children may enter directly into the preference functions of the changing set of individuals involved in making decisions—and considerations such as the number as well as quality of children, life expectancies, discount rates, and risk aversion are likely to play roles in the preference functions or affecting the variables that enter into the preference functions of those involved in such decisions. Estimates of key parameters of preference or objective functions are difficult, particularly because utility is not observable but also because many of the arguments in such functions are determined by behavioral choices. Conditional on some strong assumptions, nevertheless, some estimates are available. For instance, conditional on a CES parental welfare function over the nutritional status of their children, Behrman (1988) estimates the inequality-productivity tradeoff exhibited by parental preferences to weight productivity more than in the "neutral" Cobb-Douglas case and that parents exhibit "unequal concern" favoring males in rural India when food is relatively scarce. Binswanger (1980), based on experiments with substantial amounts at stake in terms of local wages, finds risk aversion to be inversely associated with income also in rural India.

2.1.2 Budget constraint(s)

Three basic questions about the relevant full-income budget constraints are (1) What determines the share of child full income to which parents have access as the children age? (2) What are the expected future returns to knowledge and skills? (3) What are the options for transferring resources over time?

With regard to (1), generally it is assumed that parents have full access to their children's full income through childhood, but subsequently much more limited or no access to their adult children's full income.⁶ The access to children's full income early in their lives means that there is likely to be an opportunity cost from the parental point of view of sending children to school in terms of foregone child contributions to the family resources in addition to any direct resource costs due to tuition, transportation, books, etc. The expected extent of access to their children's adult full income affects the direct economic incentives for parents to invest in their children's human capital. If there are expected to be differences by birth order or by gender or by personality (e.g., some children appear to care more about their parents than their siblings) in the extent to which parents are expected to have access to children's full income and if parental motivation to invest in their children is in part to obtain future resources from these investments, then the expected differential parental access to such resources results in differential incentives to invest in different children.

With regard to (2), the expected returns to investments in child knowledge and skills are important if (a) the parents are altruistic and parental altruism depends on expected child command over resources (Section 2.1.1) or (b) if the resources generated by children's full income to which parents have access depends on the total full income of the children, which in turn depends in part on the children's knowledge and skills. Note that in forming such expectations about the returns to child knowledge and skills, parents (and, when the children are older, the children themselves) must have expectations about how labor, marriage and possible other markets are going to evolve given various changes expected in the larger economy in which they are situated.⁷ Improved information about these expected returns would be expected to improve investments in education from the parental point of view. Of course, this may lead to increased or lessened investments in education, depending on whether the imperfect information before the improvements under or overestimated the returns to such investments.

With regard to (3), the nature of the options for transferring resources over time seems central if investments in children's education are motivated substantially by parental expectations of obtaining subsequently, such as in their old age, part of the returns on such investments. Presumably investing in children's education for such a purpose is desirable for risk-neutral⁸ parents only if the expected returns on those investments are at least equal to the expected returns on alternative investments. Therefore if returns on other investments become more attractive (e.g., if there are high expected rates of return to holding land or other physical or financial assets because the economy is expected to grow rapidly, if pension options become more attractive), the opportunity costs of investing in children's education increases and the incentives for such investments probably fall.

A related question is what is the nature of capital market access for investments in education. If there is capital market access for investments in education, then

risk-neutral investors will invest in education until the expected rate of return equals the market rate of interest (e.g., Becker, 1967). This will be the optimal investment in education whether the investors (e.g., a children's parents, the children themselves) have considerable assets or not. For children with equal innate ability, investments in education in such a case do not differ depending on family background (though if the returns to such investments differ with innate ability and ability is correlated inter-generationally and with parental characteristics such as parental schooling and income, there may be a positive association between parental characteristics and investments in child schooling nevertheless).

Policies may change investments in education through a number of aspects of the budget constraints: prices of investments in education may be changed through subsidies, rates of return to such investments may be changed through altering the characteristics of educational service related providers, rates of return to such investments may be altered through changing labor or marriage markets, incentives for such investments may be changed through altering capital or insurance markets. Some empirical evidence on these effects is summarized in Section 2.1.4.

2.1.3 Production functions

Production functions give the relations between inputs and output(s), the marginal products of the inputs, and the extent of substitution or complementarity among inputs. The nature of the production function technology may be informative about some components of responses to policies and about some of the pathways for such responses. But in general they do *not* on their own permit inferences about total policy impacts, as noted, because in general all household or individual inputs into production functions adjust in response to a policy change, not just the one toward which the policy is directed. However, production functions often are critical components of structural models, which can be used to investigate the impact of counterfactual models (e.g., see the discussion of Todd & Wolpin, 2006 below). The central production functions for considering the impact of incentives and inputs on education are those for knowledge and skills and marginal labor productivities or wages or whatever other outcomes are affected by adult knowledge and skills.⁹

Knowledge and skills production functions. Knowledge and skills (k_t) in the t th period can be considered to be determined by a production function in which the inputs are all previous and current inputs ("educational experiences" that are flows within each period from t (e_t) back to conception) that affect the accumulation of the stock of knowledge and skills, genetic (and other) initial endowments (e_0) and a stochastic term (u_{tk}) to reflect all other idiosyncratic, and assumed exogenous, learning experiences that affect knowledge and skills:

$$k_t = k^p(e_t, e_{t-1}, \dots, e_0, u_{tk}), \quad (1)$$

where the superscript p refers to the function being a production function. There may be important interactions and nonlinearities in this production function (and in other relevant production functions). For example, individuals with better preschool nutrition may learn more from their school-age experiences (so that the cross-derivative of Eq. (1) with respect to e_1 and e_2 is positive in the case in which the periods are defined to be the three life-cycle stages noted above). This production function also may reflect that some processes are not likely to be reversible at reasonable costs. For example, nutrition early in the life cycle may establish basic patterns of neural development and of other aspects of human development and that may be quite costly or impossible to offset later in life (e.g., Barker, 1992; Engle et al., 2007), which implies that within the three life-cycle version of Eq. (1) e_2 and e_3 can only substitute imperfectly and to a limited extent for e_1 .¹⁰ For each period i , moreover, the relevant experiences are likely to reflect a vector of inputs, including, for example, family-determined inputs such as nutrients or parental time spent helping with school homework (h_i), community inputs (c_i), and inputs that refer to characteristics of other relevant entities such as preschool, school or training programs, or employers (s_i).

Good estimates of Eq. (1) may be informative about questions relating to what are the marginal products of variations in inputs in period i on knowledge and skills in period t . For example, what would be the marginal products of variation in micronutrient supplements for very young children or of stimulation in preschool programs or of class size in school or of training processes in employment. Good estimates of Eq. (1) are difficult to obtain, however, because many of the experiences/inputs reflect behavioral decisions in response in part to persistent unobservables such as e_0 . Very few, if any, data sets, moreover, have information on all the major inputs from conception to adulthood. If only school-age experiences are included in the estimation of Eq. (1) for adult knowledge and skills as in many studies, for example, the coefficient estimates for those school-age experiences are likely to be biased even if they are treated as behaviorally determined because of the excluded pre- and postschool years experiences that, within the model sketched out in Section 2.1.4, are determined in important part by the same or related variables (h, c, s). Estimates of Eq. (1) typically make strong assumptions in order to attempt to estimate the production function parameters depending on what data they use: For example, a specification that uses only contemporaneous data assumes that contemporaneous inputs are perfectly correlated with earlier inputs and that observed inputs are independent of unobserved ones including innate endowments. A value-added specification that conditions on an earlier measure of knowledge and skills that is taken to be a sufficient statistic for unobserved input histories assumes that the production function parameters, including those for unobserved endowments, are not age dependent or that baseline knowledge and skills are orthogonal to unobserved endowments. Todd and Wolpin (2003, 2007) give specific details of the assumptions underlying these and other specifications, as well as some specification tests.

Empirical estimates. There are some, but a relatively small number of estimates of Eq. (1) for developing countries that are based on systematic specification and that attempt to deal with the estimation problems that arise because of unobserved endowments and endogenous input choices. One example follows (and two more are given below in Section 3.2.1).

Estimates for Guatemalan adults aged 25–42 of the production function in Eq. (1) in Behrman, Hoddinott, et al. (2008) with reading comprehension scores and Raven's tests as alternative measures of adult skills that treat all three life-cycle stages as endogenous (using shocks such as an early-life nutritional experiment and natural policy experiments for identification, with a range of tests of robustness and of instrumental variable (IV) diagnostics) find: (1) endogenous schooling attainment has a significant marginal product for reading comprehension scores (but not for Raven's test scores) that is substantially smaller than if schooling attainment is treated as predetermined or the pre- and postschooling experiences are excluded from the specification and (2) endogenous pre- and postschooling experiences (related, respectively, to stunting as of the age of entering school and to tenure in skilled jobs) have significant and fairly substantial marginal products that are larger when they are treated as endogenous (consistent with inverse correlations in unobserved ability and physical endowments as suggested for the United States in Behrman & Rosenzweig, 2004).¹¹ These estimates suggest that focus on schooling alone as in much of the literature, all the more so if schooling is treated as predetermined, leads to a substantial overemphasis on the role of schooling and underemphasis in the role of pre- and postschool experiences in determining what adults know.

Marginal productivities of knowledge and skills on wages and other outcomes. Education is of interest from the point of view of development because presumably it increases productivity in various activities, as represented in a production function such as

$$\gamma_t = \gamma^p(k_t, e_0, \dots, u_{ty}), \quad (2)$$

where γ_t is full labor income or some other measure of the outcome of interest in period t (e.g., resources available to an individual through his/her own work and through other family members), which is posited to depend on the stock of knowledge and skills in period t (k_t), innate endowments (e_0), and an income shock in period t (u_{ty}), as well as possibly other factors (e.g., other components of human capital such as health and work experience, productive physical capital stock). The most common form of such a relation that is used in the literature is a semilog wage function with k_t represented by highest grade or level completed in school (schooling attainment) and actual or potential postschooling work experience. Such a specification gives the marginal product of schooling attainment and work experience on income under the assumptions that schooling attainment and work experience are orthogonal to

(a) genetic endowments (e_0) and (b) any other inputs into the production of knowledge and skills that are not controlled. There is a substantial literature concerned with (a) as related to schooling attainment at least for developed economies, with some ongoing debate about how large the upward “ability” bias¹² due to endowments is and to what extent random measurement error in reported schooling results in OLS biases toward zero (e.g., Behrman & Rosenzweig, 1999; Card, 1999). Relatively little attention has been paid to possible ability bias in coefficient estimates for work experience, though at least one study finds that control for such endowments changes the estimated coefficients for actual work experience substantially (Behrman & Rosenzweig, 2002b). There also is a fairly substantial literature concerned with (b) but focused primarily on school characteristics (“school quality,” e.g., see discussion in Section 2.1.4). Very little of the literature deals with pre- or postschool ages learning experiences beyond actual or more usually potential work experience and in some cases formal training programs (and most of the studies that include these latter indicators of postschooling experience treat them as predetermined and not determined within the life-cycle framework of this chapter with persistent unobserved factors such as ability and motivation).

While wage earnings functions are the primary focus of the economic literature on the impact of education—or at least the impact of schooling, there also is a substantial literature on the impact of some indicator of education (again, almost always schooling attainment in empirical applications) on household production of fertility, health, nutrition, and education for the next generation. Of course for individuals who are at the margin between participating in labor market or household productive activities, the returns to their time are equated in the two types of activities. But that does not mean that the returns to knowledge and skills are equated in the two activities. Moreover, there typically are many individuals who are not engaged in labor markets. For such reasons, it may be desirable to be able to specify how education affects household productivity. There are many estimates of associations between one input into education, schooling attainment once again, and various household outcomes. But very few of the available estimates deal with points (a) and (b) noted in the previous paragraph. Recently, however, there has been a small set of studies that have attempted to deal with (a) in the context of the impact of parental schooling on child schooling in developed countries using samples of identical twins or of adoptees or “natural” policy experiments to attempt to break the correlation between parental schooling attainment and endowments (Behrman & Rosenzweig, 2002a, 2005; Black, Devereux, & Salvanes, 2005; Plug, 2004; Plug & Vijverberg, 2003). These studies find that such controls alter substantially the estimated impacts of parental schooling on child schooling, generally lowering the estimated impact considerably, particularly for mother’s schooling.

From the point of view of the private incentives for investing in education, the question is to what extent more education increases private control over resources or other outcomes that increase the private decision maker’s utility (Section 2.1.1). These

may include own income in labor markets through increasing the probability of employment and increasing wages given employment, income obtained through marriage both by increasing the earning power of the spouse obtained and by increasing one's own bargaining power in intrafamilial bargaining, and household production leading to better own and familial health and nutrition. These possible effects on private incentives do not necessarily coincide entirely with social incentives from a development perspective for several reasons. First, wage rates may differ from marginal productivities due to market or policy failures. Second, unless there are positive complementarities between the education of spouses in household production, the private gain from obtaining a more educated spouse does not increase overall production. Third, likewise, the private gains from being able to bargain better within the household is not necessarily associated with a social gain—though there may be a case if, say, women tend to have preferences that weight more than do those of men investments in the human resources of children and the social interest in the next generation is greater than the private interest.

Further, it should be noted that common estimates of production functions such as in Eq. (2), particularly for adult outcomes, are interpreted as if the marginal products are the same as the total impacts. The marginal impact of schooling in wage relations, for example, often is interpreted to be the rate of return to increasing schooling. But, as noted at the start of this section, such an interpretation is incorrect if there are other inputs that adjust to whatever caused the change in schooling. For example, if there is an exogenous policy change that affects schooling through reducing costs or increasing quality, it also may affect health through household reallocations and learning through work experience. The total impact of the policy change on wages, thus, must include the effects through all of the inputs in the wage production function, not just the marginal product of schooling.

Empirical estimates. There are literally hundreds of studies that report generally positive associations between schooling attainment and (usually potential) experience and wages or other indicators of productivity that might be viewed as representing wage production function estimates (see the reviews in Psacharopoulos, 1985, 1994; Psacharopoulos & Patrinos, 2004). Most of these estimates do not control for the endogenous determination of schooling attainment or work experience or include health, so they may be subject to omitted variable bias due to the failure to control for ability and motivation endowments and any other factors such as physical capital and nonschool forms of human capital that may be correlated with the observed schooling and experience. Under the strong assumptions necessary to give these associations causal interpretations, there appear to be significant private returns to schooling attainment and to work experience (or perhaps general maturity) and, in a small subset of these studies, to health and nutritional status. Under Mincerian assumptions that the only private cost of schooling is the forgone time in the labor market, the average private returns to schooling

implied by these estimates generally are in the 8–15% range, though in some studies the estimates are considerably outside of this range.

There are a limited number of studies that have gone beyond these standard wage estimates in various respects, and that suggest somewhat different implications for investments in education:

- (i) *Including school quality as well as quantity in wage production function estimates.* Some studies have investigated including directly in the wage production function not only schooling attainment and experience, but also school quality. For example, Behrman and Birdsall (1983), first show how not including school quality in standard wage production function estimates is likely to cause upward biases in the estimated impact of schooling attainment (because if school quality is complementary in production with schooling attainment, incentives for investing in schooling attainment are positively correlated with schooling quality). Then they present estimates of wage functions for Brazilian males aged 15–35 with quality represented by the average schooling attainment of teachers in the area in which the sample member was schooled. These estimates are subject to qualifications because they treat schooling attainment as predetermined and they do not consider the possibility that there are other omitted inputs (e.g., other school characteristics, health), but nevertheless are suggestive and imply that (a) the estimated private rate of return to schooling attainment in the quality-inclusive specification is only about half of that in the standard specification without quality, indicating that there is a substantial omitted variable bias if quality is excluded; (b) the social rate of return to school quality exceeds the social rate of return to school quantity, thus implying an equity-productivity tradeoff; and (c) most of the regional variation in standard estimates of rates of return to schooling attainment as well as the differences estimates for migrants versus nonmigrants.¹³
- (ii) *Including cognitive skills, and perhaps health, in wage production function estimates.* Examples include Boissiere, Knight, and Sabot (1985) for samples in urban Kenya and Tanzania; Glewwe (1996) in Ghana; Alderman, Behrman, Ross, and Sabot (1996) in rural Pakistan; and Behrman, Hoddinott, Maluccio, and Martorell (2009) in rural Guatemala. The first three of these studies present some estimates that attempt to control for innate ability through including Raven's scores; the second, third and fourth control for selective wage employment; the last two control for simultaneity for the included human capital variables; and the last explores the impact of including adult health human capital indicators in addition to those for adult knowledge and skills. All four studies report significant coefficient estimates for cognitive skills (only the mathematics components for the Ghanaian estimates) and that the estimated association of schooling with wages declines in all four cases and becomes insignificant in all but the second study when cognitive skills are added to the standard wage

production specification. These estimates are consistent with time in schooling being an input, but not the only input, into the cognitive skills that are rewarded in these labor markets. The preferred estimated impact of cognitive skills on wages is fairly considerable—for example, an increase of 38% in wages in Pakistan and of 43% in Guatemala for a one standard deviation increase in cognitive skills. The fourth one reports that indicators of adult health and strength (e.g., fat-free body mass, adult height) are significantly associated with wage rates, but do not have significant marginal products in the wage production functions if they too are treated as endogenous.

- (iii) *Allowing for heterogeneous schooling attainment impacts.* Sakellariou (2005) summarizes estimates from 26 countries on the possibility of heterogeneous returns across the earnings distribution by examining the pattern of returns across quantiles to schooling attainment.¹⁴ This compilation of the available evidence suggests that there is a pattern with high-income countries associated with increasing marginal products of schooling by quantile and low-income countries associated with a decreasing pattern of marginal products of schooling for higher quantiles. Under the assumption that such estimates are unbiased, this pattern suggests that with development the rates of return to higher schooling attainment may increase, perhaps due to higher demands for more skilled labor, which would seem to increase the incentives for investing in higher levels of schooling—but *vice versa* in developing countries. Under the alternative assumption that most of the differences in the quantile estimates is due to omitted ability or school quality bias, the pattern for the developed countries suggests that unobserved ability and/or school quality is a complement with schooling attainment in the production of wages, but a substitute for the developing countries (an interpretation apparently dating back at least to Mwabu & Schultz, 1996). Other alternatives, of course, are that there are other omitted variables such as pre- and postschooling experiences that are correlated with schooling differentially in developing versus developed economies (e.g., the nature of occupations are such that postschooling experience tend to reinforce what is learned in schooling in developed, but not in developing countries).
- (iv) *Treating education as endogenous.* As noted, two of the studies mentioned above that use adult cognitive achievement to represent education in wage production functions, the ones for Pakistan and Guatemala, treat cognitive skills as endogenous using IV estimates. *A priori* such a procedure may affect the point estimates in either direction, depending on the signs of the correlations with unobserved variables and the extent to which the measures of education have measurement error that is controlled by the IV estimates. In both of these cases, the IV estimates indicate much stronger impacts on wages of cognitive skills than do OLS estimates; therefore, if the IV estimates are the preferred estimates, the OLS estimates are misleading indicators of the impact of cognitive skills on wages (by about 60%

and 30%, respectively, in the two cases). As also noted, most of the studies that use schooling attainment as the measure of education that affects wages treat schooling attainment as if it is predetermined and orthogonal to any relevant unobserved variables. But there are some exceptions. Duflo (2001), for example, in a well-known study (discussed further below in Section 3.2.2) of the “natural experiment” of rapid school expansion in Indonesia, uses that school expansion for IV estimates in wage functions. Depending on the other controls that she includes, her estimates imply that OLS estimates range from 26% below to 15% above the IV estimates.

- (v) *Including international migration.* As noted, the expected returns to education include returns in domestic and in international markets. Stark, Helmenstein, and Prskawetz (1998) suggest that expectations for international labor markets may increase education enough so that those who end up not migrating are more educated than they would be in the absence of international migration because the possibility of migrating might create greater incentives even for those who do not end up migrating. Rosenzweig (2006b) presents some related estimates using aggregate data. He finds that the patterns of student flows across countries are consistent with the hypothesis that students acquire schooling abroad to obtain jobs in the host countries: there are larger *per capita* numbers of foreign students in the United States from lower skill-price countries than from high skill-price countries and host countries with higher skill prices attract the most foreign students, controlling for the quality of tertiary schooling in sending and receiving countries.

Summary and implications.

- The adult stock of education generally seems to have significant private marginal products on wages, thereby creating incentives for private investments in education.
- The adult stock of education seems to be determined by more than schooling attainment—probably pre- and postschool experiences and school quality in addition to schooling attainment—with the result that empirical wage production function studies that limit themselves to schooling attainment are likely to be misleading regarding the incentives for investments in education, probably overstating those for time in school and understating those for school quality and for pre- and postschooling investments in schooling.
- Most studies in the literature permit characterization of only the private returns and incentives to investments in schooling, not the differences between the private and social returns and incentives that are germane for the efficiency policy motive discussed above.¹⁵ Therefore, though they may provide guidance for the distributional (e.g., antipoverty) policy motive, they do not provide much insight into efficiency concerns.
- There is some suggestion that the context matters (also see Section 2.1.4)—the extent of rewards for the learning capacities that at least broad education creates, as seems plausible *a priori* and long has been emphasized by scholars such as Rosenzweig (1995), Schultz (1975), and Welch (1970).

- The estimates often seem sensitive to specifications and estimation procedures and thus are conditional on the underlying assumptions that in many cases are not spelled out clearly nor tested for robustness adequately. Some *a priori* interesting questions, such as the extent of substitution and irreversibilities among experiences in different life-cycle stages in producing adult knowledge or wages in developing countries, are unexplored. There would seem to be substantial possibilities for advancing knowledge about these key production function relations through further more systematic studies with careful attention to the robustness of the results to alternative underlying assumptions.

2.1.4 Constrained maximization, dynamic reduced-form demand relations related to inputs into and incentives for investing in education, estimates and policy implications

As noted above, the estimation of some of the parameters of objective functions and of production functions may provide useful insights into aspects of the demand side for education, the underlying production function technologies, and some of the pathways for impacts of incentives. Obtaining such estimates, conditional on the underlying models of behavior, requires proceeding empirically in a manner consistent with those models. (As discussed further below in Section A.2, experiments generally do not permit the identification of these parameters.) In addition to such estimates of parameters in the structural objective and production functions, moreover, it is desirable to obtain reduced-form demand relations that capture, again, within the model specified, the *total* response of education and education inputs to changed incentives.

Assume a very stylized model in which the “dynasty” (first the parents through intrahousehold bargaining between themselves and perhaps other relatives, then the children themselves increasingly as they age into youth though with intrahousehold bargaining with their parents and other relatives and into adulthood usually with a spouse that involves further bargaining) make decisions so as to maximize a welfare function that includes y_3 for each individual. This welfare function is maximized sequentially subject to the constraints at each life-cycle stage related to relevant current and expected production functions, resources allocated to this individual, community characteristics including community services and markets that affect household decisions, and stochastic factors. Among these expectations, of course, are the expected returns to education, both in domestic and in international markets.

It is straightforward to write the reduced-form demand relations for the educational inputs selected by the household (h_i), education-related experiences (e_i), and end-of-stage human capital stocks (k_i) for each of the three life-cycle stages. But such expressions are somewhat tedious and repetitious in spirit, if not in details. Therefore following are these expressions only for the school-age stage, stage 2:

$$h_2 = h_2^d(i_1, f_1, c_1, s_1, e_0, i_{12}^e, f_{13}^e, f_{12}^e, f_{13}^e, c_{12}^e, c_{13}^e, s_{12}^e, s_{13}^e, i_2, f_2, c_2, s_2, i_{23}^e, f_{23}^e, c_{23}^e, s_{23}^e, u_1, u_2, u_{12}^e, u_{13}^e, u_{23}^e), \quad (3a)$$

$$e_2 = e_2^d(i_1, f_1, c_1, s_1, e_0, i_{12}^e, i_{13}^e, f_{12}^e, f_{13}^e, c_{12}^e, c_{13}^e, s_{12}^e, s_{13}^e, i_2, f_2, c_2, s_2, i_{23}^e, f_{23}^e, c_{23}^e, s_{23}^e, u_1, u_2, u_{12}^e, u_{13}^e, u_{23}^e), \quad (3b)$$

and

$$k_2 = k_2^d(i_1, f_1, c_1, s_1, e_0, i_{12}^e, i_{13}^e, f_{12}^e, f_{13}^e, c_{12}^e, c_{13}^e, s_{12}^e, s_{13}^e, i_2, f_2, c_2, s_2, i_{23}^e, f_{23}^e, c_{23}^e, s_{23}^e, u_1, u_2, u_{12}^e, u_{13}^e, u_{23}^e), \quad (3c)$$

where the superscript d refers to reduced-form demand relations, i_i includes individual exogenous (to the household) factors, f_i includes family exogenous factors, c_i includes community exogenous factors (including prices and policies), s_i includes human capital services provider exogenous factors, u_i is a stochastic disturbance term, the subscripts for the variables with one subscript refer to the three life-cycle stages, and the variables with a superscript e are the expected values for that variable held during the stage represented by the first subscript for the stage indicated by the second subscript. These expressions have many right-side terms because they include the actual and expected future values during the first period (that determined the end of the first period human capital stock k_1) and expected values of from the perspective of the second period. With added assumptions, the end of first period human capital stock (k_1) is a sufficient statistic for all the right-side variables that determined outcomes in the first period and therefore can be substituted into Eq. (3a)-(3c) for all of those variables ($i_1, f_1, c_1, s_1, e_0, i_{12}^e, i_{13}^e, f_{12}^e, f_{13}^e, c_{12}^e, c_{13}^e, s_{12}^e, s_{13}^e, u_1$) in conditional demand relations (with a superscript c instead of d).

Good estimates of Eq. (3a) (and similar relations for other life-cycle stages) permit answering a number of important questions about how changed incentives or resources alter inputs into educational production functions. Good estimates of Eq. (3b) (and similar relations for other life-cycle stages) permit answering questions about how changed incentives or resources alter educational experiences in different life-cycle stages. Good estimates of Eq. (3c) (and similar relations for other life-cycle stages) permit answering questions about how changed incentives or resources alter stocks of knowledge and skills at the end of the different life-cycle stages.

Various types of data limitations, however, are likely to make obtaining good estimates of these relations challenging: (1) absence of longitudinal data for sufficiently long periods of time; (2) absence of information on expectations held earlier on subsequent outcomes including returns to various human capital investments; and (3) absence of information on some right-side variables including genetic endowments but also unobserved aspects of communities and service providers that may be correlated with administrative decision makers' placement of programs and provision of inputs into public service providers (Rosenzweig & Wolpin, 1986).¹⁶ For these

reasons, some right-side variables that often are treated as predetermined are likely to be correlated with unobservables that are in the disturbance term so that some estimation procedures may be needed to break that correlation (e.g., fixed effects, instrumental variables).

In some respects, the conditional demand relations mentioned at the end of the paragraph with Eqs. (3a)–(3c) have promise of requiring less longitudinal data, but they do require stronger assumptions for the end of the previous period human capital stocks to be a sufficient statistic for past relevant decisions and in estimation these end of the previous period human capital stocks should be treated as endogenous because they are likely to be correlated with other unobserved right-side variables that are in the disturbance term. Potential instruments for IV estimation are the period-specific idiosyncratic parts of the variables that they replace (the parts of these variables that are systematically related over time do not provide good instruments because they are correlated with the systematic parts of these variables that are on the right-side of the conditional demand relations parallel to Eqs. 3a–3c).

Reduced-form demand relation estimates and policy implications. Changes in any of the right-side variables in the reduced-form demand relations (e.g., Eq. 3b and the parallel relations for the other life-cycle stages) change the incentives for investing in education. Some of these right-side variables affect incentives for investing in education through directly affecting other dimensions of household constraints—which is the subject of this section (others work through changing supply-side characteristics, which are considered below in Section 3).

Expected returns to human capital investments Standard models of human capital investments imply that increased expected returns in human capital investments *ceteris paribus* increase human capital investments either because they directly increase the investors' welfare because of altruism (Section 2.1.1) and/or indirectly increase investors' welfare through the budget constraint (Section 2.1.2). But there is relatively little direct evidence on these effects because most data sets do not have good representations of expected returns to human capital investments. However two studies on India do provide some support for the importance of such effects.

Foster and Rosenzweig (1996) use panel and time series data describing the Indian green-revolution period to assess the effects of: exogenous technical change on schooling returns, schooling on profitability of technical change, and technical change (as well as changing school availability) on household schooling investments. The results indicate that the returns to primary schooling increased during the period of rapid technical progress due to the green revolution with its differential exogenous impact in different rural areas of India due to different agronomic conditions (which is key to identifying these impacts), particularly in areas with the highest growth rates. The increased

growth rates induced significant additional private investment in schooling, net of changes in wealth, wages, and the availability of schools.

Munshi and Rosenzweig (2006) investigate roles of an important traditional institution—the caste system—in shaping schooling investment and career choices in Mumbai using survey data that they assembled on school enrollment and income over 20 years. They find that male working-class lower-caste networks continued to channel boys into local language schools that lead to traditional occupations, despite substantially increased returns to nontraditional white-collar occupations in the 1990s. Thus the traditional caste employment network precluded or at least discouraged schooling investments in boys being responsive to changing labor market opportunities. In contrast, schooling investments in lower-caste girls, who historically had low schooling and low labor market participation rates and so did not benefit from the caste network, switched rapidly to English schools that permitted them to exploit the opportunities that became available in the rapidly changing economy due to globalization. This is a very interesting result because of the interaction between the caste system and gender divisions in labor and schooling investments leading to schooling investments in girls, but not in boys, responding to expected higher returns for English-language schooling.

Income and capital markets Many studies report associations of family income or wealth with investments in child schooling, but often the reported associations are fairly weak (see the review of 42 studies related to schooling in Behrman & Knowles, 1999; Paxson & Schady, 2007 present some results for preschool education in Ecuador and give citations to other studies, though they emphasize that parenting style is likely to be important, not just resources). However as also noted in the Behrman and Knowles review, most of these studies use fairly short-run measures of expenditures and income from cross-sectional data for a period of time that is not necessarily one in which households are making marginal educational investment decisions. Therefore the measures used may not represent well the household resource constraint when the relevant decisions are made. Behrman and Knowles present estimates for Viet Nam that use predicted income based on asset data to represent longer-run income, control for censoring, and represent additional channels through which there may be effects beyond just the number of completed grades. Their estimates indicate that the income associations for child school success are fairly considerable—for example, the income elasticity of completed grades is five times the median in the 42 earlier studies. This association, moreover, is strongest *not* for completed grades on which most of the previous literature has focused, but for grades completed successfully per year of school. There are some gender differences, with the most important one being a smaller income association with grades completed per year of schooling for males than for females, which implies that schooling of females is treated as more of a “luxury” (less of a “necessity”) than is schooling of males. Therefore, much of the previous literature

may underestimate the true associations of household income with child school success and overstate true intergenerational social mobility and equality of opportunity.

However, even if the true associations of educational investments with longer-run family income are larger than implied by much of the previous literature, that does not mean that increases in income will induce increases in schooling or other forms of investment in human capital. The large associations with income may only mean that income is correlated with other factors, such as intergenerationally correlated ability or motivation. Within standard models of human capital investments such as Becker's (1967) Woytinsky Lecture, as noted above, if all households have equal access to capital markets in which human capital investments can be financed at the market rate of interest, then investments in the education of all individuals will occur until the expected rate of return on these investments is equal to the market rate of interest, independent of household income levels—and changes in income do not change the investments. There still will be associations of the educational investments with income if income is correlated with intergenerationally correlated endowments and if those endowments are correlated with the educational investments in producing the outcome of interest (e.g., wages). In contrast, if capital markets for financing human capital investments are imperfect or nonexistent so that households have to self-finance educational investments or if access to capital markets is positively associated with household income, then at least part of the association of educational investments with income may imply that income increases would increase such investments (though part may still be due to intergenerationally correlated endowments).

Limited or no access to capital markets also is likely to mean that income shocks disrupt or terminate educational investments. If capital markets constrain choices of poorer households more than those of better-off households, the nature of capital market access might underlie part of the pattern noted above for Viet Nam that grades completed successfully per year of school are inversely associated with income. Many poor households in developing countries, particularly those in rural areas with weather fluctuations but also those working in other activities such as commerce or construction, experience considerable income fluctuations. There are a set of empirical studies that suggest, however, that such households generally are able to insure themselves fairly well against at least household idiosyncratic income shocks. A common finding for rural households in developing countries is that consumption elasticities with respect to permanent income are near unity and that propensities to save out of transitory income are much larger than propensities to save out of permanent income (Jacoby & Skoufias, 1997; Paxson, 1992; Townsend, 1994; Wolpin, 1982).

One mechanism through which households may smooth consumption, however, is through altering the time allocation between educational investments and work. Jacoby and Skoufias (1997) use monthly longitudinal rainfall data over two and a half years from the six International Crop Research Institute for the Semi-Arid Tropics

(ICRISAT) Village-Level Studies (VLS) communities in arid semitropical India to investigate the extent to which child school enrollments are insulated from exogenous full-income shocks, as well as how important is the distinction among anticipated changes based on household attributes, unanticipated household idiosyncratic shocks, and unanticipated community idiosyncratic shocks. Their estimates suggest that anticipated income changes do not have much impact on child schooling enrollments but that household idiosyncratic transitory shocks do affect school enrollments, particularly for poorer households. These poorer households apparently use increased child labor and reduced school time to help smooth consumption in the face of adverse household idiosyncratic income shocks, which their estimates suggest disadvantages the education of these children some, though not by a large amount. Thus this study suggests that the large cross-sectional association between longer-run income and the schooling enrollment component of educational investments only in small part is likely to reflect differential mechanisms for smoothing income shocks across households with different incomes. The Foster and Rosenzweig (1996) and Munshi and Rosenzweig (2006) studies summarized above also suggest that budget constraints are not likely to be major limiting factors: the former finds no significant wealth effects and the latter finds that the poorest, low-caste families responded most quickly to the changes in the returns to investing in English-language schooling (at least for girls).

Parental schooling Parental schooling *a priori* may be an input into child education through increasing learning at home particularly during the preschool and school-age life-cycle stages. A number of studies report positive associations between parental schooling and dimensions of child education (e.g., for the preschool life-cycle stage in the Philippines in Ghuman, Behrman, Borja, Gultiano, & King, 2005, in Ecuador in Paxson & Schady, 2007; for the school-age life-cycle stage in Ghana in Glewwe & Jacoby, 1994; for rural Pakistan in Alderman et al., 1996, in South Africa in Case & Deaton, 1999, in north India in Drèze & Kingdon, 2001).¹⁷ Recent studies for developed countries as noted in Section 2.1.3, however, bring into question the extent to which parental schooling, particularly maternal schooling, in such estimates is representing the causal effects of parental schooling on child education or in substantial part intergenerational ability and motivation endowments by using “natural experiments” of maternal twins or policies or adoption data to control for such endowments (e.g., Behrman & Rosenzweig, 2002a, 2005; Black et al., 2005; Plug, 2004; Plug & Vijverberg, 2003). Whether parental schooling has associations that are robust to controlling for possible omitted variable bias has not been explored much in the developing country literature. But a few studies suggest the need for further explorations. Studies using Nicaraguan adult sister data to control for common family background of adult siblings, for instance, finds much different (generally weaker) associations with within-adult-sister estimates, though without control for random measurement error

that may attenuate such estimates (Behrman & Wolfe, 1984, 1987a,b). A preliminary study using a nutritional experiment and “natural experiments” from policy changes as instruments to identify maternal schooling effects on children in Guatemala suggest that OLS estimates may underestimate slightly the impact of maternal schooling on child grades of schooling but that including maternal schooling as the only measure of maternal human capital appears to bias upward the impact of schooling because it proxies in part for mothers’ biological human capital (Behrman, Murphy, Quisumbing, Ramakrishna, & Young, 2008). These results suggest the value of further explorations of what causal role parental education plays as inputs into child education in developing countries.

Reductions in private costs of education *Conditional cash transfers.* In recent years, a number of countries in Latin America and elsewhere have implemented large-scale conditional cash transfer (CCT) programs that provide monetary incentives to poor families for investing in the human capital of their children. The *Oportunidades* program (formerly called *PROGRESA*¹⁸) has been operating since 1997 in small (with populations less than 2500) rural Mexican communities and since has spread to urban areas and covers over 30 million Mexicans. It is one of the earlier CCT programs and probably the best-known and most studied because of a strong commitment to systematic data collection and analysis (Behrman, 2007). Prior to the program initiation an initial evaluation sample was established with about 25,000 households residing in 506 communities, with random assignment of 326 of those communities to treatment in 1998 and 180 of those communities as controls, though subsequently they were incorporated into the program.¹⁹ The initial evaluation of the program, based on data from the first 2 years of the program (1998–2000), demonstrated significant short-run impacts in improving some important dimensions of preschool and school-age education. Using difference-in-difference estimators, for example, Behrman and Hoddinott (2005) report significant increases in early-life nutritional status equal to about one sixth of mean growth for children aged 12–36 months and a significantly reduction in stunting, which is associated with neural and cognitive development in the nutrition literature (also see Rivera, Sotres-Alvarez, Habicht, Shamah, & Villalpando, 2004), and which, based on other estimates, could increase lifetime earnings by 2.9%.²⁰ Difference-in-difference estimates of schooling enrollment probits and transition matrices imply increases in progression rates, reductions in dropout rates, increases in re-entry rates and long-run (if the estimated short-run relations are sustained through the school life of a student) increases in schooling attainment of about 0.7 grades (Behrman, Sengupta, & Todd, 2005; Schultz, 2004).²¹ In 2003 there was additional data collection on the original evaluation sample of 506 communities (by then all incorporated into the program) and on a new control sample obtained through matching among communities that had not yet been incorporated into the program. Estimates based on comparing the original randomly assigned treatment and control groups who had differential

exposure to the nutritional supplement during the critical early years of life even though the control group later was incorporated into the program reveals that the improved nutrition during those critical years due to the program is associated with improved initial schooling performance, consistent with the literature on the impact of early-life nutrition on cognitive development (Behrman, Parker, & Todd, 2009b). Similar estimates for schooling attainment imply long-run impacts similar to those extrapolated from the results found for the first year or two of the program with the experimental data (Behrman, Parker, & Todd, 2009a). Matching propensity estimates between the original 1998 treatment group and the 2003 group 5.5 years after the program began again show positive impacts on schooling, but few effects on cognitive achievement (though information is not available to control for preprogram conditions that might affect cognitive achievement). Benefit-cost simulations based on these estimates show that program benefits are several times higher than program costs in terms of resources under nearly all scenarios, with the exception of a very high discount rate and a low estimated return to education. Overall, then, even if the program were considered as only a schooling investment program, the overall benefits would seem to significantly outweigh the resource costs.²²

Probably the most interesting studies of *PROGRESA* and education utilize structural models to explore counterfactual policies. Todd and Wolpin (2006), for example, develop a dynamic structural model of household behaviors relating to schooling and other behaviors (e.g., fertility) based on the baseline data.²³ They then test their model against the subsequent experimental results and find that their model performs relatively well, which gives more confidence in the validity of using the model for other simulations. They then use the model to simulate the impact of counterfactual policies (e.g., what would have happened had the program focused on critical postprimary school grades with the same governmental budget constraint rather than including grades 3–6 for which preprogram enrollment rates already were high? What would have happened if children were exposed to the program all their lives rather than just for the 2 years of experimental data?). This is a good example of the potential gains of using economic modeling to go beyond the black box assessment of actual policies to consider policies that have not yet been implemented—at the costs, of course, of the assumptions necessary to develop the structural model.

There remain a number of issues for further examination of the educational impacts of even the already much-studied *PROGRESA/Oportunidades* program. For example, what are the long-run effects on labor market and marriage market outcomes? What underlies the very limited evidence on impacts on cognitive achievement and what are the implications of such evidence? How did (should?) the educational supply side respond best to the changes in demand? And, of course, as for other empirical studies, what are the implications of using such programs in other market and policy contexts? But the effort to date by the Mexican government to collect information on a

large-scale program that permits systematic evaluation, despite some shortcomings, is an outstanding example that merits much wider emulation.

Educational vouchers. Vouchers typically cover part of all of the direct monetary costs (price) of enrolling in formal educational institutions, though typically not other costs (e.g., the opportunity costs of time spent in school and not working) as might be covered by some scholarship or CCT programs (and, indeed, was a central part of the calculation for the benefit schedule in the *PROGRESA* program discussed above). Because vouchers are portable, they hopefully increase incentives for more efficient provision of educational services through increasing competition among potential providers. Vouchers could be offered for any formal educational program, but in fact have largely been used for formal schooling.

Chile has been a leader in market-oriented schooling innovations and has had universal primary and secondary voucher programs for over a quarter of a century. When the voucher program was first introduced in 1981, the percentage of students enrolled in private-subsidized schools increased rapidly from about 15% to over 30%, with a corresponding decline in public school attendance. Currently schooling attendance in Chile is distributed among three types of schools: municipal or public schools, with about half the enrollment; private-subsidized schools that accept vouchers and in some cases impose additional tuition charges, with about 40% of the enrollment; and private nonsubsidized schools with the remainder (the last group includes the premiere private schools that are financed solely by tuition, which on average in these schools amounts to about four times the *per capita* voucher payment).

Most of the questions that have been addressed in research to date on the Chilean voucher system, however, have been somewhat limited by the nature of the data available. Most of this research (e.g., Contreras, 2001; Sapelli & Vial, 2002) analyzes the relationship between test scores and attendance at different types of schools, using test score data aggregated to the school level. With such data, there are multiple selection problems, primarily that the types of children attending each school are self-selected and that test scores are unavailable for children who are not attending school. Hsieh and Urquiola (2006) consider whether the voucher program resulted in better school performance, exploiting the fact that the impact of the voucher program and consequently private school enrollment was different in different communities. They argue that communities that experienced a greater increase in private school enrollment should have experienced better schooling outcomes compared to other communities. Using community-level data, they find that average test scores did not rise any faster in communities where the private sector enrollment expanded more, and that average repetition and grade-for-age actually worsened in such areas relative to other communities. They also examine the relative performance of Chilean students in international tests in science and mathematics (TIMSS), in which Chile participated in 1970 and 1999. The comparison shows that despite nearly two decades under an unrestricted

school choice regime, the performance of the median Chilean student had not improved relative to that of the median student in other countries. McEwan and Carnoy (2000) estimate a relationship between average school test scores and a measure of competition that they define as the percentage of total enrollment in private schools at the community level for 1988–1996. They include fixed effects to control for time-invariant unobserved school characteristics. They conclude that public schools that faced more competition had lower average test scores, in part because of the mobility of the better students to private schools and that nonreligious voucher schools are no more effective than public schools, whereas Catholic voucher schools are more effective. They document that average per pupil expenditure is lower in private schools than in public schools. Auguste and Valenzuela (2003) also analyze the relationship between test scores and competition, using an IV approach that uses community population and distance to the closest city as instruments for competition. They find positive effects of competition on average test scores. McEwan (2001) examines the effects of attendance at a public or private voucher school on test score outcomes, using individual-level data for eighth graders and using a control function approach to account for selectivity into type of school. He finds no important differences in achievement between public and nonreligious voucher schools, but that Catholic voucher schools exhibit a small advantage in test scores over most public schools. He does not examine whether public schools might have improved due to increased competition from voucher schools. Using fourth-grade achievement test scores, averaged at the school level, and a similar estimation strategy, Bravo, Contreras, and Sanhueza (1999) and Mizala and Romaguera (1999) also examine the gap in test score performance between municipal and subsidized private schools and conclude that the test score gap is small or nonexistent after controlling for geographic location and socioeconomic background. Tokman (2002) examines the relationship between primary school test scores and type of school, allowing the impact of attending private schools to differ by average socioeconomic status (using school-level data). Her results indicate that public schools are neither uniformly worse nor uniformly better than private schools. Rather, public schools appear to be relatively more effective for students from disadvantaged family backgrounds, which is a finding similar to that of Neal (1997) for United States Catholic schools. All of the papers cited above examine how type of school influences test score outcomes, but do not consider important questions on how introduction of the voucher program alters school attendance choices, completed schooling attainment, and labor market outcomes, nor do they control for the selectivity in which individuals attend school and take the tests.²⁴

Bravo, Mukhopadhyay, and Todd (2008) advance beyond the previous literature on vouchers in Chile with a research strategy and data that permits them to answer some of these questions and that results in a more positive assessment of the system. They develop and estimate a dynamic model of schooling and work decisions using

data from the 2002 Historia Laboral y Seguridad Social and the 2004 Enuesta Protección Social (EPS) surveys. The dataset includes rich demographic information as well as contemporaneous and retrospective schooling and work information covering a 35-year time frame. Some individuals in the sample completed their schooling before the voucher program was introduced, while others had the option of using the vouchers over part or all of their schooling. The impacts of the voucher program are identified from the differences in the schooling and work choices made and wage returns received by individuals differentially exposed to the program. Simulations based on the estimated dynamic model indicate that the school voucher program induced individuals affected by the program to attend private-subsidized schools at a higher rate, achieve higher educational attainment, receive higher wages, and participate more in the labor force. Returns to both public and private education increased after the introduction of vouchers. An examination of distributional effects shows that the voucher program benefited individuals from both poor and nonpoor backgrounds, but that the nonpoor experienced greater benefits.

While most of the studies on vouchers in developing countries have focused on Chile because of its early leadership instituting a large-scale program, there also are a few studies on related programs elsewhere. The most prominent of these studies probably is Angrist and Lavy (2002), which investigates the impact in selected Colombian cities of the *Programa de Ampliación de Cobertura de la Educación Secundaria* (PACES) voucher program. This program was introduced in 1991 and was targeted at poor secondary school-age children with the goal of enabling them to attend private secondary schools. By 1996, this program covered about 100,000 students. Prior to the program, enrollment in primary school (grades 1–5) was almost universal but the enrollment rate in secondary school (grades 6–11) was only 73% overall and 55% for students from the poorest quintile of families. The vouchers were designed to cover about one-half of the cost of private secondary schools and could be renewed as long as students maintained satisfactory academic performance. Evaluation of the PACES program was facilitated by the fact that vouchers were initially awarded by lottery in municipalities in which the number of eligible students applying for vouchers exceeded the estimated shortfall in the number of spaces available in public schools. Previous year enrollment was used by local authorities to decide how many new vouchers to make available in a given school year. The excess demand in some municipalities made it possible to estimate the effect of the vouchers on the target population under the implicit assumption that the excess demand for vouchers occurred randomly across municipalities (and therefore was not associated with other municipality-level factors that might affect schooling). If, for example, excess demand for vouchers more likely occurred where recent expectations were higher regarding the returns to schooling because of greater economic activity than elsewhere, the estimated impacts may not be applicable to other municipalities with lower expectations regarding the returns to schooling. Unfortunately, the

study does not provide information with which to assess whether the “excess-demand” municipalities were similar to or different from other municipalities. Under this assumption, Angrist and Lavy (2002) did not find any significant impact on enrollment. However, they did find that lottery winners were 15% more likely to attend a private school. After 3 years, lottery winners had completed 0.12–0.16 more grades of schooling (primarily due to lower repetition rates) and were about 10% points more likely to have completed the eighth grade. Although the program had no effect on dropout rates, lottery winners scored 0.2 standard deviations higher on standardized tests. In addition, lottery winners reported working 1.2 h per week less than nonwinners and were less likely to be either married or cohabiting as teenagers (however, this last difference affected only about 1% of the sample). Angrist, Bettinger, and Kremer (2006) use administrative records on registration and test scores on a centralized college entrance examination and find that the lottery program increased secondary school completion rates by 15–20% and that the program increased test scores by about 0.20 standard deviations, with somewhat larger increases for boys (who have lower scores than girls in this population), particularly in mathematics. Angrist et al. do not include a formal cost-benefit analysis, concluding that the benefits would clearly exceed the costs using any plausible discount rate. Knowles and Behrman (2005) build upon the Angrist et al. study and calculate what they characterize as conservative benefit-cost ratios of 3.8 with a discount rate of 3%, 2.7 with a discount rate of 5%, and 1.4 with a discount rate of 10% (they also present much higher estimates with less conservative assumptions regarding the impacts of accelerating progress through school and the lifetime labor market value of increased test score performance). While the lottery deals relatively well with selection and omitted variable problems among those in the lottery, it does not permit confident inferences about effects on the larger population of which the lottery participants are a selected subset. For such purposes, population-based information would seem to be required.

Subsidies for other private educational costs. Government-subsidized school meals have been provided in a number of countries in efforts to increase school enrolment, attendance, and performance (World Food Program, 2002).²⁵ The Drèze and Kingdon (2001) study on four states in north India discussed in Section 3.2.2 reports an association between providing midday meals and increased primary school for girls but not for boys. Vermeersch and Kremer (2004) conducted a randomized evaluation of the impact of school meals on participation in western Kenyan preschools, and found that preschool participation was 30% greater in the 25 Kenyan preschools in which free breakfasts were provided than in the 25 comparison preschools. In preschools in which teachers were relatively well trained prior to the program, the meals program led to higher test scores on academic tests (0.4 of a standard deviation). But there were no effects on tests of general cognitive skills, suggesting that the school meals program

did not improve children's nutritional status and that the academic test score increases probably were due to more time spent in school.

Kremer, Moulin, and Namunyu (2002) evaluated a randomized trial in rural western Kenya to assess a program in which a NGO provided uniforms and textbooks and built classrooms for seven schools randomly selected from a pool of 14 poorly performing schools. Dropout rates fell considerably in the seven schools selected for participation, and after 5 years pupils in those schools had completed about 15% more grades of schooling. In addition, many students from nearby schools transferred into program schools, raising class size by 50%. Apparently students and parents were willing to trade off much larger class sizes for the benefit of free uniforms, textbooks, and improved classrooms. The authors argue that the main reason for the increase in schooling likely is the financial benefit of free uniforms. A randomized trial of textbook provision in western Kenya (Section 3.2.2) showed almost no impact of textbooks on schooling attainment, and the first new classrooms were not built until the second year of the program, while dropout rates fell dramatically in the first year. Anticipation of later classroom construction may have influenced these results, but the authors doubt it, because effects were present for students in the upper grades who would have finished school by the time the classrooms were built. Of course part of the problem in making inferences is that there were a bundled package of benefits, as in some other experiments (e.g., the Mexican *PROGRESA* CCT program discussed above, the investigation of teacher absence in Rajasthan, India discussed below in Section 3.2.2), rather than a design that would permit identifying the impacts of the various program components.

Preschool experiences and subsequent education The framework at the start of this section emphasizes that what happens in the schooling ages and later is likely to be conditioned on preschool education-related experience. The inputs into education obtained through schooling, for example, importantly include children and their pre-school levels of cognitive and physical development. Improving those inputs is likely to increase the productivity of subsequent education and thereby the incentives to invest in such education. Heckman (2006), for example, has suggested that for poor children the rates of return to improved preschool development may exceed significantly those for investments later in the life cycle.

Galiani, Gertler, and Schargrodsky (2009) investigate the empirical case for universal preprimary schooling by examining the effect of a large expansion of universal preprimary education on subsequent primary school performance in Argentina. They exploit the variation introduced by the program's expansion over time that generated differences in exposure by cohorts and municipalities to estimate program impacts. They estimate that 1 year of preprimary school increases average third-grade test scores by 8% of a mean or

by 23% of the standard deviation of the distribution of test scores and positively affects third-grade students' self-control as measured by behaviors such as attention, effort, class participation, and discipline. They test for alternative causes by a falsification test using older students, for whom they find no effects and by examining but rejecting the possibility of selective migration.

For developing country contexts, a body of literature, some of it outside economics, has explored the relationship between preschool nutritional status and the education of school-age children and adolescents. Malnutrition is widespread amongst children in developing countries (United Nations ACC/SCN, 2000). Compared to better-nourished children, malnourished children: score lower on tests of cognitive functioning; have poorer psychomotor development and fine motor skills; have lower activity levels; interact with others less frequently; have lower enrollment rates; and complete fewer grades of schooling (Alderman, Behrman, Lavy, & Menon, 2001; Alderman, Hoddinott, & Kinsey, 2006; Ghuman et al., 2005; Glewwe & Jacoby, 1994; Glewwe, Jacoby, & King, 2000; Glewwe & King, 2001; Grantham-McGregor, Fernald, & Sethuraman, 1999a,b; Grantham-McGregor, Walker, Chang, & Powell, 1997; Johnston, Low, de Baessa, & MacVean, 1987; Lasky, Klein, Yarborough, Engle, Lechtig, et al., 1981; Paxson & Schady, 2007). Some of these studies have used "natural experiments" such as weather fluctuations (Alderman, Hoddinott, & Kinsey) or market fluctuations (Alderman et al.) to control for the endogeneity of preschool nutritional status. The adverse effect of undernutrition on fine motor control, for example, suggests that physical tasks associated with attending school, such as learning to hold a pencil, are more difficult for undernourished children. It is believed that all these effects reflect, in part, the biological pathways through which undernutrition affects neurological development since controlled experiments with animals suggest that early-life undernutrition results in irreversible damage to brain development (Yaqub, 2002). Behrman, Alderman, and Hoddinott (2004) attempt to synthesize these and many other studies and estimate, subject to a number of qualifications, benefit-to-cost ratios for interventions to improve nutrition early in life in poorly nourished developing country populations to be considerably greater than one, in substantial part because of the estimated effects on cognitive development and schooling attendance that translate into higher adult productivity.

Two recent studies contribute to this literature in part by investigating a longer part of the life span than most previous studies.

Maccini and Yang (2009) examine the effect of early-life rainfall deviations from local mean rainfall on the health, schooling, and socioeconomic outcomes of Indonesian adults in 2000 who were born between 1953 and 1974. They have linked historical rainfall for each individual's birth year and birth location (tying births to agricultural seasons and using deviations from long-run trends in rainfall) with current adult outcomes (again using deviations from long-run trends) from the 2000 wave of the

Indonesia Family Life Survey. They report that higher early-life rainfall has large positive effects on the adult outcomes of women, but not of men (e.g., women with 20% higher rainfall in their year and location of birth attain 0.14 centimeters greater height, finish 0.15 more years of schooling, live in households with 5.2% higher expenditures *per capita*, and have spouses with 5.1% higher earnings). The authors suggest that these patterns most plausibly reflect a positive impact of rainfall on agricultural output, leading to higher household incomes and better health for infant girls but not boys, given that the elasticities of health and nutritional inputs for girls with respect to income, etc., are higher for girls than for boys. They also present suggestive evidence that eventual benefits for adult women's socioeconomic status are mediated by improved schooling attainment, which leads to higher spousal quality, which in turn improves socioeconomic status.

Maluccio, Hoddinott, Behrman, Quisumbing, Martorell, et al. (2009) advance beyond previous literature that examines the effect of early childhood nutrition on education in developing countries by using longitudinal data begun during a nutritional experiment during early childhood with educational outcomes measured in adulthood at ages 25–42 for a sample of adults originally from four Guatemalan villages in which the experiment was conducted. Estimating an intent-to-treat model capturing the effect of exposure to the intervention from birth to 36 months, their results indicate significantly positive, and fairly substantial, effects of the randomized nutrition intervention a quarter century after it ended: increased grade attainment of 1.2 grades by women (reducing a preexisting male advantage in average grades completed) via increased likelihood of completing primary school and some secondary school; speedier grade progression by women; a one-quarter standard deviation increase in a test of reading comprehension at age 25–42 years with positive effects for both women and men; and a one-quarter standard deviation increase on nonverbal cognitive tests scores at age 25–42 years for both men and women. There is little evidence of heterogeneous impacts, with the exception being that exposure to the intervention had a larger effect on grade attainment and reading comprehension scores for females in wealthier households. The findings are robust to an array of alternative estimators of the standard errors and controls for sample attrition.

These studies thus add to the literature (some of which is cited above and in Section 3.2.1 below) on the importance of preschool experiences on subsequent education and knowledge, in the latter two cases by following the same individuals over more than a quarter century.

Summary and implications.

- Though some dimensions of family background seem fairly strongly associated with educational investments and it seems possible to a number of observers that family background in turn is associated with access to capital, information and insurance markets, there is but fairly limited evidence that policy-manipulable aspects of family

background or of what would seem to be critical markets lead to substantial increases in education.

- Standard models of investment in education imply that such investments respond to increases in expected returns because of altruism on the part of the investors and/or budget constraints that depend on the expected returns. A few available relevant studies indicate that there are significant responses to changes in technology or markets that alter the expected returns. The recent increased collection of expectational data in household surveys in developing countries should facilitate further explorations of these effects.
- The controlled experimental evidence for CCTs in Mexico and the natural policy experiment evidence for school vouchers in Colombia both indicate fairly high economic benefits relative to costs (or internal rates of return) to these demand-side policies in these particular contexts, suggesting that there probably is high value to considering and carefully evaluating such possibilities in other contexts.
- A number of studies in different contexts suggest that improved preschool investments in physical, mental and social development leads to greater capacity for subsequent education, particularly doing the school years but in a few cases well into adulthood.
- The *PROGRESA/Oportunidades* CCT program perhaps is more important as a model of the importance of systematic data collection and analysis of a large-scale program than of a particular program to be blindly emulated, with the evaluation being important not only for validating some components of the program and modifying others, but apparently for maintaining the program in the face of a skeptical congress and a historic change in political parties (after over six decades of dominance by one political party).
- While there is some research on the demand side for preschool education-related investments, the vast majority of studies focus on formal schooling, with little attention to other ways in which knowledge might increase.
- Despite improvements in the nature of the data (e.g., with more experiments and longitudinal data) and in the analysis, present results often are predicated on some strong assumptions and are for but a few contexts, so there is considerable potential for new empirical research contributions on these topics.

3. SUPPLY SIDE—MODELS, ESTIMATES, AND POLICY IMPLICATIONS

Decisions regarding investments in education and responses to incentives that may affect those decisions also occur in part on the supply side, which feed back to the demand-side decisions and outcomes through the actual and expected s_i variables in the relations in Section 2.1.4. The supply-side decisions depend very much on the institutional structure, as well as the objective function of the relevant decision maker(s) and constraints under which the supply-side decisions are made.

3.1 Supply-side modeling

3.1.1 Private provision and profit maximization for educational service providers²⁶

First, consider an existing profit-maximizing educational service that has an existing capital stock²⁷ K^s and that chooses variables inputs I^q and the quality of those variable inputs Q^i at prices that may reflect market power in the input market (and therefore depend on the quantities purchased) in order to maximize profits within a one-period framework given a demand function that depends on the quality of the educational services provided Q^s and possibly the number of individuals being educated N^s by that entity (if the entity is large enough relative to the market in which it operates to have market power in the educational services market)²⁸:

$$\Pi = P^s(N^s, Q^s)N^s - P^i(I^q, Q^i)I^q. \quad (4)$$

Assume that the indicator of the quality of the educational services provided Q^s that affects the price of the educational services provided by the program P^s is the mean knowledge of the number of individuals being educated N^s by that entity at the end of the period under consideration. For each such individual, then, the production function in Eq. (1) gives the knowledge, and substitution of the relevant production function for the experience during the time period during which educational services are being received gives for each individual:

$$k_t = k^p(h_t, c_{tp}, s_{tp}, e_{t-1}, \dots, e_0, u_{tk}) \quad (5a)$$

into which can be substituted the relevant characteristics of the educational service provider to obtain

$$k_t = k^p(h_t, c_{tp}, I^q, Q^i, N^s, K^s, e_{t-1}, \dots, e_0, u_{tk}), \quad (5b)$$

which can be averaged over the N^s individuals to obtain the function that determines Q^s . Thus Q^s depends in part on choices made by the educational service provider in the short run (I^q , Q^i , N^s), as well as in the long run (K^s). The first-order conditions for profit maximization (under the assumption that the functional forms are such that there is a maximum), thus, include

$$\partial\Pi/\partial I^q = (\partial P^s/\partial Q^s)(\partial Q^s/\partial I^q)N^s - P^i - (\partial P^i/\partial I^q)I^q = 0, \quad (6a)$$

$$\partial\Pi/\partial Q^i = (\partial P^s/\partial Q^s)(\partial Q^s/\partial Q^i)N^s - (\partial P^i/\partial Q^i)I^q = 0, \quad (6b)$$

$$\partial\Pi/\partial N^s = (\partial P^s/\partial N^s)N^s + P^s = 0. \quad (6c)$$

Inputs of a given quality are purchased and used until the marginal cost of inputs (including that due to any market power in the input market) equals the marginal impact of those inputs on the price of the educational service (through improving the indicator of the quality of the educational service) times the number of individuals receiving the educational service (Eq. 6a). The quality of inputs selected is that that equates the marginal impact of input quality on the price (through affecting the quality of the educational service) times the number of individuals receiving the educational service and the marginal impact of input quality on the price of inputs times the quantity of inputs (Eq. 6b). The number of individuals receiving the educational service is adjusted to equate the marginal impact on the price of the service of increasing the number of individuals receiving the educational service times the number of individuals receiving the service with the price of the educational service. Changes in output or input markets or in capital stocks or the relevant functional forms create incentives for educational service providers to change their supply behaviors to re-establish these equilibrium first-order conditions. Some observations about these first-order conditions are:

- (1) The first-order conditions for the quantity and quality of the input depend on the price of educational services being dependent on some function of the quantity and quality of the educational services. Without such a dependency, the optimal quantity and quality are driven to a corner solution of zero. Either the prices of the educational services or the number of individuals who purchase such services must depend on the quantity and the quality of inputs for a profit-maximizing firm to use positive quantities of such inputs.
- (2) The first-order conditions for the number of students depends on the price of educational services being dependent on the number of individuals using such services, that is, on the provider of the services having some market power. Without such a dependency, the optimal number of individuals purchasing such services is at a corner solution of infinity. Of course it would seem that even if the price of educational services offered at initial levels of individuals purchasing such services were not dependent on the number of individuals purchasing such services, if such a provider expanded enough such dependence would result.
- (3) Increases in the prices of inputs conditional on quality create an incentive in equilibrium to reduce the inputs of that quality to re-establish the first-order condition in Eq. (6a) by increasing $(\partial Q^s / \partial I^i)$ enough given the increase in the price of inputs of a given quality.
- (4) Increases in the responsiveness of input prices to input quality $(\partial P^i / \partial Q^i)$ increase the input quality in order to increase $(\partial Q^s / \partial Q^i)$ to re-establish firm equilibrium.
- (5) The impact on the short-run equilibrium of the educational service provider having a greater capital stock is to increase the price and therefore the short-run profits through the quality of the service provided, but the impact on the inputs

depends on whether this capital stock is a complement or a substitute with inputs of a given quality.

- (6) The private educational service provider considered here is not efficient even if there are no externalities because it has market power in both the product and input markets, and in equilibrium operates by producing less output and using less inputs than it would without such market power. And if the educational service provider considered here does not have market power, then input and output choices appear indeterminant.
- (7) Beyond the characteristics of the educational service provider, in this case both (a) the history of educational experiences of the individuals who purchase this educational service (and therefore all the actual and expected right-side variables in reduced-form relations for past periods such as Eq. 3c and similar relations for other life-cycle stages in Section 2.1.4) and (b) the current inputs into current education provided by the family and the community affect the knowledge of individuals. Under plausible assumptions about the production technologies, the better the past educational experiences and the more the current-period family and community inputs that are devoted to the education of individuals purchasing the services of this educational provider, the greater the short-run profits of this educational service provider. Therefore there is an incentive for the educational service provider to attract clients who have better past educational experiences and greater current familial and community educational inputs (i.e., for “cream skimming”). If the educational service provider can use resources to recruit individuals with better characteristics to purchase educational services, then the objective function in Eq. (4) needs to be modified to include the cost of recruitment, a recruitment production function needs to be added, and there is an additional first-order condition related to optimal recruitment. For example, for recruitment based on innate genetic endowments e_0 ²⁹:

$$\Pi = P^s(N^s, Q^s)N^s - P^i(I^q, Q^i)I^q - P^rR, \quad (7a)$$

$$e_0 = f(R), \quad (7b)$$

$$\partial\Pi/\partial R = (\partial P^s/\partial Q^s)(\partial Q^s/\partial e_0)(\partial e_0/\partial R)N^s - P^r = 0. \quad (7c)$$

Thus, under profit maximization, resources are devoted to recruiting individuals with better genetic educational endowments to the point at which the educational service product price gain through increasing the quality through increasing the endowments times the number of individuals enrolled equals the marginal price of recruitment.³⁰

- (8) The assumption to this point is that the indicator of the quality of the educational services provided Q^s that affects the price of the educational services provided by

the program P^s is the mean knowledge of the number of individuals being educated N^s by that entity. An alternative, and perhaps more sophisticated from the point of view of potential clients, assumption is that the indicator of the quality of the educational services provided Q^s that affects the price of the educational services P^s is the mean increase in knowledge of the number of individuals being educated N^s by that entity at the end of the educational program. If the relevant production technologies are linear, this “value-added” assumption eliminates (a) the history of educational experiences of the individuals who purchase this educational service but not (b) the current inputs into current education provided by the family and the community that affect the knowledge of individuals. Therefore there would not be an incentive to recruit on the basis of past knowledge, but still on the basis of expected current family and community inputs (which may lead to similar recruitment patterns given serial correlations over time). If the relevant production function technologies include interactions between past knowledge and skills and current educational service provider inputs, then—as discussed in point (6)—there is an incentive to recruit on the basis of past knowledge and skills and the experiences and endowments that led to them.

3.1.2 Public sector educational service providers

The public sector plays a large role in the provision of educational services. In most countries the public sector provides the majority, some times the vast majority, of schools, considerable preschool services and a much smaller proportion of postschooling educational services, though often important training services.

The objective functions of these public sector educational services providers often are not well specified in the economics literature. Some of the analytical literature just assumes that public sector educational quality is an increasing function of public resources provided for education in almost a production-function sense (e.g., Epple & Ferreyra, 2006; Ferreyra, 2003 in their analysis of schooling). In such a case the usually implicit behavior seems to be basically to maximize something like W , the quality of educational services provided (equal to the average quality of educational services provided times the number of individuals who receive the educational services, with the latter a function of the quality under the assumption that the institution has to accept any individuals who choose to attend it) subject to the quality production function and the budget constraint:

$$W = Q^s N^s (Q^s) - \lambda (P^i (I^q, Q^i) I^q - B), \quad (8)$$

where B is the budget constraint that is determined by the political process and λ is the Lagrangian multiplier on the exogenously provided public budget. If Q^s is assumed to be the knowledge and skills of the individuals who attend this educational institution as above in Eqs. (6a)–(6c), the first-order conditions are

$$\partial W / \partial I^q = (\partial Q^s / \partial I^q) N^s + Q^s (\partial N^s / \partial I^q) - \lambda (P^i + (\partial P^i / \partial I^q) I^q) = 0, \quad (9a)$$

$$\partial W / \partial Q^i = (\partial Q^s / \partial Q^i) N^s + Q^s (\partial N^s / \partial Q^i) - \lambda ((\partial P^i / \partial Q^i) I^q) = 0, \quad (9b)$$

which imply

$$\begin{aligned} (\partial W / \partial I^q) / (\partial W / \partial Q^i) &= (\partial Q^s / \partial I^q) N^s + Q^s (\partial N^s / \partial I^q) / (\partial Q^s / \partial Q^i) N^s + Q^s (\partial N^s / \partial Q^i) \\ &= (P^i + (\partial P^i / \partial I^q) I^q) / ((\partial P^i / \partial Q^i) I^q), \end{aligned} \quad (9c)$$

subject to the budget constraint. If the budget constraint B changes, the variable inputs, and quality of those inputs change in the same direction, which alters the average quality of the educational services and the number of individuals who enroll in the same direction. Given employment relations in many public institutions, there may be a question of the extent to which staff (teachers, trainers) can be adjusted in the short run, but typically there are other variable inputs that can be adjusted relatively quickly. Whatever are the variable inputs for the relevant time horizon, their quality and quantity are adjusted in equilibrium to satisfy the first-order conditions in Eqs. (9a)–(9c). The resulting equilibrium of course differs from that of a private provider of educational services because of the different objective function (i.e., maximizing $Q^s N^s$ instead of maximizing Π) and different constraints (e.g., the public sector provider has a politically determined budget constraint and must accept individuals who wish to enroll). If there are no externalities, market power or other market failures, this means that the public sector provider must be inefficient because its equilibrium differs from the efficient equilibrium of a private sector provider with the same technology, same capital stock and facing the same market prices. If there are externalities, market power or other market failures so that the private sector provider is not efficient, generally the public sector provider with the objective function used here also will not be efficient (except in a epsilon probability case in which it happens by chance) because this constrained maximization is not designed to simulate net revenue (profit) maximization with true shadow prices rather than market prices that reflect market distortions. Of course the behavior of public sector educational services providers might reflect different objectives, such as the value added in clients' knowledge, somewhat parallel to the discussion in Section 3.1.1. Also there may be other considerations parallel to those discussed in Section 3.1.1.

Educational service provider incentives and effort. The discussion to this point assumes that there is no agency problem in assuring that, for example, contracted teachers or trainers perform their duties per their contracts. As discussed below, several recent studies question whether the failure to have incentives to assure such performance may not have a

widespread cost at least for public schools that many poor students in developing countries attend (e.g., Banerjee & Duflo, 2006; Banerjee et al., 2007; Duflo & Hanna, 2006). Within the general framework above, teachers can be considered a variable input into the production of educational quality and an important component of the quality of the teachers can be considered to reflect their effort. Teacher effort may in part reflect teacher altruistic interest in the knowledge of students or teachers may in part gain direct satisfaction from educating students well—and, for such reasons, their effort may be allocated more or less consistently with the stated objectives of educational service providers. However if there is a cost to monitoring effort so that monitoring is not perfect and if there is a cost (perhaps the opportunity cost, for instance, in using time for private tutoring) of providing time and effort, the staff of educational service providers are likely to be responsive to incentives that may lead them to devote less time and effort to the educational service provider than would be the case with costless and perfect monitoring. In such a case in equilibrium the marginal cost of monitoring is equated to the marginal gain in terms of the objective function through increased effort/time due to the monitoring.

Decentralization. In recent years, there has been considerable emphasis on decentralization of some public educational services to give greater ownership to local communities and to make service provision more efficient because it is responsive to local market and other conditions that affect education about which there is limited information centrally. Within the kind of models sketched out above, the latter consideration is effectively to use the correct local prices and capital stocks (including management and other heterogeneities) rather than national averages (or some other more aggregate representation) to make the allocation decisions, which is likely to lead to incentives to use resources more efficiently. On the other hand, there are arguments made against decentralization including (1) possible capture by local monopolies or elites, (2) increased risk of worse monitoring or imposing of standards, and (3) lessened possibility of attaining equity across space. Of course it is possible that gains from decentralization could be obtained by decentralizing some decisions, but some of the downsides could be limited by not decentralizing others (e.g., with regard to the second and third points in the previous sentence, by maintaining national standards and by maintaining interregional transfers in the interests of equity).

3.2 Estimates and policy implications

3.2.1 Knowledge and skill production functions, critical inputs, and incentives

Most estimates of what are interpreted to be educational production functions in the literature are estimates of the knowledge or achievement of school children as a function primarily of school inputs (e.g., students per teacher, books per student, school physical facilities) though in some cases with limited measures of parental characteristics included as well. Hanushek (1995) reviews 96 such studies for developing countries,

Table 1 Summary of 96 studies on the estimated effects of resources on education in developing countries

Input	Number of studies	Statistically significant		Statistically insignificant
		Positive	Negative	
Teacher-pupil ratio	30	8	8	14
Teacher's schooling	63	35	2	26
Teacher's experience	46	16	2	28
Teacher's salary	13	4	2	7
Expenditure per pupil	12	6	0	6
Facilities	34	22	3	9

(source: Hanushek, 1995)

summarizing the findings for six educational inputs: teacher-pupil ratio, teacher's education, teacher's experience, teacher's salary, expenditure per pupil, and physical facilities (Table 1). He concludes that, except for physical facilities, observed schooling inputs are not systematically related to student performance. Kremer (1995) points out that an alternative interpretation of these studies is that almost all of the school inputs raise test scores (the exception being the teacher-pupil ratio) because the probability is very small that several studies will find statistically positive coefficient estimates if the coefficient in fact is zero or negative. But from the point of view of the framework in Section 2.1.4, it is hard to know what to make of most of these estimates. The (usually implicit) maintained assumptions include that the observed right-side inputs are orthogonal with all of the unobserved variables, including preschool education, out-of-school learning, and ability and motivation and that contemporaneous inputs are perfectly correlated with earlier inputs. Further, as emphasized by Case and Deaton (1999), the studies apparently vary substantially in quality and in a number of cases it is difficult to reconstruct what actually was done because the information and data used are not available.

Glewwe and Kremer (2006) claim that "Since the mid-1990s, the most significant recent retrospective studies of the determinants of learning in developing countries are: the research of Ghanaian middle schools by Glewwe and Jacoby (1994); the study of Jamaican primary schools by Glewwe, Grosh, Jacoby, and Lockheed (1995); the investigation of grade 8 students in India by Kingdon (1996); and the paper on Philippines

primary schools by Tan, Lane, and Coustere (1997)." Though Glewwe and Kremer do not clarify why these are "the most significant recent studies," these do seem to be four relatively strong studies that attempt to estimate cognitive skills production functions with a number of schooling characteristics represented and concern about some estimation problems such as sample selectivity and, in the case of Kingdon's study, the endogeneity of initial child knowledge. In all four studies, the coefficients of most schooling and teacher variables are not significantly nonzero though each study reports that some teacher characteristic and some physical input (except for the Jamaican case) has significant impacts. Glewwe and Kremer conclude first of all that these studies provide "no general results regarding which teacher and school variables raise learning from these studies" and secondly that though there has been progress in econometric sophistication and measurement, the estimation problems remain substantial due to omitted variable biases (e.g., student ability, teacher motivation, factors that affect endogenous program placement or program characteristics), sample selection bias, and measurement errors so that "the results of these and other retrospective studies need to be interpreted very cautiously." They then argue that the use of natural experiments and randomized evaluations offer the best means for improving knowledge about the impacts of exogenous changes in supply-side educational inputs on knowledge and skills (though, as noted in Section 2 and Appendix, natural and controlled experiments generally cannot identify production function parameters).

Certainly obtaining estimates of educational production functions is difficult for reasons that Glewwe and Kremer elaborate upon and that also are discussed in Section 2.1.3 and Appendix. And it may be particularly difficult to estimate the impacts of a large number of detailed endogenously determined inputs into such production, for which reason approaches may have to be taken that use more aggregate indicators for constructs such as, for example, composite indicators of educational institution staff quality, of physical attributes of educational supply providers, of home environment, of an individuals' entering level of education. And it also is the case that reduced-form demand estimates of the sort that controlled experiments can explore can tell us a lot that is important about the gross impact (after all adjustments within the time period of the experiment) of specific detailed changes in educational inputs—conditional, of course, on the particular context. Yet in my view efforts to obtain better estimates of educational production functions are worth undertaking because they may be informative in themselves about the relative importance of different inputs into the production process and the extent of substitution/complementarity among these inputs and because they may be important components of structural approaches that can help increase our understanding of counterfactual policies. Obtaining good estimates of such production functions is a challenge, but a challenge probably worth undertaking with care to be clear what are the necessary assumptions to obtain estimates.

My judgment also is that some recent production function estimates have been undertaken for education in developing countries that have yielded new insights or raised useful questions. One example is the estimates of adult cognitive skills as dependent on schooling and pre- and postschooling experiences that are discussed in Sections 2.1.3. Two other recent examples follow:

- (1) *Marginal products of teacher subject knowledge on child learning in Peru, controlling for unobserved student and teacher endowments.* Metzler and Woessmann (2009) use unusual data from the 2004 Peruvian national evaluation of student achievement (*evaluación nacional del rendimiento estudiantil*) to estimate a variant of Eq. (1). These data include subject-specific student and teacher knowledge in both mathematics and reading for the sixth grade. Metzler and Woessmann control for both student and teacher unobserved endowments (and for school unobserved endowments) by estimating a within-student, within-teacher production function. They limit their basic sample to schools with only one sixth-grade classroom so that students have the same teacher for both subjects and there is not selective teacher assignment to subjects (though they also provide other estimates that suggest that this basic sample limitation does not affect their substantive conclusions). They find a highly significant marginal productivity of teacher knowledge, indicating that a one standard deviation increase in teachers' scores increases students' test scores by 0.04 standard deviations *ceteris paribus*, or somewhat larger with control for measurement errors. This effect is robust for subsamples defined by urban only, rural only, public only, multigrade schools, complete schools, first language of students' native (Spanish), male (female) student, female teacher, years of teacher with the class, and teachers having an institute degree. The effect is statistically weak in subsamples with only private schools or university teachers, which may reflect in part the small subsample sizes, and smaller in subsamples for student-teacher gender differences and for male teachers. Quantile estimates do not indicate significant nonlinearities. Their production function estimates therefore persuasively indicate that there are significant basically linear marginal productivity impacts of teacher knowledge on student learning, though they do not assess the costs or mechanisms for improving teacher knowledge or to what extent there are other effects such as increasing schooling attainment.
- (2) *Peer effects in Chinese secondary schools in Jiangsu Province.* Ding and Lehrer (2007) estimate a production function for college entrance examination scores using rich data on individual student records including the secondary school entrance examination results for all students and information on all teachers in a set of secondary schools, among other information. The authors argue that with their data and their approach they have (a) valid outcome measures due to the comprehensiveness of the exams and the strong incentives for students to perform well on the exams

(to attain admission into better colleges), (b) no attrition problem (since no students dropped out of the sample), (c) reduced omitted variable problems on the supply side due to good measures of teacher quality, and (d) very little omitted variable problems with regard to students because they have complete information on the secondary school assignment algorithm, the secondary school entrance examination and the cutoff rule for admission to different schools. Their alternative estimates of cognitive achievement production function (e.g., using the admissions cut off for regression discontinuity estimates, semiparametric estimates to explore the curvature in peer effects), indicate significant and fairly substantial peer effects (i.e., the impact of a 1% increase in mean peer entering examinations increases the college entrance examination score by 8–15% as much as a 1% increases in own entering score), but suggest that peer effects operate in a heterogeneous manner with students with high entering scores benefiting more from having higher-achieving schoolmates and from having less variation in peer quality than students with lower entering scores. While their observed data are not likely to eliminate all problems with unobservables, they probably do so much better than in many data sets and their careful exploration suggests interesting results for peer effects.

3.2.2 Reduced-form estimates of policy-related educational supply-side changes

Increases in the quantity of suppliers of education-related services *Preschool education.* The cognitive, social, and language developments of preschool children are increasingly seen as a critical factors in schooling attainment and skill acquisition and in turn for socioeconomic well-being in later life. In a recently published *Lancet* symposium on early childhood development (ECD), Grantham-McGregor, Cheung, Cueto, Glewwe, Richter, et al. (2007) estimate that there are over 200 million children under 5 years of age in developing countries who do not reach their developmental potential, which likely means that they are substantially less able to learn and take advantage of educational opportunities later in life and less productive and poorer as adults than if they had reached their full developmental potential. Delayed development is a cumulative process that starts as early as in the womb and once begun is difficult to reverse during school years and particularly in adulthood. Thus, Heckman (2006) and others have argued that policies that aim to improve child development have the largest impact and are the most cost effective if they intervene as early as possible and reach the most disadvantaged groups.

Within this context, interest in both new ECD programs and in conducting more systematic evaluations of such programs in developing countries has increased greatly. In the same *Lancet* symposium, Engle et al. (2007) summarize evidence showing effective programs for improving ECD in developing countries based on 19 systematic studies of ECD programs in developing countries that met six criteria: “(a) randomized controlled

trial or matched comparison group; (b) intervention before age 6 years; (c) effectiveness or program evaluations (not efficacy trials); (d) child development assessed; (e) targeted disadvantaged children; and (f) developing country” (Engle et al., 2007, p. 232). These interventions are summarized in Table 2. Almost all of these evaluations are for introducing new ECD programs (the Philippines case is an exception in which the new ECD policy primarily built on existing programs). The estimated effect sizes in cases in which they can be calculated indicate fairly substantial but varying impacts on cognitive skills, ranging from 0.19 to 1.8. These estimates suggest considerable but varying value in terms of cognitive development for the range of ECD programs examined in the contexts in which they were introduced. Because currently preschool enrollment rates, though growing, are fairly low in comparison with primary school enrollment rates, there would seem to be considerable potential for having important impact on education in developing countries through expanding ECD programs.

One shortcoming of these evaluations, however, is that they generally do not consider the economic costs of the programs, so that it is not possible to calculate benefit-cost ratios for these interventions.³¹ Another limitation is that there are no efforts to address possible efficiency reasons for policy interventions by identifying differences between private and social returns. Moreover, most of these studies do not address the implications of scaling up these programs (only one of these ECD intervention evaluations included in this table is based on a national sample and over a third are based on less than 10 communities). Further, they present evidence generally about fairly short-run effects on cognitive development and not much on longer-run effects (though two studies reviewed in Sections 2.1.3 provide some somewhat longer-run evidence—the Argentinean study by Galiani et al., 2009 considers effects on third-grade students’ attention, effort, discipline, and test scores and Maluccio et al., 2009 provides evidence of effects of early-life nutrition on adult performance on cognitive skills tests).

That this review found that there had been so few ECD interventions in all the developing world over a decade and a half that had been systematically evaluated, and that many of these cases were based on very few communities, indicates that there are likely to be high rates of return in terms of knowledge and in terms of the foundation for policy formation from expanded evaluation of ECD programs of different types in different developing country contexts.

School-age education. Many school-age children in developing countries must travel long distances to attend school, so one policy option is to construct new schools. Several studies using behavioral (nonexperimental) data are consistent with the possibility that school enrollments or attendance are inversely associated with distances from homes to school. For example, Glewwe and Jacoby (1994) present evidence on the significantly negative impact travel time had on schooling attainment and the positive impact on late enrollment for Ghanaians aged 11–20 using data collected in

Table 2 Summary of 19 available systematic evaluations of ECD programs in developing countries since 1990 based on Engle et al. (2007, Tables 2 and 3)

Country	Intervention	Child age	Outcome measures	Effect size of cognitive measure	Sample size for evaluation	Scale ^a
<i>I. Primarily center-based programs</i>						
1 Argentina	Increase in preschool places	3–5 years	Third-grade mathematics and Spanish achievement tests	0.23	>125,000	3
2 Bangladesh	Preschool run by NGO, feeding	4.5–6.5 years	(1) Cognitive development from WPPSI-III (2) School readiness (3) Play observation scale	0.20–0.23 1.4 0.19–0.72	208	1
3 Cape Verde	Formal preschool	3–6 years	Cognitive development (simplified Boehm basic concept test) at 5 years	0.29; 0.48 ^b	803	3
4 Colombia	Day care enter-based feeding and stimulation; five groups: food alone, and food + different time periods of stimulation, high SES control	42–75 m	Stanford-Binet test initially	NA	333 children (170 at follow-up)	1
5 Guinea	Informal community-based early learning centers	2–6 years	Cognitive development (simplified Boehm basic concept test) at 5 years	0.33; 0.66 ^b	877	2

6	Myanmar	Community-based ECD center and community support	3–5 years	(1) Primary school pass rate (2) Repetition rate for grade 1 (3) Test performance	NA	(1) 3484 (2) 1880 (3) 268	2
7	Nepal	Community-based ECD center (education and health)	3–6 years	(1) Primary school pass rate (2) Repetition rate for grade 1 (3) Annual drop out rate after 4 years	NA	935	2
8	Vietnam	Center and home (education, parenting, nutrition)	0–3 years for nutrition 4–5 years for education	Raven's colored progressive matrices at 6.5–8.5 years	0.25	313	1
<i>II. Parenting and parent-child interaction training</i>							
9	Bangladesh	Parent groups that meet weekly for 1 year; mean attendance of 12 sessions (range 0–42; assessment 2 m after end of program)	2–3 years	(1) Maternal knowledge (2) Home scale and subscales (3) Receptive vocabulary (4) Weight/height (5) Five preventative health behaviors (6) Mother-child picture and puzzle task	(1) 0.31 mother knowledge (2) 0.34 on home	329	2

Continued

Table 2 Summary of 19 available systematic evaluations of ECD programs in developing countries since 1990 based on Engle et al. (2007, Tables 2 and 3)—Cont'd

	Country	Intervention	Child age	Outcome measures	Effect size of cognitive measure	Sample size for evaluation	Scale ^a
10	Bolivia	Adult literacy programs and home visits (parenting, health, nutrition)	24 m; some older	Psychosocial development (rating of 1–4); fine motor, gross motor, hearing and language, personal and social assessment		454	2
11	Colombia	Nutritional supplement and/or a stimulation (home visit) program	Prenatal to 3 years, follow-up at 6 years	Griffiths at 4, 6, 12, 18, 24, and 36 m Locomotor, personal-social, speech & language, eye-hand coordination Einstein scale applied through 18 m	NA	433 families	1
12	Jamaica	Home visits by roving caregivers (health, nutrition, parenting, income-generating)	3–36 m	Griffiths mental developmental scales	0.5 ^b	163	1
13	Jamaica	Home visits by health aides (parenting)	9–30 m	(1) Griffiths mental developmental scales (2) Mothers' knowledge and practices of childrearing	0.8 ^b	130	1

14	Turkey	3 (center) × 2 (mother training) design; center = educational, custodial, or none; mother training = MT, NMT	3–5 years	(1) School attainment (2) School achievement (3) WISC-R vocabulary test	(2) 0.45	217 families	1
<i>III. Comprehensive programs for ECD</i>							
15	Bolivia	Child care centers in home (feeding health and nutrition monitoring, education)	6–72 months	Gross and fine motor skills, language and auditory skills, and psychosocial skills	0.4–1.5	1198	2
16	India	Integrated childcare center; support for pregnant and lactating mothers, growth monitoring, feeding	3–6 years	(1) Motor and mental development using WHO milestones assessment (2) Binet-Kamat IQ tests	NA	3724	3
17	Peru	Preschool and nonformal preschool	3–5 years	Grades (A–C) in mathematics and language (Spanish) as assessed by the first-grade teacher	NA	304	3
18	Philippines	Home (family day care programs, home visits)	0–4 years	ECD checklist of gross and fine motor skills, receptive and expressive language, socioemotional skills, cognitive skills, and self-help skills	0.5–1.8	6693	2

Continued

Table 2 Summary of 19 available systematic evaluations of ECD programs in developing countries since 1990 based on Engle et al. (2007, Tables 2 and 3)—Cont'd

	Country	Intervention	Child age	Outcome measures	Effect size of cognitive measure	Sample size for evaluation	Scale ^a
19	Uganda	Communication on ECD, child health days, village grants on nutrition, ECD centers	0-6 years	(1) Ugandan version of the British abilities scale (2) Parenting practices (3) Nutritional status	NA	2010	2

Sources: 1 Berlinski, Galiani, and Gertler (2006); 2 Aboud (2006); 3 Jaramillo and Tietjen (2002); 4 McKay, Sinisterra, McKay, Gomez, and Lloreda (1978) and Pollitt and Escamilla (1996); 5 Jaramillo and Tietjen (2002); 6 Save the Children (2004); 7 Save the Children (2003); 8 Watanabe, Flores, Fujiwara, and Tran (2005); 9 Aboud (2006); 10 Morenza, Arrazola, Seleme, and Martinez (2005); 11 Super, Herrera, and Mora (1990) and Waber et al. (1981); 12 Powell, Baker-Henningham, Walker, Gernay, and Grantham-McGregor (2004); 13 Powell (2004); 14 Kagitcibasi, Sunar, and Bekman (2001); 15 Behrman, Cheng, et al. (2004); 16 Vazir and Kashinath (1999); 17 Cueto and Diaz (1999); 18 Armechin et al. (2006) and Ghuman, Behrman, Gultiano, and King (2006); 19 Alderman and Engle (2007).

^aScale = 1 coverage < 10 communities; 2 coverage > 10 communities or district, but not national; 3 national coverage.

^bControlling for SES.

1988–1989 with ordered probits that allow for right censoring, but only a modest impact on test scores. Lavy (1996), with the same data, reports negative associations with distance to secondary school on primary school enrollment, which is consistent with forward-looking decisions and complementarities between primary and secondary schooling. Alderman et al. (1996) report that the probability of attending primary school in a rural Pakistani sample for 1989 was significantly inversely associated with distance from primary schools, though (surprisingly) also significantly positively associated to a lesser degree with distance to middle schools. Bommier and Lambert (2000) report significantly negative associations between distance to schools and schooling attainment in 1992–1993 for 5–25 year olds in Tanzania. For the most part these studies are consistent with the *a priori* plausible possibility that greater distance from school reduces school enrollment so constructing new schools closer to potential students who otherwise would have to travel far would increase school enrollment. But these studies do not control for the previous placement of schools and of households (though for the Pakistan study it is reported that school availability is not associated with any observed village characteristics), so distance may in part be proxying for other characteristics (e.g., families with less resources or less interest in market activities may tend to live further away from schools).

Duflo (2001) took advantage of the “natural experiment” of a rapid school expansion program in Indonesia to estimate the impact of building schools on schooling attainment while controlling for fixed community characteristics that might affect program placement.³² The Indonesian government launched a major school construction program in which more than 61,000 primary schools were built in the next 5 years and enrollment rates among children aged 7–12 years increased from 69% to 83%. Duflo examined schooling (and wages, see Sections 2.1.3) for men born between 1950 and 1972 by linking the 1995 intercensal survey (SUPAS) with district-level data on the number of new schools built in the district of birth of each individual. She found that each new school built per 1000 children was associated with an increase of 0.12–0.19 grades of schooling and an increase of 12% in the probability that a child would complete primary school. Trends across regions were parallel before the program and shifted clearly for the first cohort exposed to the program, which raises confidence in the identification assumption. Thus this study seems to advance knowledge about the impact of a school construction programs at least in this particular context. However, the approach does not permit identification of the impact of schooling construction from other correlated changes such as in the expansion of health services that may have occurred with similar targeting toward relatively poor areas. Moreover, as Oratzem and King (2008) observe, one cannot infer from these results that generally building schools will raise enrollment because the schools were built where they were most needed, suggesting that the enrollment response might have been atypically large and because migration of families most interested in schooling their children may occur

in response to the expansion of school construction, as noted by Rosenzweig and Wolpin (1988). Finally, this study does not provide information with which to assess the benefit-cost ratio or any differences between the social and the private returns to the program.

Changes in the quality of educational supply-related services *Preschool education.* As noted in the discussion of Table 2 above, most of the evaluations that are summarized in that table refer to new ECD programs, in many cases with fairly limited coverage. In contrast, the Philippine project did not introduce new services; rather, its innovation was to adopt an integrated, multisectoral approach to delivering a combination of services that include center-based (e.g., day care centers, preschools, health stations) and home-based (e.g., family day care programs, and home visits by health workers) interventions. Therefore if a simple dichotomy is made between increasing the quantity and the quality of educational services, arguably it fits into the latter category. Armeen et al. (2006) evaluate this Philippine ECD initiative using 3 years of longitudinal data that they collected on children age 0–4 years at baseline and “intent-to-treat” difference-in-difference propensity score matching estimators with differential impacts by child age and duration of program exposure. Their results indicate significant improvements in cognitive, social, motor and language development, and short-term nutritional status of children in ECD program areas, particularly for younger children. The impact on proportions of children with worms and diarrhea is mixed. They also suggest that having the actual initiation dates of programs in locales rather than the official or scheduled startup dates with which researchers often work changes the understanding significantly not only of the magnitude of program effects but of the impact of different durations of program exposure. Subject to the qualifications necessary for such data and estimation strategy, their results suggest fairly substantial impact of quality improvements in ECD programs.

School-age education. There are a number of studies of the impacts of school quality on outcomes related to education using behavioral (nonexperimental) data to estimate reduced-form demands similar to Eq. (2) or, in some cases, conditional demands similar to those discussed immediately after Eq. (2). The study by Glewwe and Jacoby (1994) on Ghana mentioned above with regard to distance, for example, is one of the more thorough available studies in some respects because of its multitude of indicators of school quality (18 school and teacher variables), consideration of several outcomes (school attainment, age of starting school, and cognitive skills), and the possibility of choice among local schools. The estimates suggest that schooling attainment is reasonably associated with school quality; if they are interpreted as reflecting causal relations, schooling attainment could increase by 2–2.5 years by raising average teacher experience (from 2 to 10 years), repairing leaking roofs, reducing travel time (from 2 h to a few minutes), or providing blackboards to schools without them. The authors suggest

that, since repairing roofs and providing blackboards is much less expensive than building new schools, repairing classrooms is a more cost-effective means of increasing Ghanaian schooling attainment than building new schools to reduce travel time. For another example, Drèze and Kingdon (2001) report that several school quality variables have statistically significant associations with primary school attainment for children 13–18 years of age (current enrollment for children 5–12 and ever-enrolled for children 5–18 also are considered) in four northern states covered by the PROBE survey in India: both provision of a midday meal and “waterproof” classrooms have strong positive association for girls (though not for boys); teacher absences due to nonteaching duties have a negative association for boys (though not for girls); a parent-teacher cooperation index has a positive association for both boys and girls; and class sizes have negative associations for both boys and girls. For a last example, Behrman, Ross, and Sabot (2008) estimate the “social” rates of return (in the sense of incorporating both private and social costs, but only private returns without any externalities) to expanding school quantity versus school quality in rural Pakistan in the late 1980s. To do so they estimate wages as a function of endogenously determined cognitive skills, where those skills depend on endogenously determined schooling attainment, reasoning ability (assumed predetermined), teacher/class ratios and teacher quality (the weighted average of teachers’ reasoning ability, cognitive test scores, in-service training, other training, and whether they were born in the village, with the weights determined simultaneously with the estimation of cognitive achievement and schooling attainment).³³ Their estimates suggest the highest “social” rate of return to having low-quality primary schools versus no schools (18.2%), then to increasing the quality of current low-quality primary schools to a high-quality level (13.0%), and last to increasing middle schooling subsequent to low-quality primary schooling (2.8%).

Studies such as these appear to be careful in a number of aspects of their research strategies (e.g., controlling for sample selectivity, in the first and third case treating some key behaviorally determined variables as endogenous and attempting to estimate relative returns, including a number of indicators of school quality), have some plausible implications and are suggestive with regard to the relative advantages of investments in alternative components of schools within particular contexts. But caution in interpretation nevertheless is advisable, as the authors of these studies seem to realize, because of possible important unobserved factors (e.g., student ability, teacher motivation, determinants of program location) and problems in identifying impacts of each of a fairly large set of correlated variables.

In light of such qualifications, some have tried to improve our knowledge through “natural experiments” that have resulted in policy variations. Case and Deaton (1999), for example, investigate educational outcomes in South Africa using data collected in 1993 when they claim that blacks did not influence funds provided to their children’s schools and that tight migration controls limited their possibility of migrating to areas

with better schools. They show that pupil-teacher ratios varied widely across black schools, and argue that this variation, combined with migration barriers and black South Africans' lack of control over their schools, generates a kind of natural experiment. Their estimates indicate that raising teacher-student ratios increases enrollment rates for and schooling attainment for blacks but not for whites. Because blacks had much larger class sizes than whites, this is consistent with the possibility that there are diminishing returns to reductions in class size. They estimate large effects from reducing class size at black schools: decreasing the student-teacher ratio from 40 to 20 (the approximate means in black and white schools, respectively) is simulated to increase schooling attainment by 1.5-2.5 grades.

The assumptions necessary to give a causal interpretation for identifying teacher-student affects still seem strong. As Case and Deaton note, there cannot be any other unobserved variables that affect educational outcomes such as teacher ability and motivation that are correlated with teacher/student ratios. Also, program placement and program characteristics must be orthogonal to unobserved determinants of education, a condition that could be violated easily by any conscious decisions about such matters, whether or not blacks had much say in these matters. Moreover, there are some measurement questions because, as Glewwe and Kremer (2006) note, the teacher-student ratios that are used for the study are not very correlated ($R^2 = 0.15$) with data from the Ministry of Education.

Chin (2005) also takes a natural experiment approach to evaluate a reform in India, "Operation Blackboard," that sought to provide a second teacher to all one-teacher primary schools. The central government paid for 140,000 teachers, equal to 8% of the prereform stock of primary-level teachers. Chin finds that less than half of these teachers were sent to the intended places. Also, teachers per school did not increase and class size did not decrease so that the only effect on school inputs appears to have been the redistribution of teachers from larger schools to smaller schools. This nevertheless generated increases in girls' primary school completion rate of 3-4% and in girls' literacy rates by 2-3%. Identification is based on the fact that cohorts participating in the program are assumed to be identified because only children attending primary school after 1987 were exposed, but there is some fuzziness if there was late entry into school or grade repetition.

Urquiola (2006) also attempts to use natural policy and geographical experiments to identify class-size effects, in this case in rural Bolivia, with two different research designs. The first focuses on variations in class size in rural schools with fewer than 30 students and therefore only one classroom. The second exploits regulations that allow schools with more than 30 students in a grade to obtain an additional teacher (akin to Angrist & Lavy's, 1999 well-known study of "Maimonides' rule"). Both designs suggest negative effects of class size on student test performance, with possible nonlinearities suggested by the larger estimates for the second than for the first design.

Such results are suggestive, though, as Urquiola emphasizes in his conclusions, they must be interpreted with care, are context specific, and particularly because of the possible nonlinearities, even if taken at their face value do not support across-the-board class-size reductions.

Other researchers increasingly have turned to evaluations based on controlled experiments, often in collaboration with NGOs operating in limited geographical areas of India or Kenya. Banerjee, Jacob, Kremer, Lanjouw, and Lanjouw (2000), for example, report that provision of additional teachers (usually female) in nonformal educational centers in rural Rajasthan, India, increased school participation by girls. Glewwe, Ilias, and Kremer (2004) and Glewwe, Kremer, and Moulton (2007) find that programs designed to improve school quality by providing inputs like textbooks or flip charts had no detectable effect on school participation, and limited effects on test scores. Their analysis exploring flip charts has an interesting comparison of different approaches. There are significant associations between having flip charts and school tests performance in 83 schools in two agricultural districts in western Kenya, with higher test performance by 0.20 standard deviations, which might plausibly reflect causality in that context in which school supplies were very limited so having a flip chart might make teachers more effective. But the associations between school performance and having flip charts also might reflect in part or in whole other factors, such as the quality of the teachers or of the administration or the support of parents for education. A difference-in-difference estimator with behavioral data suggests a modest or negligible impact of flip charts on test scores, suggesting that the positive association between having flip charts and tests scores in the cross-sectional behavioral data is due to unobserved fixed factors that affect test scores and that are correlated with having flip charts. And when flip charts were allocated randomly across schools following an experimental design in 178 schools, no significant impact was found between having them and school performance. The observed positive associations between having flip charts and school performance therefore apparently reflected *not* that in this context flip charts positively improved school performance, but that schools that had flip charts tended to also have other characteristics and the latter caused better school performance.

Banerjee et al. (2007) report on two randomized programs in schools operated by a NGO in Vadodara (for both programs) and Mumbai (for the first program) in Western India. (1) The first program was a remedial education program in which young women (“balsakhi”) from the communities work on basic skills about half of the school day with children who have reached grades 3 or 4 without attaining these basic skills. It was undertaken in 2001–2002 and 2002–2003 in 98 of Vadodara’s 122 governmental primary schools with random assignment for half of the schools to have a balsakhi to work with students in grade 3 in the first year and grade 4 in the second year (and *vice versa* for the other half). It also was undertaken in one ward in Mumbai in the same 2 years, with 32 schools in the treatment and 35 schools in the control group (though in the second year a third of the treatment group apparently did not receive treatment).

(2) The second program was a computer-assisted learning (CAL) program in which children in grade 4 were offered 2 h of shared computer time per week during which they play games that involve solving math problems the level of difficulty of which adapts to their ability to solve them. This program was implemented in the fourth grade in 55 municipal public schools in Vadodara in 2002–2003 (with 56 schools randomly assigned to the control group). In contrast to a number of earlier studies, both programs had fairly substantial significant positive short-run effects on student test score performance. The remedial education program increased average test scores by 0.14 standard deviations in the first year and 0.28 standard deviations in the second year, with the entire gain among the children with the balsakhi treatment rather than the reduced-class-size and more-homogeneous class treatment that the other children received (also suggesting no spillovers). The CAL program increased math scores by 0.35 standard deviations in the first year and 0.47 standard deviations the second year and was equally effective for all students. Such gains, however, are short-lived: 1 year after the program, initially low-scoring students in the balsakhi schools scored about 0.1 standard deviations higher than the control group and students in the CAL program scored 0.1 standard deviations more than the control group. The authors' enthusiasm about having found low-cost schooling interventions that work by adapting to the situations of individual students (they emphasize the advantages of the common background of the balsakhi with poor-performing students as a probably important factor) possibly “to dramatically increase the quality of education in urban India” (p. 20 in pre-publication draft) therefore should be tempered due to apparent short-run duration of most of the effects. Longer-run follow-ups and replications in other contexts with other conditions (e.g., regarding the availability of individuals to play the role of the balsakhi, particularly if being from the same community is important as conjectured) would be valuable.

Postschooling education. Karlan and Valdivia (2006) explore whether basic entrepreneurship skills can be taught, or are they fixed personal characteristics? Most academic and development policy discussions about microentrepreneurs focus on their access to credit, and assume their human capital to be fixed. The self-employed poor rarely have any formal training in business skills. However, a growing number of microfinance organizations are attempting to build the human capital of microentrepreneurs in order to improve the livelihood of their clients and help further their mission of poverty alleviation. Karlan and Valdivia use a randomized control trial to measure the marginal impact of adding business training to a Peruvian group lending program, FINCA, for female microentrepreneurs. Of 239 banks participating in the program, 104 were assigned to mandatory treatment groups, 34 were assigned to voluntary treatment groups in which clients were allowed to depart meetings after their loan payment was made but before training began, and 101 were assigned to control groups (with no training). Treatment groups received 30–60 min entrepreneurship training sessions

during their normal weekly or monthly banking meeting over a period of 1–2 years. Control groups remained as they were before, meeting at the same frequency but solely for making loan and savings payments. They find that the treatment led to improved business knowledge, practices, and revenues. The program also improved repayment and client retention rates for the microfinance institutions. Larger effects were found for those that expressed less interest in training in a baseline survey, which may have important implications for implementing similar market-based interventions with a goal of recovering costs. The estimated marginal costs of the training are smaller than the marginal gains due to a higher retention rate, which suggests that the training is worthwhile from the point of view of the banks. The study to date has not examined whether these effects persist over a longer period of time.

Incentives for supply-side providers and employees of such providers³⁴ Advocates of incentive pay for teachers and other educational service provider employees claim that these individuals face weak incentives to educate their clients, with pay determined almost entirely by schooling and training credentials and tenure and not by how much their clients learn. Opponents of teacher incentives based on indicators such as students' test scores argue that, since teachers' tasks are multidimensional and only some aspects are measured by such indicators, linking compensation to these indicators creates incentives for educational service sector employees to discourage weak clients from taking tests that measure levels rather than value added (by limiting their enrollment or encouraging them to drop out) and sacrifice promoting other capabilities such as creativity in order to teach to the tests used for the indicators of performance (Holmstrom & Milgrom, 1991). There is but limited empirical evidence on the effectiveness of monetary incentives for educational service provider employees in developing countries, though there are a few recent studies for teachers in schools.

Lavy (2002) evaluates a program in Israel that offered selected schools monetary incentives based on their students' achievements as measured by three indicators: average number of credits per student, proportions of students receiving matriculation certificates and dropout rates. The program was started in 1995 in 62 nonrandomly selected secondary schools that were the only school of their kind in a community (religious girls' and boys' Jewish schools, secular Jewish schools, and Arab schools). Levy compares the three indicators for these treatment schools with control group schools in which there were more than one kind of school in the same community. Fixed-effects estimates indicate that the program had positively significant effects after 2 years on two of the three indicators: average credits were 0.7 units higher and the proportion of students sitting for matriculation examinations increased by 2.1%, with larger effects for students whose mothers had less education. These results are consistent with the possibility that incentive pay affects school performance, that incentive pay causes schools to teach more strictly to the test, or that other changes that positively

affect these indicators were greater in the sort of communities that had only one school of their kind (which probably tended to be relatively smaller communities) than in communities that had multiple schools of their kind.

Glewwe et al. (2004) develop a model in which teachers can invest both in short-run teaching to the tests and in promoting longer-run learning and investigate the model implications using data from a randomized experiment on the impact of a teacher incentives program in Kenya. This program offered prizes based on schools students' average scores on district-wide exams, but penalized schools for dropouts by assigning low scores to students who did not take the exam. Student average scores increased significantly in treatment schools (0.14 standard deviations) during the 2 years that the program was in operation, but this gain did not persist a year after the program ended. Teacher attendance and student dropout and repetition rates did not improve, and no changes were found in either homework assignment or pedagogy. However, teachers were more likely to conduct test-preparation sessions outside of normal class hours and exam participation rose (presumably because schools wanted to avoid penalties for no-shows at exams). Thus, by looking at a number of details of teacher and students behaviors, this study separates better than most others whether the incentives improved longer-run performance or intensified teaching to the test, and the latter appears to have dominated.

Duflo and Hanna (2006) test whether improved monitoring and incentives based on teacher presence can reduce teacher absence, which is a widespread problem particularly in rural areas, in both formal and informal schools and whether such improved monitoring and incentives can induce more teaching activities and more learning. An improved monitoring and financial incentive program was initiated to reduce absenteeism in 57 informal one-teacher schools in rural Rajasthan in India, randomly chosen out of 113 schools. Baseline comparisons indicate that treatment schools did not differ significantly from control schools in terms of observed teacher attendance, teacher qualifications, infrastructure and number of students. Teachers in the 60 treatment schools were given a camera with a tamper-proof date and time function, along with instructions to have one of the children photograph the teacher and other students at the beginning and end of each school day. The time and date stamp on the photographs were used to track teacher attendance. Teachers' salaries were a direct function of their attendance. The remaining 60 schools served as comparison schools. The introduction of the program resulted in an immediate decline in teacher absence. The absence rate (measured using unannounced visits in both treatment and comparison schools) changed from an average of 43% in the comparison schools to 24% in the treatment schools. When the schools were open, teachers were as likely to be teaching in both types of schools, and the number of students present was roughly the same. The program positively affected child achievement levels: a year after the start of the program, test scores in program schools were 0.17 standard deviations higher than in

the comparison schools and children were 43% more likely to be admitted into regular schools. Instrumental variable estimates suggest that reducing teacher absence rates by 10% point increases test score by 0.10 standard deviation. Attrition was fairly large at 21% despite considerable efforts to track students who were not available at the first effort of the posttest, though the attrition for the posttest did not appear to be related to treatment. The estimated budgetary cost per 0.1 SD improvement in test scores is \$3.58 (though it would appear that this is an underestimate of the true economic cost if teachers have positive opportunity costs for their time when they are absent), which is comparable to or better than other recent estimates for programs to increase test scores. The cost per year of additional schooling is \$60, much higher than the deworming program in Western Kenya investigated by Miguel and Kremer (2004), but lower than that of other programs evaluated in Western Kenya. The authors conclude that their estimates suggest that high teacher absence rates contribute to low school quality, with probable long-run effects (based on earnings-schooling associations). While scaling up and longer-run effects are discussed, there is no mention of whether there is a longer-run follow-up of these interesting experimental results to see, for example, whether the effects increase or fade over time. The approach also does not permit comparison of other alternatives in the same environment, such as payments to teachers directly for improved student performance that would unbundle the effects of improved monitoring from those of increased incentives.

The studies reviewed in this section so far are focused on the proximate determinants of supply. But in the background is the question of what determines the extent to which, if at all, the public sector provides subsidies for education. Galor and Moav (2006) present a theoretical model in which capitalists have an incentive to invest in education of workers because of complementarities between physical and human capital in manufacturing so that human capital is of increasing importance in sustaining their profit rates. Galiani, Heymann, Dabús, and Tohmé (2008), in contrast, emphasize the demand for human-capital-intensive services by high-income groups. The demand for workers' education, voluntarily financed by a landholding elite, can be generated in their model with heterogeneous goods, where consumption preferences are nonhomothetic and the demand for skill-intensive commodities emerges at comparatively high levels of income. This model seems capable of accounting for salient features of the development of Latin America in the nineteenth century, where, in particular, land-rich countries such as Argentina established extensive public educational systems and developed sophisticated service sectors before starting significant manufacturing activities.

Decentralization and local community participation The decentralization of public services has been advocated widely as a means for making them more efficient and more responsive to heterogeneous conditions across localities by reducing information

problems. Many observers advocate decentralization and community participation to improve incentives for better use of educational inputs for educational service providers including preschools and schools (World Bank, 2004). Local communities arguably have the best knowledge about the needs of those being educated, strong incentives to monitor the performance of educational sector service providers, and a comparative advantage in conducting this monitoring. Decentralization reforms have been increasingly adopted.

Jimenez and Sawada (1999) evaluate the EDUCO decentralization program for schools in El Salvador. In this program, school committees are responsible for contracting and removing teachers and closely monitoring their performance and for equipping and maintaining the schools. All of their resources come from the central government and international organizations. Jimenez and Sawada find that the program successfully expanded education in poor rural areas and also reduced student absences by 3–4 days in a 4-week period. No effect was found on student achievement test scores. However, identifying the impact of the EDUCO program is difficult because there was no random assignment nor other means of persuasively controlling for selection into the program. Jimenez and Sawada attempt to control for selective program implementation through district dummies variables, but it is not clear why there are no direct unobserved district effects in the second stage relations of interest.

Reinikka and Svensson (2003) examine the effect of local community empowerment through an information campaign on delivery of nonwage funds from the Ugandan central government to schools. They estimate that the percentage of central governmental funds that actually reached the schools increased from 20% in 1995 to 80% in 2001, which they argue was mainly the result of better monitoring of local officials' handling of resources by the schools, stimulated by a governmental information campaign. In this campaign, data on monthly capitation grant transfers to districts were published in major newspapers and broadcast on the radio. Fixed-effects estimates indicate that schools with access to newspapers increased their funding on average by 12% points more than schools with no access to newspapers, despite the fact that the two groups had similar funding levels in 1995. This estimation strategy, of course, assumes that there were no unobserved time-varying changes that affected the receipt of funds at the school level and that were correlated with changes in newspaper access (e.g., Glewwe & Kremer, 2006 suggest that geographical variations in economic development might have been such a factor).

Galiani, Gertler, and Schargrodsky (2008) claim that decentralization can also degrade service provision relatively in poor communities that lack the ability to voice and defend their preferences. They examine such a possibility in Argentina, where two systems of secondary schools, one administered by the provinces and the other by the central government, existed side-by-side for over a century. Then in 1992–1994 the central government transferred all its secondary schools to provincial control.

The authors use the resulting differential exposure of different cohorts of students to decentralization in this “natural experiment” to analyze the impact of decentralization, with a number of robustness tests. The estimates suggest that this school decentralization had an overall positive impact on student test scores (math test scores increased 3.5% and Spanish test scores increased 5.4% on average after 5 years of decentralization), but that the decentralization gains did not reach the poor. The authors conclude that although bringing decisions closer to the people may help the good get better, the already disadvantaged may not receive these benefits and that in decentralized systems, central governments need to monitor service quality and guarantee a minimal level of service provision through targeted interventions to lagging areas. This study, however, does not provide any direct evidence on whether the decentralization changed the distribution of resources among schools to favor those with better-off clients. If there was not such a redistribution toward schools with better-off clients, it is hard to understand why schools with poorer clients were not able to use their resources better to improve student scores when centralized constraints on their actions were removed. The discussion of the participation of parent associations suggests that funds available were positively associated with the activity of these associations, but direct evidence is not provided. Also, the authors conclude that “In decentralized systems, the central government needs to monitor school quality, and to guarantee a minimal level of service provision through targeted interventions to lagging areas” (p. 23). It is not clear how this relates to the mechanism that the authors conjecture underlies their results, and therefore the underlying model. If the problem is that participation in parents associations is positively associated with income, how will better monitoring resolve the problem? Should the government somehow force poorer parents to participate in these associations—or provide incentives for them to do so?

Kremer et al. (2002) examine Kenya’s mix of centralized and decentralized control over different aspects of schools in a study described in Section 3.2.2. They argue that the system creates incentives for misallocation because of the reservation of some inputs for the centralized government (teacher assignments and teacher salaries) and others for local governments and schools (school buildings, material inputs such as textbooks). Local communities had strong incentives to build new schools, because once they had built one, the central government provided the teachers, which accounted for more than 90% of the present discounted cost of operating schools. Thus, many small schools, with small classes, were built close together. The system also led to excessive spending on teachers relative to nonteacher inputs. This system also created incentives for schools to set high fees and other attendance requirements, which kept many children out of school because increasing enrollment did not bring any more resources from the central government since a new teacher was assigned only when class size surpassed 55, and most classes were substantially smaller. Setting fees that lead to smaller classes also eases teachers’ workloads and could increase the school’s average score on

the national exams, the main criterion used to judge schools and headmasters. The inefficiencies of this schooling system are reflected in that the government could have financed the package of textbooks, classroom construction, and uniforms examined in the NGO experiment that Kremer, Moulin, and Namunyu study by using the savings that could be generated from much smaller increases in class size than those associated with this experiment. Thus, their study suggests that the details of decentralization are critical. Inefficiencies arose in this case from a mismatch between decision-making power and financial responsibilities since local communities had authority to start new schools while covering only a small fraction of the cost.

Summary and implications.

- Most of the large numbers of studies using cross-sectional behavioral data to estimate educational production functions are difficult to interpret because they do not control for behavioral choices in the presence of unobservables and in many cases use hybrid specifications with a mixture of production function inputs and determinants of those inputs included, so it is hard to be sure how much they support the well-known skepticism of Hanushek (1995) and others regarding what we know that works to improve education.
- A small number of some recent studies using better data, specifications and estimation procedures, however, more persuasively identify the impact of production function inputs on education and suggest in the particular contexts studied that, for example, the returns to improving the quality of primary school are likely to be higher than to extending postprimary schools in low-schooling environments such as rural Pakistan, that there are important peer effects in Chinese secondary schooling, and that there are significant impacts of improving teacher knowledge on student knowledge in Peru. But knowledge of educational production technologies in different contexts, including questions of to what degree different inputs are substitutes or complements, remains very rudimentary.
- There also is a growing number of reduced-form demand studies, using a range of data and approaches, which indicate fairly substantial positive impacts of expanding the availability of educational service providers on education both at the preschool and at the school levels. Given the current pattern of enrollment rates, in most developing countries this potential would seem to be greatest for preschool education and postprimary education (possibly including postschool education) rather than for basic primary schooling (though in some low-income developing countries expanding primary schools may be high priority). There remain many questions that are not explored in the literature about the potential advantages and disadvantages of expanding public school systems versus public subsidies for expansion independent of ownership as well as to what extent such expansion is justified by efficiency or distributional policy motives.
- There also is a growing number of reduced-form demand studies, again using a range of data and approaches, which suggest important possible educational gains from improving the quality of educational service providers at pre-, post-, and particularly school-age levels. Generally these studies refer to fairly specific interventions, such as

improving knowledge about the importance of stimulation for parents in preschool programs; increasing teacher-student ratios and reducing teacher absence in schools with a bundle of improved monitoring and incentives; and introducing business training in postschooling microcredit programs. There are also some studies that demonstrate that increases in some inputs that have been conjectured to be bottlenecks, with textbooks being the leading example, do not have positive impacts on education as measured, for example, by test scores and that incentives based on student test performance may create incentives to teach to the tests without longer-run gains in learning. These studies, of course, are for particular often fairly limited contexts, so it is not clear how much they generalize. They also tend to be of fairly short duration so it is not clear what are the longer-run impacts. They also do not illuminate to what extent the particular input(s) on which a study is focusing is effective or not because of interactions (either substitutive or complementary) with other inputs as would be possible with good studies of educational production functions. The studies to date also generally do not attempt to exploit any insights from generating hypotheses from systematically modeling educational service provider supply decisions along the lines sketched out in Section 3.2.1 or alternative lines. So they point to needs to undertake substantial future research on inputs and incentives in other contexts, particularly for the pre- and postschooling life-cycle stages.

- Decentralization of educational service providers *a priori* seems to have potential advantages given heterogeneity in local conditions and asymmetric information. Some studies suggest that these potential gains can be realized, but others raise questions about the distributional implications of decentralization and the distortionary incentives that might be established if there is partial decentralization. Again, further systematic research in other contexts and over longer time periods would be useful.

4. MARKETS FOR PRODUCTS AND INPUTS—MODELS, ESTIMATES, AND POLICY IMPLICATIONS

The previous two sections consider aspects of microbehaviors for the demand and supply for education. These microentities—individuals and households on the demand side and educational service providers on the supply side—are embedded in larger market and aggregate economies, to which this section turns. To date, however, there has been very little work on these questions for developing countries. Therefore this section is relatively short and mostly points to needs for further research.

4.1 Sorting in educational markets

If educational service providers and individuals seeking educational services both are heterogeneous then there are likely to be incentives to sort individuals systematically among providers. This tendency is likely to be reinforced if there are peer effects in the production functions for knowledge or for the quality of educational services (i.e., if one of the educational service provider characteristics that increases knowledge acquisition is the average knowledge of the other clients). This possibility has led to

concerns about better schools or private schools “cream skimming” the best students, to the detriment of the students in other schools.³⁵ I am aware of one recent study that explores some of these issues for developing countries.

Urquiola and Verhoogen (2006) examine how schools choose class size and how households sort their enrollment of their students in response to those choices in the highly liberalized Chilean school market. They develop a model in which schools are heterogeneous in an underlying productivity parameter, class size is a component of school quality, a class-size cap applies to some schools, and households are heterogeneous in income and therefore willingness to pay for quality. The model offers an explanation for two interesting empirical patterns (the second of which is published in Urquiola & Verhoogen, 2009): (1) There is an inverted-U relation between class size and household income in equilibrium, which tends to bias cross-sectional estimates of the effect of class size on student performance if household income is associated with students’ capacity for learning in school as seems plausible because of complementary education at home before entering school and while at school and because of intergenerationally correlated ability endowments. The upward-sloping portion in such a pattern reflects the fact that low-productivity schools may have trouble filling their existing classroom(s) to achieve the desired class size. The downward-sloping portion reflects the fact that among the higher-productivity schools used by higher-income households, quality considerations dominate: these schools find it profitable to restrict class size and charge higher tuition.³⁶ (2) Some schools at the class-size cap adjust prices and/or enrollments to avoid adding another classroom, which produces stacking at enrollments that are multiples of the class-size cap, which results in discontinuities in the relationship between enrollment and students’ families’ income at those points, violating the assumptions underlying regression-discontinuity research designs. Therefore, regression discontinuity approaches should not be applied in contexts in which parents have substantial school choice and schools are free to set prices and influence their enrollments.

4.2 Supplies and sectoral choices of childcare workers, teachers, trainers, and other educational service sector staff

The staff is the central input for most educational services. The questions of what determines the supplies of such staff and how are such staff sorted among various types of educational service providers (e.g., private preschools, public preschools, private schools, public schools, private training programs, public training programs—all varying by location, particularly between rural and urban area) are critical for understanding inputs and incentives in the production of education. Yet I am unaware of studies that address these questions for developing country labor markets.

Indeed, there are only a few papers that develop behavioral models of the decisions to become a teacher and to stay in the teaching force for developed economies, such as

the United States. Some recent studies examine teacher labor supply decisions using a competing risks framework (e.g., Boyd, Lankford, Loeb, & Wyckhoff, 2005; Dolton & Van der Klaauw, 1999; Hanushek, Kaine, & Rivkin, 2004). For example, Boyd et al. (2005) study the decisions of New York public school teachers to stay in the same school, transfer to another school or leave teaching in the state of New York. They find salary and proximity to home to be important factors in explaining why teachers choose to stay or move. They also find that teachers in their data also sort based on nonpecuniary factors such as class size, time available for preparation, student characteristics, and school facilities and that teachers have heterogeneous preferences for these characteristics. Using a similar modeling framework and data from Texas, Hanushek et al. (2004) find that, controlling for other student characteristics, teachers prefer high-achieving students.

Stinebrickner (2001) estimates a dynamic discrete choice model of occupational choices of teachers in elementary and secondary schools. Like the aforementioned studies, Stinebrickner's analysis is also based on individuals who became certified to teach and does not examine the decision to become a teacher. Stinebrickner finds that individuals who leave the teaching profession tend to be selective of higher ability individuals (as measured by SAT scores and college grades) and hypothesizes that rigidities in the public teacher salary schedule may in part account for the difficulty in retaining good teachers.

4.3 Educational expenditures and age distributions

As is well known, the world population is aging fairly quickly and in some parts of the developing world, particularly East Asia, most populations already are fairly old in comparison with the mid-twentieth century, with the result being that the biggest cohorts currently are of prime-adult working age. One result is that in relative terms, the demands for early-life education have fallen substantially and the number of working age adults per preschool and school student has increased significantly as part of what has been labeled the “demographic bonus” or “demographic dividend” (e.g., Bloom & Canning, 2004; Bloom, Canning, & Sevilla, 2002). What may be less well known is that this is a much broader phenomenon and that the projected most rapidly aging populations in coming decades in the world are not in developed regions nor in East Asia, but in Latin America and the Caribbean and in South Asia (Behrman, Duryea, & Székely, 2003). These shifts in the age distributions raise questions about which counteracting tendency regarding resources, particularly from the public sector, for education will prevail. On one hand, the reduced dependency ratio creates the potential for increased resources being available for early-life education. On the other hand, the relatively small younger birth cohorts may reduce their (or their advocates') political power with political processes shifting increasingly toward pension support and noncommunicable diseases that become more salient with aging populations.³⁷

Research on the implications of these changes for education in developing countries is noteworthy primarily by its absence.

5. CONCLUSIONS

Education occurs over the life cycle, not just in formal schooling. A focus on formal schooling alone is likely to miss important options for improving education of individuals, both before they initiate their schooling and after they complete their schooling. Investments in education should be viewed as occurring within a dynamic life-cycle perspective in the presence of unobserved heterogeneities, state variables that reflect the education to any particular point in the life cycle, and expectations regarding further developments, including future returns to education. Educational outcomes are determined by a combination of household demand decisions and educational supply provider decisions, with implications for educational product markets and educational input markets.

There have been a number of recent studies that have improved our knowledge of inputs and incentives relating to education within particular contexts. These studies have exploited a range of techniques to establish the counterfactual, from natural policy experiments to controlled experiments to propensity score matching to fixed-effects/instrumental variable estimates to structural models. There are a number of both demand-side policies (e.g., conditional in-kind and cash transfers, educational vouchers) and supply-side policies (monitoring and creating incentives for reducing teacher absences, expanding supply, possibly more individualized education) that appear to have some positive at least short-run impacts in particular contexts that are discussed in some detail in this chapter. But the limited number of studies of direct responses to incentives that are designed to directly influence particular behaviors reinforces that entities respond to incentives in pursuit of their own objectives, which is not necessarily how policy makers or academics thought they would or should, so that ongoing monitoring and evaluation is essential.

And, there also are a number of other possibly important gaps in the literature. There is a need to undertake further research on what appear to be promising educational interventions to see how robust they are to scaling up and to different contexts, explore the relevance of expectations on incentives for investments in education, investigate longer-run effects rather than just effects over a year or two since effects may be cumulative or dissipate over time, direct attention to education other than just schooling—particularly postschooling learning, investigate market and aggregate effects, exploit the insights of economic modeling, develop evidence on efficiency and distributional policy motives, and go beyond impact estimates and governmental budgetary estimates to calculate economic benefit-cost estimates for possible policy interventions. Just as there are not likely to be “magic bullet” interventions that improve education at all levels in all developing country contexts, there is not a “magic bullet”

research strategy for evaluating the alternatives. More controlled experiments with longer follow-ups in varying contexts, though with well-known limitations, would be desirable for evaluating many specific proposed interventions. But use of nonexperimental data with other means of establishing controls and of structural models to explore counterfactual experiments, with careful and explicit attention to the assumptions being made and their probable implications, also is likely to add importantly to our knowledge of inputs and incentives related to education and development.

APPENDIX: ESTIMATION ISSUES, POSSIBLE METHODOLOGICAL RESOLUTIONS, AND DATA

Questions about what impact incentives and inputs have on education in developing countries are basically empirical questions. Data limitations, no matter how good the data, lead to probable estimation problems. In all of the right-side relations above there are vectors of variables, and a number of the components of those vectors are likely to be unobserved or poorly measured.³⁸ For the production function relations and the conditional demand relations in Sections 2.1.3 and 2.1.4, moreover, some of the right-side variables are determined endogenously within the life-cycle framework.³⁹ As a result of these estimation issues—unobserved variables, measurement errors and endogeneity—the disturbance terms in the relations to be estimated are likely to include not only the stochastic terms (μ_s) but also components that are correlated with the right-side variables in the relations. For example the disturbance term in Eq. (8) is likely to include unobserved parental abilities and parental innate health and parental preferences and family connections, unobserved individual abilities and innate health and unobserved community characteristics such as the disease environment that may be related to program placement. These unobserved characteristics are likely to be correlated with the observed ones; for instance, parental schooling is likely to be correlated with their innate abilities, preferences and with family connections. As a result the OLS estimation of relations such as Eq. (8) is likely to lead to biased estimates of some key parameters of interest because in the estimation, for example, parental schooling proxies in part for correlated unobserved parental abilities, preferences, or family connections.

A.1. ECONOMETRIC APPROACHES TO DEALING WITH ESTIMATION PROBLEMS

Better data always help deal with such problems. Section A.2 addresses different types of data that may be used for the investigation of inputs and incentives related to education and development, and the better the data the less are likely to be such estimation problems. But there exist standard methodologies for dealing at least in part with these problems with existing or new data. Some examples follow:

Instrumental variable (IV) or two-stage least squares (2SLS) estimates. To break the correlation between the observed right-side variables and the compound disturbance terms that include unobserved determinants in addition to stochastic terms, one estimation strategy is to use IV or 2SLS. Good instruments must (1) predict well the variable being instrumented and (2) not be correlated with the disturbance term in the second-stage relation of basic interest. The model should indicate the set of potential instruments. The three reduced-form demand relations for the three life-cycle stage educational experiences in Eqs. (4), (6b), and (8), for example, give the potential instruments to be used to identify the three life-cycle educational experiences in the adult knowledge and skills production function in Eq. (1).⁴⁰ Note that these include experiments (e.g., receiving the Mexican PROGRESA treatment with random assignment by rural communities as examined in Behrman & Hoddinott, 2005; Behrman, Sengupta, et al., 2005; Behrman et al., 2009a,b; Schultz, 2004; the random assignment by communities of different nutritional supplements in the INCAP Guatemalan data as examined in Behrman, Hoddinott, et al. 2008; Maluccio et al. 2009; the random assignment of treatment of worms among Kenyan school children as examined in Miguel & Kremer, 2004; random assignment of remedial education programs, computer-assisted learning programs, and photograph-based monitoring of whether teachers are present in 98 to 120 Indian schools as examined in Banerjee et al., 2007; Duflo & Hanna, 2006) and so-called “natural experiments” in the form of natural events (e.g., weather fluctuations that occurred when the individual was a child that are used to identify schooling and health impacts on access to resources in Indonesia by Maccini & Yang, 2009) and policy changes (e.g., the Indonesian school-building program investigated by Duflo, 2001, 2004). Lagging such relations a generation suggests the potential set of instruments for the parental family background variables on the right side of Eq. (8). Good IV estimates, thus, can eliminate problems due to omitted (unobserved) variables, endogeneity and random measurement error.

Finding good instruments, however, is often not easy. Not all of the potential instruments that are suggested by the model structure, for example, are likely to be independent of the second-stage disturbance term. For the estimation of the adult intellectual capabilities production function in Eq. (1), for example, the reduced-form relations (4), (6b), and (8) suggest that family background characteristics are potential instruments. But if unobserved genetic ability endowments affect adult intellectual capabilities as posited in Eq. (1), if unobserved parental ability endowments affect their schooling attainment and income and if there are significant correlations between parental and child ability endowments, then parental schooling attainment and income may not satisfy the second condition for good instruments (and indeed do not in recent estimates of such a relation for Guatemala in Behrman, Hoddinott, et al., 2008). It may also be difficult to find instruments that predict sufficiently well the second-stage right-side variables. The econometric literature has been evolving recently in the

development of diagnostic tests for good instruments (e.g., Stock & Yogo, 2002 on the use of the Cragg–Donald statistic for the extent of bias due to “weak instruments” that do not satisfy the first condition for good instruments as well as would be desired). Overidentification tests of whether the instruments are independent of the second-stage estimates are thought to be relatively weak tests, but there are several recent examples in the development literature in which they do seem to reject including *a priori* questionable instruments (e.g., Behrman, Hoddinott, et al., 2008; Pitt, Rosenzweig, & Hassan, 2006).

Propensity score matching (PSM) estimates. Recently, there has been increasing development and use by economists (e.g., Heckman, Ichimura, & Todd, 1998) of PSM methods that were developed originally in the statistical literature (e.g., Rosenbaum & Rubin, 1983). These methods have been developed primarily in the context of the program evaluation literature. They try to find the best comparison for someone exposed to the program (“treatment”) among those not treated on the basis of observed characteristics.⁴¹ An increasing number of studies have been undertaken to estimate in particular program impacts in developing countries that are consistent with the general life-cycle framework presented in Section 2.1.1 (e.g., the impact of ECD programs in Bolivia in Behrman, Cheng, et al., 2004, in Mexico in Behrman et al., 2009a,b, and in the Philippines in Arme Cin et al., 2006; Ghuman, Behrman, Gultiano, Arme Cin, et al., 2006).

Fixed-effects (FE) estimates. Some of the unobserved variables that are likely to cause problems if they are not controlled in the estimates are fixed across observations in the data. From a longitudinal perspective (i.e., fixed over time) these include variables such as individual and parental genetic ability and innate health endowments, some aspects of community culture and environment, and longer-run fixed characteristics of educational service providers. From a cross-sectional perspective (i.e., fixed across observations in some group such as members of the same family or the same community) these include the family and community environments and endowments shared by siblings and other members of the same family, the preschool or school environment shared by students in the same preschool or school and the community environment shared by residents of the same community. Such factors that are fixed across observations can be controlled so that they do not bias estimates of observed variables through using dummy variables for each group of observations for which the control is desired (i.e., individuals or families over time, siblings or community members at a point of time). Such methods have been used extensively to investigate aspects of the framework in Section 2 (e.g., adult sister sibling estimates to control for shared childhood background in the estimation of the impact of mother’s schooling attainment on child schooling in Nicaragua in Behrman & Wolfe, 1987b; individual fixed effects to control for unobserved malnutrition that determined which children received nutritional supplements in the Mexican PROGRESA program in Behrman & Hoddinott, 2005

or which children were admitted to preschool programs in the Bolivian preschool PIDI program in Behrman, Cheng, et al., 2004 or family background when the exposure of sibling to a new program such as the PROGRESA schooling transfers depends on the age of the children as in Parker, Todd, & Wolpin, 2006). They have the advantage of controlling for unobserved fixed characteristics that otherwise might bias the estimates and numerous studies suggest that controlling for fixed effects changes the estimates substantially. For example, Behrman and Rosenzweig (2002a, 2005) present a dramatic example regarding intergenerational schooling effects for the United States. Controlling for fixed characteristics including genetic endowments at conception between adult identical twins changes the estimated impact of maternal schooling on child schooling from significantly positive in OLS estimates to negative in FE estimates—apparently because, controlling for endowments such as innate abilities, women in that society who receive more schooling tend to spend more time in the labor market and less time caring for their children (there are not parallel changes in the estimated impact of paternal schooling—which is consistent with fathers not changing their time spent in child care much if they have more schooling).⁴²

FE estimates have several limitations. First, they do not control for unobserved varying characteristics (e.g., time-varying prices in longitudinal estimates that may affect endogenous behaviors), for which reason in some studies they are combined with IV estimates (e.g., the investigation of the impact of nutrition on labor allocation in Bangladesh in Pitt, Rosenzweig, & Hassan, 1990 and in Pakistan in Behrman, Foster, & Rosenzweig, 1997). Second, they tend to increase the importance of noise relative to the signal, which tends to cause a bias toward zero. For this reason, FE-IV estimates have been used in some studies (e.g., using other respondents' reports for schooling attainment in the United States in Ashenfelter & Krueger, 1994; Behrman, Rosenzweig, & Taubman, 1994). Third, they do not permit estimates of the first-order impact of observed fixed variables, but only of variables that vary across the observations for which the fixed effects are used (though these may include interactions between fixed variables and variables that vary across the observations for which the former variables are fixed). Therefore, for example, they do not permit estimating the impact of parental schooling on child schooling unless parental schooling varies over time (as in the Rosenzweig & Wolpin, 1995 study of the impact of young mothers' schooling on ECD in the United States).

Construction of standard errors. Most household sample surveys collect data from clusters (e.g., census tracts, villages, neighborhoods)—or perhaps samples within clusters—because the fixed costs of data collection in a locale mean that a cluster design is much cheaper than would be, for example, a random sample of households in the overall population. The cluster design means that there are likely to be correlations across observations in the stochastic terms that, if not accounted for in the estimation of standard errors, might bias test statistics toward inferring greater significance to the results than is warranted. Estimation strategies that utilize within-family estimates may be

further subject to this problem. Moulton (1990), for example, notes, “[i]t is reasonable to expect that units sharing an observable characteristic . . . also share unobservable characteristics that would lead the regression disturbances to be correlated.” These correlations, if positive, may cause the estimated standard errors to be biased downwards. Therefore, it is important to assess the sensitivity of the results to the construction of the standard errors. A starting point is to test for heteroscedasticity and correct, where appropriate, standard errors using established methods (e.g., Huber, 1967; White, 1980). Recent studies by Angrist and Lavy (2002) and Wooldridge (2003), however, suggest that the usual corrections for clustering available in standard software are valid only when the number of units or groups or clusters of observations is large, say on the order of magnitude of 70 or greater. For many data sources this does not pose a problem, but for some it may because, for example, the data are from a relatively small number of samples. In such a case alternative standard error estimators can be constructed as indicated in Bertrand, Duflo, and Mullainathan (2004) by block bootstrapping the t statistics. Another approach is to aggregate all covariates up to their group means and carry out estimation on the average data (Wooldridge, 2003) at the cost of a considerable loss in degrees of freedom as the sample size drops from the number of households (or other entities of interest) to the number of clusters. Explorations of such alternatives in a recent study using 16 birth-year cohorts from four villages in Guatemala suggest that at least in that case these methods do not change substantially the inferences from the estimates (Maluccio et al., 2009).

Sample selection. Selection may take many forms: only having data on wages for those who participate in the labor force, only having data on test scores or other outcomes for those attending school, only having data on the impact of early childhood programs for those who survived infancy and earlier childhood, only having data on those who do not attrite in longitudinal data. The general problem is that those who are selected are not likely to be a random subsample. A general solution is to model the selection rule and to use it to correct for selection in the estimates, such as in the well-known Heckman (1974, 1979) two-step procedure or other methods such as maximum likelihood estimates.

Because sample attrition is a major concern for one major type of data, some elaboration on this type of selection is provided here. Sample attrition has the potential to invalidate inferences that can be drawn from longitudinal data if the attrition is non-random with respect to the behavior being studied. Consider the following canonical selection model:

$$L_t^* = b_2 + b_3x_t + b_4z_t + u_t^* \quad (10)$$

and

$$y_t = b_0 + b_1x_t + u_t^{**} \quad (y_t \text{ observed only if } L_t^* < 0). \quad (11)$$

Equation (11) is the model of interest (e.g., a simplification of Eq. 8). The outcome variable, γ_t , is observed only for a subset of the entire sample, those for whom the latent index variable, L_t^* , is less than zero. Equation (10) is a selection function depending (possibly) on the same independent variables in Eq. (11) as well as on additional factors. In practice, it is known only whether an observation is observed or not, that is, $L_t = 1(L_t^* < 0)$ if observed and $L_t = 0(L_t^* \geq 0)$, if not. If the error terms, u_t^{**} and u_t^* , are correlated, estimation of Eq. (11) on the observed sample, ignoring Eq. (10), may lead to inconsistent parameter estimates and thus incorrect inferences. Often attrition appears to be selective in the sense that mean values differ between those who attrite and those who do not (e.g., with respect to schooling attainment in the baseline). However, what is of concern is not the level of attrition nor such mean differences but whether, and to what extent, it invalidates the inferences that can be made using the data. It is desirable to attempt to address sample attrition, even if such efforts must be limited to considering attrition on observable variables. Some options (1) include testing with baseline data whether the coefficients in multivariate relations differ significantly for those who subsequently attrite and those who do not. Simple tests using data from both developing and developed countries often find no evidence of significant differences even if mean characteristics do differ significantly (e.g., Alderman, Behrman, Kohler, Maluccio, & Watkins, 2001 for Bolivia, Kenya and South Africa; Fitzgerald, Gottschalk, & Moffitt, 1998b for developed countries); (2) include in the specification of Eq. (11) all the plausible covariates, some of which may be associated with attrition. Conditional on the maintained assumptions about the functional form, attrition selection on observed right-side variables does not lead to attrition bias (Fitzgerald, Gottschalk, & Moffitt, 1998a); (3) implement correction procedures for attrition on observed variables that might relate to attrition even if they are not directly in the model, such as interviewer characteristics and whether other family members remain in the original sample unit (Fitzgerald, Gottschalk, & Moffitt, 1998a,b). Recent studies for developing countries find that most key results are not influenced by sample attrition on observed variables (Behrman, Hoddinott, et al., 2008; Maluccio et al., 2009). Given the potential importance of attrition in confounding the results, nevertheless, it is desirable for studies of education and development to test to the extent possible for attrition biases—and in new data collection, to try to limit the extent of attrition as much as possible (the Indonesian Family Life Survey, available on the web, provides an excellent model).

A.2. STRENGTHS AND LIMITATIONS OF ANALYSIS OF VARIOUS TYPES OF DATA AND RELATED METHODS

The discussion above points to considerable challenges in undertaking empirical estimates of causal relations pertaining to incentives and inputs related to education in developing countries. Better data lessens such challenges. The ideal would be

representative panel data with substantial detail updated frequently on every member of the family over several generations, substantial detail on the context (markets, public services, environment, kin and social networks, relevant educational services institutions) also updated frequently over the same time period, and a series of experimental and quasiexperimental shocks over the same time period that would permit identification of the short- and long-run causal effects. Such data are not available for any society, and the data that are available generally tend to be less satisfactorily (though not always) for developing than for developed economies. While it would always be desirable to obtain better data, it is also desirable to gain as much understanding as possible from existing data. Most data permit at least some examination of how robust the estimates of education and development are to some major assumptions regarding possible data limitations. This section considers various types of data and related analytical techniques in turn and how they can be informative about particular causal mechanisms pertaining to education, with some references to their use for developing countries.

A.2.1 Some major characteristics pertaining to data quality

Before turning to different major types of data, it is useful to note five critical aspects of data quality that are common across different data options:

- (1) *Representativeness.* How representative are the data for the population of interest? Can inferences be made for some population of interest beyond the sample, perhaps through weighting the observations appropriately? Some potentially very interesting data, such as individual and family histories, school- or training program-based data and much (though not all) qualitative data may raise interesting questions and conjectures for more systematic study, but be difficult to interpret with regard to their implications for broader populations because of the selection into the sample. The broader population need not be a nation (though some analysts seem to argue that, e.g., Jenkins & Siedler, 2006), but it should be clear of what is the sample representative.
- (2) *Power, sample size, and sample design.* Power refers to whether the sample is large enough to identify the effect of interest at a given significance level. The sample size necessary to have any particular level of statistical power, of course, varies depending on what question is being asked. For instance a larger number of households is required the more fine-tuned the question is with respect to demographic groups—so many more households will be needed to investigate the possibility of a given impact with given significance between parental schooling and school enrollment among 8-year-old girls than to investigate the possibility of the same percentage impact with the same significance between parental schooling and schooling attendance for all 6–12-year-old children (even with correction of the standard errors for clustering at the family level). If the sample design involves clustering, the number of clusters and the intracluster correlations are important in addition

to the number of households (see discussion in Section A.1). Data that in other respects might appear very promising for the analysis of education and development may not warrant analysis if the power is too low.

- (3) *Coverage of relevant variables.* To state the obvious, data are of value for the analysis of education and development only if they include some information on variables that capture critical elements of the demand side links across the life-cycle stages that are discussed in Section 2 or the supply-side issues that are discussed in Section 3.
- (4) *Measurement errors.* Data typically are imperfect representations of the underlying constructs of interest. Even for data such as self-reported completed schooling in developed countries, the noise-to-signal ratio⁴³ has been estimated to be on the order of magnitude of 10% (e.g., Behrman et al., 1994). Random measurement error in right-side variables tends to cause biases in the estimated coefficients toward zero—intuitively the noise masks part of the effect of the signal so the absolute magnitude of the coefficient is underestimated. This effect tends to be exacerbated in fixed-effects estimates because controlling for fixed effects tends to increase the noise-to-signal ratio. Random measurement error can be eliminated if there are multiple reports on a variable and the measurement error across the reports are not correlated (e.g., schooling attainment as reported not only by the individual but by others, as in Ashenfelter & Krueger, 1994; Behrman et al., 1994). IV estimates, as noted in Section A.1, may also eliminate this bias toward zero due to random measurement error. However, measurement is not only random, it might also have systematic components, particularly on what might be perceived as sensitive topics (e.g., Mensch, Hewett, & Erulkar, 2003). Such systematic errors may make inferences about such behaviors, even if they are very important in understanding education and development, very difficult.
- (5) *Timing and duration of coverage related to program startup or policy change.* For evaluating policies or programs, when the evaluation occurs relative to the program initiation may be critical (Behrman & King, 2008; King & Behrman, 2009). Often, programs have startup delays so actual effective startup dates are lagged behind officially announced starting dates, there are other startup problems so that the initial effectiveness of the program is less than after the initial “bugs” have been worked out, there may be information lags in potential clients learning about the program and its effectiveness, there may be scale economies or diseconomies, programs may have differential effects depending on the duration of exposure of clients and there may be “pioneer effects” due to the dedication of those initially involved that will dissipate after the program becomes routinized. Many data sets do not permit controlling for such factors, which may result in considerable misunderstanding of program effects in either direction, depending on which of these factors predominates.

A.2.2 Some major types of data for investigating education and development

Cross-sectional household data. Cross-sectional surveys and censuses are the most common type of available data. There are many cross-sectional surveys that are representative of some larger population of interest to the subject of this chapter, often with a stratified cluster sample design. Censuses, of course, are by definition representative of the populations covered except for possible undercounting (particularly of more marginal groups). There are also many cross-sectional surveys that are not representative, but based instead on behaviors such as attending schools. These nonrepresentative data sources may have rich information—but interpretation of the implications of analysis for broader populations of interest may be difficult unless it is possible to control for the selection decision into the sample. In some cases it may be possible to control for such selectivity into the sample by using other representative or census data to estimate the selection rules on a set of variables common to the selected and the representative data sources.

Cross-sectional data sources vary considerably in their sample sizes and statistical power—and, as noted above in Section A.2.1, the required sample size for a given level of power and significance depends on the extent to which the question being asked is focused on a narrow or broader demographic group.

Typically cross-sectional data do not include much information on variables necessary to estimate directly demand-side educational relations such as those that are sketched out in Section 2.1.4. Cross-sectional data can most commonly be used to estimate reduced-form relations in the spirit of Eqs. (4) and (6b)—that is, what are the relations between parental family background and indicators of child education during preschool and the school years for children who are coresident with their parents. There are many studies in the literature, for example, between parental characteristics and child enrolment in school, child progression rates through school (often represented by the gap between completed grades of schooling and the number of grades that would have been completed had the child started at the normal or legal age and progressed one grade each year), and child schooling attainment (though this variable is right-censored for children still in school; see King & Lillard, 1987 for estimates from the Philippines and Malaysia). If the information in such data is limited to coresident children, however, the selection of which children have left the household may make such analysis difficult for older children. The majority of the available estimates, though hardly all of them, indicate greater associations between maternal schooling and command over resources as reflected in income than between similar characteristics for fathers (e.g., Thomas, 1990 for Brazil).⁴⁴

The typical cross-sectional data permit some, but limited, control for the estimation problems that are discussed above. For instance, the cluster structure of many cross-sectional data sets permits the control for unobserved cluster (e.g., community) effects

that might be correlated with family background characteristics and cause biases in the estimated impact of family background characteristics if not controlled. That information is available on a number of children also permits the investigation of time-varying changes that affect siblings differentially; this is not likely to be useful often for the first-order effects of parental characteristics,⁴⁵ but may be for evaluating program effects (e.g., Parker et al., 2006 on the impact of the Mexican *Oportunidades* program on schooling of children too old versus those of age to be affected by the program) and possibly their interaction with family background characteristics. Also, the inclusion of a number of assets in many such data sets permits the construction of more-permanent measures of parental household resources through using such assets as instruments in IV estimates rather than using current income (which often has large transitory components for poor families in developing countries; see Deolalikar & Gaiha, 1993 for rural India), which tends to lead to much larger coefficients of parental income in child schooling relations and thus greater estimated impact of family income on education than might appear with current income only (e.g., Behrman & Knowles, 1999 for Vietnam).⁴⁶

A subset of cross-sectional surveys have additional information that make them richer than most cross-sectional data sets for examining the questions considered in this chapter. Some examples include: schooling attainment for all of the parents of household members whether coresident or not, which opens up the possibility of investigating the relation between income or expenditures of current adults and their parents' schooling attainment along the lines of Eqs. (9a)-(9c) (e.g., for Brazil, see Lam & Schoeni, 1993, 1994); schooling attainment for all (not just coresident) children of the adults in the household, which permits investigation of Eqs. (6a)-(6c) without the selection problems due to older children having left the household prior to the survey (e.g., Parker et al., 2006 for Mexico); information on income by individual adults including "nonlabor earnings" that arguably are not correlated with the unobserved endowments in Eqs. (4) or (6a)-(6c) so that the impact of mothers' versus fathers' control over resources can be investigated (e.g., Thomas, 1990 for Brazil); information on assets bought into the marriage by the current adult parents that can be used to explore the impact of resources under control of mothers versus fathers on investments in children that arguably are independent of the endowments in Eqs. (4) and (6a)-(6c) (e.g., Quisumbing & Maluccio, 2003 for Bangladesh, Ethiopia, Indonesia, and South Africa).⁴⁷

Many cross-sectional data sets can be enriched by linking them with time series administrative data on public services (particularly related to health and education), communication and transportation, and weather conditions. For example: (1) even if the basic household data being used are cross-sectional, time series on available services may be informative for time periods earlier in their children's life if there is reason to believe that the impact of parental characteristics depends on the nature of such services

(e.g., mothers' schooling enhances the positive impacts of health services when children are very young); (2) such data may make possible within-sibling estimates if different siblings faced different community services during critical periods such as early childhood, again, perhaps in interaction with parental characteristics; (3) such data may provide instruments that arguably are independent of the unobserved factors on the right-side of the relations in Sections 2.1.4 and 3.2.1 but that predict sufficiently well the right-side parental characteristics that good IV estimates can be obtained (e.g., the nature of schooling options when parents were of school age might provide good instruments for parental schooling attainment, as for a different purpose for Indonesia in Duflo, 2001; levels and variations in rainfall may provide good instruments for parental income in agricultural areas, as for a different purpose in India by Wolpin, 1982 and in Thailand by Paxson, 1992 and for purposes much more directly related to this note in the study of health, schooling, and socioeconomic consequences in Indonesia by Maccini & Yang, 2009).

Cross-sectional data on educational service providers. Such data may be available in a stand-alone form if there are, for example, censuses of such providers or administrative record on such providers. There also are a number of data sets in which such data have been collected in linkage with household surveys, which permit enriched analysis of the issues of interest in this chapter (e.g., Glewwe & Jacoby, 1994 investigation of school choices in Ghana). Typically such data on educational service providers do not include information with which to control for behaviorally determined right-side variables (such as inputs in a production function) or other important unobservables (such as those emphasized in the “endogenous program” literature; see Rosenzweig & Wolpin, 1986).

Longitudinal or panel data. These data generally provide a more satisfactory means than do cross-sectional estimates of identifying estimates pertaining to education and development because: (1) the prospective data gathered in earlier rounds is likely to be less contaminated with measurement error and more complete than recall data from cross-sectional data sources; (2) the multiple observations over time in some cases permit the control for unobserved individual fixed effects such as innate ability and health; (3) the multiple observations over time permit the exploration of dynamics of effects and whether they tend to diminish over time or are enhanced over time, perhaps in part in interaction with dimensions of the environment in which the individual is developing (e.g., do early-life nutritional shocks have only short- or long-run effects, and to what extent does it depend on whether subsequently the school system or other institutions can in part or in whole compensate for them); and (4) the multiple observations over time permit exploring the impacts of possibly changing contextual factors, depending in part on how rich is the contextual information.

There currently exist relatively few longitudinal household data sets from developing countries with panels over several decades as may be needed to see how preschool

experiences affect adult knowledge and skills and other outcomes. But there are a few. Examples include: the INCAP Guatemalan data on children 0–7 years old in 1969–1977 with follow-up rounds in 1988–1989 and 2002–2004, at which time the children were 25–42 years of age (Martorell, Behrman, Flores, & Stein, 2005); the Cebu (Philippines) Longitudinal Health and Nutrition data of births in 1983 with the last follow-up in 2005 when the children were up to 20–22 years old and their mothers were from 35 to 69 years old (Cebu Study Team, 1991; Daniels & Adair, 2004); the Pelotas Brazilian data on the birth cohort of 1982 with the last follow-up in 2004–2005 when the children were up to 25 years of age (Victora & Barros, 2006; Victora, Gomes, & Barros, 1990); the NCAER rural Indian data starting in 1969–1971 with follow-up until 2002; the Bangladeshi nutritional data with follow-up after over two decades (Pitt et al., 1990, 2006).

There are many more longitudinal data sets that cover shorter, but important segments of the life-cycle stages noted above. A few examples include: The Mexican PROGRESA data for 1997–2003; a number of the Demographic Health Survey (DHS) data sets; the Vietnam Living Standard Measurement (LSMS) Survey; the Chilean *Encuesta de Protección Social* survey from 2002–2006 (Bravo, Behrman, Mitchell, & Todd, 2006); the Bolivian PIDI evaluation data (Behrman, Cheng, et al., 2004); the Malawian Diffusion and Ideation Change Project Data for 1998–2006 (Watkins, Zulu, Kohler, & Behrman, 2003); the Kenyan school-based sample (Miguel & Kremer, 2004); the Colombian *Familias en Acción* sample for 2002–2006 (Attanasio, Gomez, Gomez Rojas, & Vera-Hernández, 2004); the Philippines Early Childhood Development Survey for 2001–2006 (Armecin et al., 2006; Ghuman et al., 2005; Ghuman, Behrman, Gultiano, Armecin, et al., 2006; Ghuman, Behrman, Gultiano, & King, 2006); the Mexican Family Life Survey (Rubalcava & Teruel, 2004); the Indonesian Family Life Survey (Thomas et al., 2003).

Longitudinal data can be, and in some cases are, enriched in ways that are parallel to cross-sectional data: inclusion of questions for previous generations or other people not currently in the households, linkage to administrative data. In addition, some longitudinal data have built into their design controlled experiments with random assignment between treatment and controls groups.⁴⁸ Some prominent examples include: The Mexican rural PROGRESA program with random assignment of initial treatment versus controls for 506 communities (Behrman, Sengupta, et al., 2005; Schultz, 2004); the Kenyan random assignment of various treatments (including deworming, flip charts) among 75 schools (Miguel & Kremer, 2004); the Guatemalan INCAP data with random assignment of nutritional supplements among four participant communities (Maluccio et al., 2009); experimental assignment of remedial education and computer-assisted learning programs in 98 schools in Mumbai and Vadodara, India (Banerjee et al., 2007). Such experiments provide (a) capacity for identifying the causal effect of treatment and (b) the possibility of identifying the impact of one behavioral choice

affected by the treatment on another by using the experimental assignment as an instrument for IV estimates. But there also are limitations of experiments: some experiments may be viewed as unethical or politically unwise; selective attrition between the treatment and control areas may introduce selectivity biases; generally experiments cannot provide direct estimates of production technologies or other structural relations in Sections 2 and 3.2.1 that might be of interest; and even very good experiments only provide “black box” estimates of the impact of the specific intervention used and not of alternative counterfactuals, including longer-run impacts.^{49,50}

Time series of cross-sectional surveys. A time series of cross-sectional surveys provides a means of tracing cross-sectional associations over time as cohorts age and possibly permit controlling for cohort-specific unobserved factors. This has the advantage of using more readily available data than longitudinal data, as well as data that are representative for each cross section. Deaton and Paxson (1994) give an example in which they trace the persistence of earnings shocks experienced early in the adult life cycle as cohorts age in Taiwan and the United States. The possibilities for using such an approach to investigate education and development seem limited, but perhaps underexplored.

Qualitative data sources. Most other possible data sources for investigating education and development can be considered to fit within the categories of being either cross-sectional or longitudinal (particularly since cross-sectional and longitudinal data may be either quantitative or qualitative). The same general questions of data quality (Section A.1) apply for such data sources: representativeness, power, variable coverage, measurement errors, and time and duration relative to program initiation. Extensive family or individual histories or focus groups may provide useful insights regarding hypotheses regarding education and development whether or not they are representative or have sufficient power or whatever the nature of the measurement errors. But if inferences are to be drawn from such data sources about aspects of education and development for some population larger than the sample itself, it is necessary to know how the sample relates to the larger population and to assure that power is sufficient and to understand measurement errors.

End Notes

- *. The author thanks the editors and an anonymous referee for useful comments on earlier drafts.
- 1. William R. Kenan, Jr. Professor of Economics and Sociology, Research Associate of Population Studies Center.
- 2. This broad definition of education is consistent with some major international declarations on education and development. UNESCO (<http://www.UNESCO.org/education>; 2 May 2007), for example, states that the 1990 Jomtien Declaration on Education for All “begins by stating . . . that ‘[e]very person—child, youth and adult—shall be able to benefit from educational opportunities designed to meet their basic learning needs . . . The Declaration also . . . recogniz[ed] that learning begins at birth. It called for early childhood care and initial education to be provided through a variety of arrangements involving families, communities or institutional programmes . . .’”

3. Much of the earlier economics literature on education and development focused on estimating the returns to education, and in particular to schooling attainment (see surveys in Psacharopoulos, 1985, 1994; Psacharopoulos & Patrinos, 2004; Schultz, 1988). For the purpose of this chapter, the returns to education are relevant only in so far as their expectations affect investments in education (see Section 2).
4. Some studies focus on the governmental budgetary costs of alternative interventions to attain the same educational targets (e.g., Banerjee, Cole, Duflo, & Linden, 2007; Duflo & Hanna, 2006). But budgetary costs are *not* the same as the economic resource costs because they do not include private costs and they do include transfers (Duflo & Hanna, 2006 recognize the latter point in a note on the Mexican *PROGRESA/Oportunidades* program). See Knowles and Behrman (2005) and references therein for more discussion.
5. School ages are not identical to ages when in school because those ages reflect choices both regarding when to start and when to terminate schooling.
6. Each child's expected adult full income arguably should be adjusted by that child's expected marriage arrangements because such arrangements are likely to affect the resources under direct control of the child or the child's parents.
7. Parents must have expectations, for example, on the impact of demographic changes (e.g., relative cohort sizes may affect inversely the returns to schooling as reported in Behrman & Birdsall, 1988 for Brazil) and general economic development (i.e., the returns to investment in education may depend on how dynamic the economy is, as emphasized by Rosenzweig, 1995; Schultz, 1975; Welch, 1970).
8. If parents (or children themselves once they participate in the educational investment decisions) are risk averse and the returns to investments in education have greater variance than some alternatives, investments in education will be less than under risk neutrality.
9. Other production functions also may play important roles in the overall process. For instance, if learning capacity depends importantly on health and nutritional status, the production function for health and nutritional status is likely to be an important component of certain aspects of the determination of inputs for education.
10. The nature of the elasticity of substitution has been emphasized recently in the literature on the United States, for example, by Cunha (2006) and Cunha and Heckman (2006).
11. Most other cognitive skills production function estimates in the literature focus on schooling, in some cases with family background factors as well (that may be correlated with unobserved endowments, as found in the Guatemalan study), and are summarized in Section 3.2.1.
12. "Ability" is in quotation marks here because it refers to all relevant unobserved endowments that affect both schooling and earnings (e.g., family connections, parental role models), not just innate intellectual ability (see Behrman & Rosenzweig, 1999).
13. Subsequent studies also find that including school quality changes significantly the interpretation of such functions (e.g., Card & Krueger, 1992a,b).
14. Including his estimates for the Philippines; Girma and Kedir (2003) for Ethiopia; Mwabu and Schultz (1996) for South Africa; Fiszbein, Giovanoli, and Patrinos (2004) for Argentina; Patrinos and Sakellariou (2004) for Venezuela; Montenegro (2001) for Chile; Patrinos and Metzer (2004) for Mexico; Arabsheibani, Carneiro, and Henley (2003) for Brazil; and Lee and Lee (2002) for Korea among developing countries and his estimates for Singapore, Buchinsky (1998) for the United States; and Martins and Pereira (2004) and Pereira and Martins (2000) for Austria, Denmark, Finland, France, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.
15. There are a few, but quite few, exceptions to this statement. A notable important exception is Foster and Rosenzweig's (1995) study of spillover effects of schooling on technological adoptions in Indian agriculture. But the empirical basis for social returns and incentives for education differing is very

sparse, despite frequent claims that this is the case. And what are called social returns often are not social returns in the sense that the term is used above. For example, in well-known surveys by Psacharopoulos cited above the private estimated rates of return are reduced to account for public sector budget schooling costs to obtain what he calls “social rates of return” but these are not truly social rates of return in the sense used above because they do not account for any social impacts beyond the private impacts. Likewise, the impacts of parental (mostly maternal) schooling in household production (e.g., health, nutrition, child education) are at times referred to as social returns to schooling, but—although conceptually a differentiation between private and social returns in household production is possible (e.g., if more parental education reduces contagious diseases)—most of the related empirical literature does not directly address such possible differences between private and social returns.

16. The basic question is what choices governments make with regard to equalizing versus productivity-enhancing public sector allocations. Estimates for Brazil suggest that governmental geographical schooling allocations had a significant concern for inequality but nevertheless were productivity improving on net, though of course not as much as they would have been had there not been some concern about equity (Behrman & Birdsall, 1988). This means that the useful estimates of schooling impacts probably overstate causal schooling effects because schooling is partially proxying for unobserved geographical endowments that were associated with greater allocations of schooling.
17. There are many more such studies. In a now quite-dated survey, Behrman (1997) reviews several hundred estimates from 20 (mostly developing) countries.
18. *PROGRESA* is an acronym for the original name of the program (*Programa de Educación, Salud y Alimentación*, Program for Education, Health, and Nutrition) introduced in 1997 in the Zedillo government. When the Fox government came into power after the 2000 election, the program was modified in some details (e.g., coverage of upper secondary schooling, extension into more urban areas) and renamed *Oportunidades*.
19. Their eventual incorporation into the program was not announced as part of the initial program. Given the history of old programs being allowed to fade out with presidential elections in Mexico, it would seem that from the point of view of the control group the probability of being incorporated into the program in time to have much expected gains from the program was substantially less than one. To the extent that they did have expectations of being incorporated into the program and to the extent that these poor households could transfer resources over time, then they may have adjusted their behaviors somewhat when the program was implemented for others, which is likely to lead to underestimates of the program impact in comparisons of treatment and control households.
20. Simple comparisons of those in treatment households versus those in control households do not indicate significant effects, apparently because of selective provision of nutrient supplements despite the overall experimental design, but their preferred estimates also control for child fixed effects.
21. The schooling cash transfers are conditional on schooling attendance 85% of the time. This conditioning would seem *a priori* to put pressure on teachers to over-report attendance so that the families of students who did not attend school sufficiently were not penalized. I am not aware of evidence on this point for this program, though Linden and Shastry (2005) provide some such evidence in a different context. Conditioning the program on education or increases in knowledge, such as increases in test scores, would seem to eliminate this particular risk and make the program more directed to the desired target of increasing education rather than time in school. *Oportunidades* planning to under in 2009 an experimental evaluation of such possibilities in cities in which the program has not been introduced.
22. These benefit-cost estimates are based on alternative assumptions of key factors (e.g., rates of return to schooling, discount rates) and an attempt to compare the benefits generated by the present discounted value of future wage (productivity) increases (which may be an underestimate if schooling, say, has

positive externalities or if improved nutrition has impacts additional to those of schooling) with the resource costs (e.g., the opportunity costs of time, administrative costs, distortion costs of raising governmental revenues). Given the large transfer component of the program that was an intentional dimension from the start as part of the antipoverty concern of the program (e.g., Levy, 2006), not surprisingly these efforts at estimating the benefits and costs of the program in terms of societal resources probably result in the program looking better relative to alternatives than comparisons based on governmental expenditures (e.g., “The most expensive strategy [for increasing school enrollment] among those that are frequently recommended (for example by the World Bank, which also recommends deworming) is a conditional cash transfer program, such as Progresa in Mexico, where the mother gets extra welfare payments if her children go to school. This costs about \$6,000 per additional child per year, mainly because most of the mothers who benefit from it would have sent their children to school even if there were no such incentive.” Banerjee, 2006, p. 8).

23. Attanasio, Meghir, and Santiago (2005) also develop a structural model of PROGRESA that they use to explore counterfactuals.
24. While some of the studies make some efforts at such controls through using, for example, fixed-effects estimates. However, fixed-effects estimates do not control for time-varying unobserved characteristics and idiosyncratic shocks to schools and municipalities that may be related to private enrollment patterns (e.g., as noted by Glewwe & Kremer, 2006; if people in areas experiencing negative shocks to public schools turned to private schools in response, this would produce the correlations found by Hsieh and Urquiola).
25. Families may reduce food at home to offset in part the food received from school meals, so nutrition for the students who receive school meals may not increase as much as the meals (though Jacoby, 2002 for the Philippines and Afzidi, 2007 for India find that students’ nutrition does increase considerably from school meals). Whether or not school meals partially substitute for food at home, they constitute an in-kind income incentive for parents to send their children to school.
26. This section focuses only on domestic supply of educational services to keep the scope of this chapter manageable. There also is international migration, particularly for tertiary schooling, motivated in part by postschooling labor market alternatives (e.g., see Rosenzweig, 2006a,b). Rosenzweig (2006b), for example, presents estimates consistent with the possibility that expanding quantity holding quality constant is likely to increase international migration from developing countries to developed countries for tertiary schooling and job search, but improving quality in tertiary schooling in developing countries is likely to reduce such migration.
27. This capital stock could include physical capital and/or fixed heterogeneity in educational production due, for example, to the extent of charisma or management skills of heads of institutions.
28. This model assumes that there is no agency problem in assuring that, for example, contracted teachers perform their duties per their contracts (see Section 3.2.2). Also, it assumes for simplicity that the quality measures are continuous and that only one quality measure is selected for inputs and only one for outputs.
29. To avoid unnecessary notational clutter the distinction between individual and mean genetic endowments is blurred below. While genetic endowments are used as an example here, a similar point holds for any other aspects of the educational history and current family and community educational inputs that increase current knowledge and skills.
30. Ferreyra (2003) models private schools as clubs formed by parents under an equal-sharing rule in which households of a given endowment type have incentives to segregate themselves into a private school and reject students of a lower endowment type because of peer effects in the educational production technology.
31. The study of the Bolivian program does present estimates that the benefits-to-program cost ratios are 1.7–3.7, but these are not the same as benefits-to-economic cost ratios.

32. Pitt, Rosenzweig, and Gibbons (1993) use a similar approach to investigate the determinants of program placement for a range of programs and the impact of these programs also in Indonesia. Foster and Rosenzweig (1996) present related results for India.
33. Thus, this study is an attempt to estimate production functions, as discussed in Section 2.1.3, not reduced-form demand functions as are most of the other studies discussed in this section.
34. The studies reviewed here refer to incentives for the supply side (e.g., teachers). Section 2.1.4 reviews some programs related to incentives for the demand side (e.g., students or their families). The new Mexican ALI program (*Alineando Incentivos para el Aprendizaje*), currently being evaluated over 3 years with random assignment among schools of the program (and variants of the program) and controls, attempts to align incentives for students, teachers and principals to improve student learning in mathematics by awarding students for their own and their classmates' level and improvement on mathematics tests, mathematics teachers for their own and other classes' students' level and improvement on these tests, and nonmathematics teachers and principals for the level and improvement in mathematics tests for all the students in the school.
35. It also has led to a literature addressed to the United States schooling market on family locational choices, local property-based taxes for financing schools and educational sorting; this literature seems of limited relevance for most developing countries because of different and generally more-centralized school financing mechanisms.
36. Urquiola and Verhoogen claim that these mechanisms are consistent with anecdotes from Chile, where there is a widespread perception that many lower-quality voucher schools are small “mom and pop” operations that struggle to fill their classrooms. In contrast, voucher schools run by larger firms have sufficient demand to operate multiple classrooms, and are generally perceived to be of higher quality.
37. The WHO Global Burden of Disease Estimates for 2005 indicate that for the developing world as a whole the largest aggregate of DALYS is noncommunicable diseases that used to be thought to be the diseases of developed countries, not the traditional dominant category in the developing world of communicable, maternal, perinatal, and nutritional conditions (e.g., Behrman, Behrman, & Perez, 2009).
38. Even if the life-cycle experiences are treated in the estimation as behaviorally determined, if the true specification in Eq. (1) includes all the variables indicated above and includes all three life-cycle experiences but a specification is used that excludes one or more of the relevant variables (e.g., only school years experience is included), omitted variable bias is likely to result. This is likely to be the case because on the right side of each of the three reduced-form demand relations for the three life-cycle stage experiences (Eqs. 4, 6b, and 8) are the endowments and the actual or expected values of the family, community and stochastic factors for all three life-cycle stages, which means that the three life-cycle experiences may be fairly correlated, and thus the right-side variables in Eq. (1) also fairly correlated. Of course, this is hardly surprising. *A priori*, a child with better parental family background or who lives in a better community in terms of health and educational services and job options is likely not only to have more schooling but also better pre- and postschooling experiences.
39. Direct estimates of relations such as Eqs. (1) and (2) without controlling for the behavioral determinants of the three life-cycle experiences are likely to be biased because (as indicated in the reduced-form demand relations 4, 6b, and 8) each of the three life-cycle experiences depends on all the endowments. These biases could be in either direction. For instance, the “ability bias” on which the schooling literature has focused is consistent with e_2 (schooling) being correlated positively with e_0 with the result that the coefficient of schooling is likely to be upward biased in OLS estimates of Eq. (1). On the other hand if the summary measure of preschool experience is some variable such as child stunting, and if ability and physical endowments are negatively correlated as suggested by

- Behrman, Hoddinott, et al. (2008) and Behrman and Rosenzweig (2002a, 2004), then OLS estimates of Eq. (1) may lead to biases toward zero in the coefficient estimate for schooling.
40. Note that on the right side of each of these, three reduced-form demand relations are the same endowments and the actual or expected values of the family, community, service provider, and stochastic factors for all three life-cycle stages. That means that, though there may be instruments that seem *a priori* to have first-order effects on particular life-cycle experiences (e.g., preschool programs or nutrition on e_1 , school characteristics on e_2 , labor market characteristics on e_3), it would not be correct to assert *a priori* that a particular instrument identifies a particular life-cycle educational experience. Instead, there is a potential set of instruments that hopefully identifies the set of life-cycle educational experiences. This also means that it would not be a test of the plausibility of the instruments to see if subsequent life-cycle stage family or community variables are significant (e.g., if schooling characteristics or postschooling labor market characteristics significantly determine preschool educational experience e_1) because the expected value of those variables should be included. Instead it might be a test of to what extent expectations are rational in the sense that the expected values for subsequent stages are equal to the realized values.
 41. The procedure is (1) to estimate a logit for whether an entity (i.e., individual, household, educational service provider, community) was exposed to treatment as a function of variables not affected by the treatment, (2) to use the estimates to predict the latent propensity for treatment for every entity, and (3) to compare each entity treated with an entity or group of entities not-treated but who are very similar in terms of the predicted latent propensity for being treated. This permits comparisons between very similar entities who have received and who have not received treatment, where similarity is defined in terms of the weighted average of *observed* characteristics used to predict the propensity to be treated. Unobserved fixed factors, such as those discussed below, are also controlled in some matching estimates (e.g., studies of the impact of preschool programs on early childhood development in Bolivia in Behrman, Cheng, & Todd, 2004 and in the Philippines in Ghuman et al., 2006a; studies of the impact of conditional cash transfers on school enrollment, achievement tests and labor market time in Mexico in Behrman, Parker, & Todd, 2007, 2009b).
 42. Other recent studies for European countries also report that OLS estimates of intergenerational schooling effects may be quite misleading. For instance, Plug (2004) uses data on adoptees to lessen problems of intergenerationally correlated endowments and Black et al. (2005) use instruments based on changes in mandatory schooling.
 43. The “noise-to-signal” ratio refers to the fact that most concepts are not measured perfectly, particularly in self-reported data, but have some random measurement error (leaving aside for the moment systematic measurement error). This measurement error is referred to as “noise” (since it disguises or hides the systematic part or “signal” in the data). The variance in variable as measured therefore can be decomposed into the variance due to noise and the variance due to the signal, with higher “noise-to-signal” indicating more contamination due to random measurement error.
 44. The conventional wisdom held by some seems to be that this evidence is overwhelming. But a now-dated survey of all the estimates that could be located of associations between parental schooling and child schooling found that larger estimates were reported for mothers’ schooling than for fathers’ schooling in 52% of the cases (Behrman, 1997). Moreover, as noted above, estimates for mothers’ schooling may be more upward biased due to omitted variable biases if women alter their time use more with schooling than do men, as seems to be the case in many societies (Behrman & Rosenzweig, 2002a, 2005 give an example for the United States).
 45. An exception is the study by Rosenzweig and Wolpin (1995) on the impact of maternal schooling on preschool cognition in the United States for young mothers who went to school between births in which they controlled for all unobserved maternal fixed characteristics.

46. Data on such assets may also permit the construction of wealth indices in the absence of income measures, such as the use of principle components of such assets for wealth as in the INCAP studies in Guatemala (e.g., Pollitt, Gorman, Engle, Martorell, & Rivera, 1993; as was subsequently popularized by Filmer & Pritchett, 2001, using the DHS data for India). The economic interpretation of such indices, at least in comparison with weighting assets by their prices, is not clear.
47. These last two examples are improvements over previous studies that used parental schooling or income to see if there were differential impacts between mothers' and fathers' resources and investments in children, but were difficult to interpret because these indicators are likely to be associated not only with control over resources but also time use and unobserved abilities and motivations. However, although these two examples probably are improvements, it is not clear that the representations that they use are independent of time uses and unobserved characteristics, so they may be subject to a weaker version of the same problem.
48. Since such experiments almost always have baseline and postintervention data rounds, they are longitudinal and not cross sectional. In principle if the treatment and control groups are randomly selected then only looking at the cross-sectional posttreatment data should be informative. But it would not be possible in such a case to test whether or not the assignment really was random (as, e.g., in Behrman & Todd, 1999a for the Mexican PROGRESA data).
49. At the cost of the assumptions necessary to estimate structural models of the behaviors such as are outlined in Sections 3.2.1 and 3.2.2, evaluations of counterfactual policies can be made (e.g., different treatments, impacts for longer time periods than observed in the data). Todd and Wolpin (2006) provide an example using the Mexican PROGRESA data. They estimate a structural model using baseline data, then test the model's predictions against the experimental results (and find that the model predicts fairly well), and the use the model to conduct counterfactual experiments (e.g., with different scholarship schedules for different grades, with the program running many years).
50. But see Deaton (2009) for a recent questioning of whether experiments have the advantages with regard to policy evaluation often claimed by their advocates.

References

- Aboud, F. E. (2006). Evaluation of an early childhood preschool programme in rural Bangladesh. *Early Childhood Research Quarterly*, 21(1), 46–60.
- Afridi, F. (2007). *Child welfare programs and child nutrition: Evidence from a mandated school meal program in India*. Syracuse, NY: Syracuse University.
- Alderman, H., Behrman, J. R., Kohler, H. P., Maluccio, J. A., & Watkins, S. C. (2001). Attrition in longitudinal household survey data: Some tests for three developing country samples. *Demographic Research*, 5(4), 79–124. <http://www.demographic-research.org>.
- Alderman, H., Behrman, J., Lavy, V., & Menon, R. (2001). Child health and school enrollment: A longitudinal analysis. *Journal of Human Resources*, 36(1), 185–205.
- Alderman, H., Behrman, J. R., Ross, D., & Sabot, R. (1996). The returns to endogenous human capital in Pakistan's rural wage labor market. *Oxford Bulletin of Economics and Statistics*, 58(1), 29–55.
- Alderman, H., & Engle, P. (2007). *The synergy of nutrition and ECD interventions in Africa. Africa's future, Africa's challenge*. (pp. 355–370). Washington, DC: World Bank.
- Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long-term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58, 450–474.
- Angrist, J., Bettinger, E., & Kremer, M. (2006, June). Long-term consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96(3), 847–862.
- Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on children's academic achievement. *Quarterly Journal of Economics*, 114(2), 533–576.
- Angrist, J. D., & Lavy, V. (2002). *The effect of high school matriculation awards: Evidence from randomized trials*. NBER Working Paper No. 9389. Cambridge, MA: NBER.

- Arabsheibani, R., Carneiro, F. G., & Henley, A. (2003). *Human capital and earnings inequality in Brazil, 1999–1998: Quantile regression evidence*. World Bank Policy Research Working Paper No. 3147. Washington, DC: World Bank.
- Armecin, G., Behrman, J. R., Duazo, P., Ghuman, S., Gultiano, S., King, E. M., et al. (2006). *Early childhood development through integrated programs: Evidence from the Philippines*. Policy Research Working Paper No. 3922. Washington, DC: World Bank.
- Ashenfelter, O., & Krueger, A. (1994, December). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review*, 84(5), 1157–1174.
- Attanasio, O., Gomez, L. C., Gomez Rojas, A., & Vera-Hernández, M. (2004, December). Child health in rural Colombia: Determinants and policy interventions. *Economics & Human Biology*, 2(3), 411–438.
- Attanasio, O., Meghir, C., & Santiago, A. (2005). *Education choices in Mexico: Using a structural model and a randomized experiment to evaluate Progresa*. (Vol. EWP05/01). London: Institute for Fiscal Studies.
- Auguste, S., & Valenzuela, J. P. (2003). *Do students benefit from school competition? Evidence from Chile*. Working Paper. Ann Arbor: University of Michigan.
- Banerjee, A. V. (2006, July/August). Making aid work: How to fight global poverty—Effectively. In A. V. Banerjee, A. H. Amsden, R. H. Bates, J. Bhagwati, A. Deanton & N. Stern (Eds.), *Making aid work. Boston review*. (pp. 111–122). Cambridge, MA: MIT Press.
- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remediating education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3), 1235–1264.
- Banerjee, A. V., & Duflo, E. (2006, Winter). Addressing absence. *Journal of Economic Perspectives*, 20(1), 117–132.
- Banerjee, A. V., Jacob, S., Kremer, M., Lanjouw, J., & Lanjouw, P. (2000). *Promoting school participation in rural Rajasthan: Results from some prospective trials*. Cambridge, MA: MIT Press.
- Barker, D. J. P. (1992). Fetal and infant origins of adult disease. Papers written by the Medical Research Council Environmental Epidemiology Unit, University of Southampton. British Medical Journal 1998. In D. J. P. Barker (Ed.), *Mothers, babies and health in later life*. (2nd ed.). Edinburgh: Churchill Livingstone.
- Becker, G. S. (1975). *Human capital and the personal distribution of income: An analytical approach*. Ann Arbor: University of Michigan. Woytinsky Lecture, republished in Becker, G. S. (1975). *Human capital* (pp. 94–117). (2nd ed.). New York: NBER.
- Behrman, J. R. (1988, March). Intrahousehold allocation of nutrients in rural India: Are boys favored? Do parents exhibit inequality aversion? *Oxford Economic Papers*, 40(1), 32–54.
- Behrman, J. R. (1997). *Women's schooling and child education: A survey*. Philadelphia, PA: University of Pennsylvania.
- Behrman, J. R. (2007). *Policy-oriented research impact assessment (PORIA) case study on the International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA anti-poverty and human resource investment conditional cash transfer program*. Impact Assessment Discussion Paper No. 27. Washington, DC: International Food Policy Research Institute.
- Behrman, J. R., Alderman, H., & Hoddinott, J. (2004). Hunger and malnutrition. In B. Lomborg (Ed.), *Global crises, global solutions* (pp. 363–420). Cambridge, UK: Cambridge University Press.
- Behrman, J. R., Behrman, J. A., & Perez, N. (2009). On what diseases and health conditions should new economic research on health and development focus? *Health Economics*, (published online in Wiley InterScience: <http://www.interscience.wiley.com>) doi:10.1002/hec.1464.
- Behrman, J. R., & Birdsall, N. (1983). The quality of schooling: Quantity alone is misleading. *American Economic Review*, 73, 928–946.
- Behrman, J. R., & Birdsall, N. (1988). The reward for good timing: Cohort effects and earnings functions for Brazilian males. *Review of Economics and Statistics*, 70(1), 129–135.
- Behrman, J. R., Cheng, Y., & Todd, P. E. (2004, February). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *Review of Economics and Statistics*, 86(1), 108–132.
- Behrman, J. R., Duryea, S., & Székely, M. (2003). Aging and economic opportunities: Major world regions around the turn of the century. In O. Attanasio & M. Székely (Eds.), *A dynamic analysis of household decision-making in Latin America* (pp. 69–100). Washington, DC: Inter-American Development Bank.

- Behrman, J. R., Foster, A., & Rosenzweig, M. R. (1997, March). The dynamics of agricultural production and the calorie-income relationship: Evidence from Pakistan. *Journal of Econometrics*, 77(1), 187–207.
- Behrman, J. R., & Hoddinott, J. (2005). Program evaluation with unobserved heterogeneity and selective implementation: The Mexican Progresa impact on child nutrition. *Oxford Bulletin of Economics and Statistics*, 67(4), 547–569.
- Behrman, J. R., Hoddinott, J., Maluccio, J. A., & Martorell, R. (2009). *Brains versus brawn: Labor market returns to intellectual and physical human capital in a poor developing country*. Philadelphia, PA: University of Pennsylvania.
- Behrman, J. R., Hoddinott, J., Maluccio, J. A., Soler-Hampejsek, E., Behrman, E. L., Martorell, R., et al. (2008). *What determines adult skills? Impacts of pre-school, school-years and post-school experiences in Guatemala*. Discussion Paper No. 826. Washington, DC: International Food Policy Research Institute (IFPRI).
- Behrman, J. R., & King, E. M. (2008). Program impact and variation in duration of exposure. In S. Amin, J. Das & M. Goldstein (Eds.), *Are you being served: New tools for measuring service delivery* (pp. 147–172). Washington, DC: World Bank.
- Behrman, J. R., & Knowles, J. C. (1999, May). Household income and child schooling in Vietnam. *World Bank Economic Review*, 13(2), 211–256.
- Behrman, J. R., Murphy, A., Quisumbing, A., Ramakrishna, U., & Young, K. (2008). *Are returns to mothers' human capital realized in the next generation? The impact of mothers' intellectual human capital and long-run nutritional status on child human capital in Guatemala*. Atlanta/Philadelphia/Washington: Emory University/University of Pennsylvania/International Food Policy Research Institute.
- Behrman, J. R., Parker, S. W., & Todd, P. E. (2007). *Do school subsidy programs generate lasting benefits? A five-year follow-up of Oportunidades participants*. Philadelphia, PA/Mexico City: University of Pennsylvania/CIDE, mimeo.
- Behrman, J. R., Parker, S. W., & Todd, P. E. (2009a). Medium-term impacts of the *Oportunidades* conditional cash transfer program on rural youth in Mexico. In S. Klasen & F. Nowak-Lehmann (Eds.), *Poverty, inequality, and policy in Latin America* (pp. 219–270). Cambridge, MA: MIT Press.
- Behrman, J. R., Parker, S. W., & Todd, P. E. (2009b). Schooling impacts of conditional cash transfers on young children: Evidence from Mexico. *Economic Development and Cultural Change*, 57(3), 439–477.
- Behrman, J. R., & Rosenzweig, M. R. (1999). Ability' biases in schooling returns and twins: A test and new estimates. *Economics of Education Review*, 18, 159–167.
- Behrman, J. R., & Rosenzweig, M. R. (2002a, March). Does increasing women's schooling raise the schooling of the next generation? *American Economic Review*, 92(1), 323–334.
- Behrman, J. R., & Rosenzweig, M. R. (2002b). *The returns to schooling and work experience for women and men: Estimates using twins*. Philadelphia, PA: University of Pennsylvania.
- Behrman, J. R., & Rosenzweig, M. R. (2004, May). Returns to birth weight. *Review of Economics and Statistics*, 86(2), 586–601.
- Behrman, J. R., & Rosenzweig, M. R. (2005, December). Does increasing women's schooling raise the schooling of the next generation?—Reply. *American Economic Review*, 95(5), 1745–1751.
- Behrman, J. R., Rosenzweig, M. R., & Taubman, P. (1994, December). Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy*, 102(6), 1131–1174.
- Behrman, J. R., Ross, D., & Sabot, R. (2008). Improving quality versus increasing the quantity of schooling: Estimates of rates of return from rural Pakistan. *Journal of Development Economics*, 85, 94–104.
- Behrman, J. R., Sengupta, P., & Todd, P. E. (2005, October). Progressing through PROGRESA: An impact assessment of Mexico's school subsidy experiment. *Economic Development and Cultural Change*, 54(1), 237–275.
- Behrman, J. R., & Todd, P. E. (1999). *Randomness in the experimental samples of PROGRESA (Education, Health and Nutrition Program)*. Washington, DC: International Food Policy Research Institute (IFPRI).
- Behrman, J. R., & Wolfe, B. L. (1984, May). The socioeconomic impact of schooling in a developing country. *Review of Economics and Statistics*, 66(2), 296–303.

- Behrman, J. R., & Wolfe, B. L. (1987a). How does mother's schooling affect the family's health, nutrition, medical care usage, and household sanitation? *Journal of Econometrics*, 36, 185–204.
- Behrman, J. R., & Wolfe, B. L. (1987b, October). Investments in schooling in two generations in pre-revolutionary Nicaragua: The roles of family background and school supply. *Journal of Development Economics*, 27(1–2), 395–420.
- Berlinski, S., Galiani, S., & Gertler, P. (2006). *The effect of pre-primary education on primary school performance*. London/Buenos Aires/Washington: University College London and Institute for Fiscal Studies/Universidad de San Andres/World Bank.
- Bernheim, B. D., Shleifer, A., & Summers, L. H. (1985, December). The strategic bequest motive. *Journal of Political Economy*, 93(6), 1045–1076.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249–275.
- Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics*, 62(3), 395–407.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005, March). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American Economic Review*, 95(1), 437–449.
- Bloom, D. E., & Canning, D. (2004). *Global demographic change: Dimensions and economic significance*. NBER Working Paper No. 10817. Cambridge, MA: NBER.
- Bloom, D. E., Canning, D., & Sevilla, J. (2002). *The demographic dividend: A new perspective on the economic consequences of population change*. Santa Monica, CA: RAND Corporation.
- Boissiere, M., Knight, J. B., & Sabot, R. H. (1985). Earnings, schooling, ability and cognitive skills. *American Economic Review*, 75, 1016–1030.
- Bommier, A., & Lambert, S. (2000). Education demand and age at school enrollment in Tanzania. *Journal of Human Resources*, 35(1), 177–203.
- Boyd, D., Lankford, H., Loeb, S., & Wyckhoff, J. (2005, May). Explaining the short careers of high achieving teachers in schools with low-performing students. *American Economic Review*, 95, 166–171.
- Bravo, D., Behrman, J. R., Mitchell, O., & Todd, P. E. (2006). *Encuesta de Protección Social 2004: Presentación General y Resultados Principales*. Santiago, Chile: Centro de Microdatos, Universidad de Chile.
- Bravo, D., Contreras, D., & Sanhueza, C. (1999). *Rendimiento escolar, desigualdad y brecha de desempeño privado/público: Chile 1982–1997*. Documento de Trabajo No. 163. Santiago, Chile: Departamento de Economía, Universidad de Chile.
- Bravo, D., Mukhopadhyay, S., & Todd, P. E. (2008). *Effects of a universal school voucher system on educational and labor market outcomes: Evidence from Chile*. Philadelphia, PA: University of Pennsylvania.
- Buchinsky, M. (1998). Recent advances in quantile regression model: A practical guide for empirical research. *Journal of Human Resources*, 33(1), 88–126.
- Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics*. Amsterdam: Elsevier Science, BV.
- Card, D., & Krueger, A. (1992a). School quality and black-white relative earnings: A direct assessment. *Quarterly Journal of Economics*, 107, 151–200.
- Card, D., & Krueger, A. (1992b). Does school quality matter: Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100(1), 1–40.
- Case, A., & Deaton, A. (1999, August). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*, 114(3), 1047–1084.
- Cebu Study Team. (1991). Underlying and proximate determinants of child health: The Cebu Longitudinal Health and Nutrition Study. *American Journal of Epidemiology*, 133(2), 185–201.
- Chin, A. (2005, December). Can redistributing teachers across schools raise educational attainment? Evidence from Operation Blackboard in India. *Journal of Development Economics*, 78(2), 384–405.
- Contreras, D. (2001). *Evaluating a voucher system in Chile: Individual, family and school characteristics*. Santiago, Chile: University of Chile.
- Cueto, S., & Diaz, J. (1999). Impacto de la educación inicial en el rendimiento en primer grado de primaria en escuelas públicas urbanas de Lima. *Revista de Psicología*, 17, 74–91.
- Cunha, F. (2006). *A time to plant and a time to reap*. Chicago: University of Chicago.

- Cunha, F., & Heckman, J. J. (2006). *Formulating, identifying and estimating the technology of cognitive and non-cognitive skill formation*. Chicago: University of Chicago.
- Daniels, M. C., & Adair, L. S. (2004). Growth in young Filipino children predicts schooling trajectories through high school. *Journal of Nutrition*, 134, 1439–1446.
- Deaton, A. (2009). *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development*. Princeton, NJ: Research Program in Development Studies and Center for Health and Wellbeing, Princeton University.
- Deaton, A., & Paxson, C. (1994, June). Intertemporal choice and inequality. *Journal of Political Economy*, 102(3), 437–467.
- Deolalikar, A. B., & Gaiha, R. (1993, December). Persistent, expected and innate poverty: Estimates for semi-arid rural South India, 1975–84. *Cambridge Journal of Economics*, 17, 409–421.
- Ding, W., & Lehrer, S. F. (2007, May). Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics*, 84(2), 300–312.
- Dolton, P. J., & Van der Klaauw, W. (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics*, 81(3), 543–552.
- Drèze, J., & Kingdon, G. (2001). School participation in rural India. *Review of Development Economics*, 5(1), 1–24.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4), 795–813.
- Duflo, E. (2004). The medium run consequences of educational expansion: Evidence from a large school construction program in Indonesia. *Journal of Development Economics*, 74(1), 163–197.
- Duflo, E., & Hanna, R. (2006). *Monitoring works: Getting teachers to come to school*. Cambridge, MA: Department of Economics and Poverty Action Lab, Massachusetts Institute of Technology.
- Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M., Gertler, P. J., Kapiriri, L., et al. (2007, January). Strategies to avoid the loss of potential among 240 million children in the developing world. *Lancet*, 369, 229–242.
- Apple, D., & Ferreyra, M. M. (2006). *School finance reform in Michigan: Assessing general equilibrium effects*. Pittsburgh, PA: Tepper School of Business, Carnegie Mellon University.
- Ferreyra, M. M. (2003). *Estimating the effects of private school vouchers in multi-district economies*. Pittsburgh, PA: Graduate School of Industrial Administration, Carnegie Mellon University.
- Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—Or tears: An application to educational enrolments in states of India. *Demography*, 38(1), 115–132.
- Fiszbein, A., Giovanoli, P. I., & Patrinos, H. A. (2004). *Estimating the returns to education in Argentina: 1992–2002*. Washington, DC: World Bank.
- Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998a). An analysis of sample attrition in panel data. *Journal of Human Resources*, 33(2), 251–299.
- Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998b). The impact of attrition in the PSID on intergenerational analysis. *Journal of Human Resources*, 33(2), 300–344.
- Foster, A. D., & Rosenzweig, M. R. (1995, December). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6), 1176–1209.
- Foster, A. D., & Rosenzweig, M. R. (1996). Technical change and human capital returns and investments: Evidence from the Green Revolution. *American Economic Review*, 86(4), 931–953.
- Galiani, S., Gertler, P., & Schargrodsky, E. (2008). School decentralization: Helping the good get better, but leaving the poor behind. *Journal of Public Economics*, 92, 2106–2120.
- Galiani, S., Gertler, P., & Schargrodsky, E. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93, 219–234.
- Galiani, S., Heymann, D., Dabús, C., & Tohmé, F. (2008). On the emergence of public education in land-rich economies. *Journal of Development Economics*, 86, 434–446.
- Galor, O., & Moav, O. (2006). Das human-kapital: A theory of the demise of the class structure. *Review of Economic Studies*, 73(1), 85–117.
- Ghuman, S., Behrman, J., Borja, J. B., Gultiano, S., & King, E. M. (2005, October). Family background, service providers, and early childhood development in the Philippines: Proxies and interactions. *Economic Development and Cultural Change*, 54(1), 129–164.

- Ghuman, S., Behrman, J. R., Gultiano, S., Armecein, G., Bas, I., Duazo, P., et al. (2006, August). Early childhood development: A review of findings from a longitudinal study in the Philippines. *East Asian Economic Papers*, 17(2), 1–24.
- Ghuman, S., Behrman, J. R., Gultiano, S., & King, E. M. (2006). *Children's nutrition, school quality, and primary school enrollment in the Philippines*. New York, NY: Population Council.
- Girma, S., & Kedir, A. (2003). Is education more beneficial to the less able? Econometric evidence from Ethiopia. Discussion Papers in Economics from Department of Economics No. 03/1. Leicester, UK: University of Leicester.
- Glewwe, P. (1996, December). The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development Economics*, 51(2), 267–290.
- Glewwe, P., Grosh, M., Jacoby, H., & Lockheed, M. (1995). An eclectic approach to estimating the determinants of achievement in Jamaican primary education. *World Bank Economic Review*, 9(2), 231–258.
- Glewwe, P., Ilias, N., & Kremer, M. (2004). *Teacher incentives*. Cambridge, MA: Harvard University.
- Glewwe, P., & Jacoby, H. (1994, Summer). Student achievement and schooling choice in low income countries: Evidence from Ghana. *Journal of Human Resources*, 29(3), 842–864.
- Glewwe, P., Jacoby, H., & King, E. (2000). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics*, 81(3), 345–368.
- Glewwe, P., & King, E. (2001). The impact of early childhood nutrition status on cognitive achievement: Does the timing of malnutrition matter? *World Bank Economic Review*, 15(1), 81–114.
- Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2). Amsterdam: Elsevier.
- Glewwe, P., Kremer, M., & Moulin, S. (2007). *Many children left behind? Textbooks and test scores in Kenya*. MIT Poverty Action Lab Paper No. 44. Cambridge, MA: MIT Press.
- Grantham-McGregor, S. M., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L. M., & Strupp, B. J. (2007, January). Over two hundred million children fail to reach their developmental potential in the first five years in developing countries. *Lancet*, 369, 60–70.
- Grantham-McGregor, S., Fernald, L., & Sethuraman, K. (1999a). Effects of health and nutrition on cognitive and behavioural development in children in the first three years of life: Part 1: Low birthweight, breastfeeding and protein-energy malnutrition. *Food and Nutrition Bulletin*, 20, 53–75.
- Grantham-McGregor, S., Fernald, L., & Sethuraman, K. (1999b). Effects of health and nutrition on cognitive and behavioural development in children in the first three years of life: Part 2: Infections and micronutrient deficiencies: Iodine, iron and zinc. *Food and Nutrition Bulletin*, 2, 76–99.
- Grantham-McGregor, S., Walker, C., Chang, S., & Powell, C. (1997). Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. *American Journal of Clinical Nutrition*, 66, 247–253.
- Hanushek, E. A. (1995, August). Interpreting recent research on schooling in developing countries. *World Bank Research Observer*, 10(2), 227–246.
- Hanushek, E. A., Kaine, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39(2), 326–354.
- Heckman, J. J. (1974, July). Shadow prices, market wages, and labor supply. *Econometrica*, 42, 679–694.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153–161.
- Heckman, J. J. (2006, June 30). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900–1902.
- Heckman, J., Ichimura, H., & Todd, T. (1998, April). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), 261–294.
- Holmstrom, B., & Milgrom, P. (1991). Multi-task principal-agent analysis: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24–52.
- Hsieh, C., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90, 1477–1503.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, 1, 221–233.

- Jacoby, H. G. (2002). Is there an intrahousehold ‘flypaper effect’? Evidence from a school feeding programme. *Economic Journal*, 112, 196–221.
- Jacoby, H., & Skoufias, E. (1997, July). Risk, financial markets and human capital in a developing country. *Review of Economic Studies*, 64(3), 311–335.
- Jaramillo, A., & Tietjen, K. (2002). *Early childhood development in Africa: Can we do more or less? A look at the impact and implications of preschools in Cape Verde and Guinea*. Washington, DC: World Bank.
- Jenkins, S. P., & Siedler, T. (2006). *Using household panel data to understand the intergenerational transmission of poverty*. Colchester, UK: University of Essex Institute for Social and Economic Research.
- Jimenez, E., & Sawada, Y. (1999). Do community-managed schools work? An evaluation of El Salvador’s EDUCO Program. *World Bank Economic Review*, 13(3), 415–441.
- Johnston, F., Low, S., de Baessa, Y., & MacVean, R. (1987). Interaction of nutritional and socioeconomic status as determinants of cognitive achievement in disadvantaged urban Guatemalan children. *American Journal of Physical Anthropology*, 73(4), 501–506.
- Kagitcibasi, C., Sunar, D., & Bekman, S. (2001). Long-term effects of early intervention: Turkish low-income mothers and children. *Journal of Applied Developmental Psychology*, 22(4), 333–361.
- Karlan, D., & Valdivia, M. (2006). *Teaching entrepreneurship: Impact of business training on microfinance clients and institutions*. New Haven, CN/Lima, Peru: Yale University, Innovations for Poverty Action, and Jameel Poverty Action Lab/Grupo de Análisis para el Desarrollo.
- King, E. M., & Behrman, J. R. (2009). Timing and duration of exposure in evaluations of social programs. *World Bank Research Observer*, 24(1), 55–82.
- King, E. M., & Lillard, L. A. (1987, November). Education policy and schooling attainment in Malaysia and the Philippines. *Economics of Education Review*, 6(2), 167–181.
- Kingdon, G. (1996). The quality and efficiency of public and private schools: A case study of urban India. *Oxford Bulletin of Economics and Statistics*, 58(1), 55–80.
- Knowles, J. C., & Behrman, J. R. (2005). Economic returns to investing in youth. In J. R. Behrman, B. Cohen, C. Lloyd & N. Stromquist (Eds.), *The transition to adulthood in developing countries: Selected studies* (pp. 424–490). Washington, DC: National Academy of Science-National Research Council.
- Kremer, M. (1995). Research on schooling: What we know and what we don’t—A comment on Hanushek. *World Bank Research Observer*, 10(2), 247–254.
- Kremer, M., Moulin, S., & Namunyu, R. (2002). *Unbalanced decentralization*. Cambridge, MA: Harvard University.
- Lam, D., & Schoeni, R. F. (1993). Effects of family background on earnings and returns to schooling: Evidence from Brazil. *Journal of Political Economy*, 101(4), 710–740.
- Lam, D., & Schoeni, R. F. (1994). Family ties and labor markets in the United States and Brazil. *Journal of Human Resources*, 29(4), 1235–1258.
- Lasky, R., Klein, R., Yarborough, C., Engle, P., Lechtig, A., & Martorell, R. (1981). The relationship between physical growth and infant development in rural Guatemala. *Child Development*, 52(2), 219–226.
- Lavy, V. (1996). School supply constraints and children’s educational outcomes in rural Ghana. *Journal of Development Economics*, 51(2), 291–314.
- Lavy, V. (2002). Evaluating the effect of teachers’ performance incentives on pupils’ achievements. *Journal of Political Economy*, 110(6), 1286–1318.
- Lee, B. J., & Lee, M. J. (2002). *Quantile regression analysis of wage determinants in the Korean labor market*. Working Paper. Notre Dame, IN: Department of Economics, University of Notre Dame.
- Levy, S. (2006). *Progress against poverty: Sustaining Mexico’s PROGRESA-Oportunidades Program*. Washington, DC: Brookings Institution.
- Levy, S., & Rodriguez, E. (2004). *Economic crisis, political transitions, and poverty policy reform: Mexico’s Progresa-Oportunidades Program*. Policy dialogue series. Washington, DC: Inter-American Development Bank.
- Linden, L., & Shastry, K. (2005). *Grain inflation: Misrepresenting student attendance in response to a school meals program*. Cambridge, MA: Harvard University.
- Maccini, S., & Yang, D. (2009). Under the weather: Health, schooling, and socioeconomic consequences of early-life rainfall. *American Economic Review* (in press).

- Maluccio, J. A., Hoddinott, J., Behrman, J. R., Quisumbing, A., Martorell, R., & Stein, A. D. (2009). The impact of nutrition during early childhood on education among Guatemalan adults. *Economic Journal*, 119(537), 734–763.
- Martinelli, C., & Parker, S. W. (2003, May). Should transfers to poor families be conditional on school attendance? A household bargaining perspective. *International Economic Review*, 44(2), 523–544.
- Martins, P. S., & Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics*, 11, 355–371.
- Martorell, R., Behrman, J. R., Flores, R., & Stein, A. D. (2005). Rationale for a follow-up focusing on economic productivity. *Food and Nutrition Bulletin*, 26(2 Suppl. 1), S5–S14.
- McEwan, P. J. (2001). The effectiveness of public, Catholic, and non-religious private schools in Chile's voucher system. *Education Economics*, 9(2), 103–128.
- McEwan, P. J., & Carnoy, M. (2000). The effectiveness and efficacy of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis*, 22(3), 213–239.
- McKay, H., Sinisterra, L., McKay, A., Gomez, H., & Lloreda, P. (1978). Improving cognitive ability in chronically deprived children. *Science*, 200(4339), 270–278.
- Mensch, B. S., Hewett, P. C., & Erulkar, A. (2003). The reporting of sensitive behavior among adolescents: A methodological experiment in Kenya. *Demography*, 40(2), 247–268.
- Metzler, J., & Woessmann (Wößmann), L. (2009). *The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation*. Munich, Germany: University of Munich and Ifo Institute for Economic Research.
- Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159–217.
- Mizala, A., & Romaguera, P. (1999). School performance and choice: The Chilean experience. *Journal of Human Resources*, 35(2), 392–417.
- Montenegro, C. (2001). *Wage distribution in Chile: Does gender matter? A quantile regression approach*. Washington, DC: The World Bank, Poverty Reduction and Economic Management Network.
- Morenza, L., Arrazola, O., Seleme, I., & Martinez, F. (2005). *Evaluacion Proyecto Kallpa Wawa* (pp. 1–130). Bolivia: Santa Cruz. http://www.unicef.org/earlychildhood/index_resources.html.
- Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro-units. *Review of Economics and Statistics*, 72, 334–338.
- Munshi, K., & Rosenzweig, M. R. (2006). Traditional institutions meet the modern world: Caste, gender and schooling choice in a globalizing economy. *American Economic Review*, 96(4), 1225–1252.
- Mwabu, G., & Schultz, T. P. (1996). Education returns across quantiles of the wage function: Alternative explanations for returns to education by race in South Africa. *American Economic Review*, 86(2), 335–339.
- Neal, D. (1997). The effect of Catholic secondary schooling on educational achievement. *Journal of Labor Economics*, 15(1), 98–123.
- Orazem, P. F., & King, E. M. (2008). Schooling in developing countries: The roles of supply, demand and government policy. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4 pp. 3475–3560). Amsterdam: Elsevier.
- Parker, S. W., Todd, P. E., & Wolpin, K. I. (2006). *Within-family treatment effect estimators: The impact of Oportunidades on schooling in Mexico*. Philadelphia, PA: University of Pennsylvania.
- Patrinos, H., & Metzer, S. (2004). *Returns to education in Mexico: An update*. Washington, DC: World Bank.
- Patrinos, H., & Sakellariou, C. (2004). *Economic volatility and returns to education in Venezuela*. World Bank Policy Research Working Paper No. 3459. Washington, DC: World Bank.
- Paxson, C. H. (1992, March). Using weather variability to estimate the response of savings to transitory income in Thailand. *American Economic Review*, 82(1), 15–33.
- Paxson, C., & Schady, N. R. (2007, Winter). Cognitive development among young children in Ecuador: The roles of wealth, health and parenting. *Journal of Human Resources*, 42(1), 49–84.
- Pereira, P. T., & Martins, P. S. (2000). *Does education reduce wage inequality? Quantile regressions evidence from fifteen European countries*. IZA Discussion Paper No. 120. Bonn, Germany: Institute for the Study of Labour.

- Pitt, M. M., Rosenzweig, M. R., & Gibbons, D. M. (1993, September). The determinants and consequences of the placement of government programs in Indonesia. *World Bank Economic Review*, 7(3), 319–348.
- Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (1990, December). Productivity, health and inequality in the intrahousehold distribution of food in low-income countries. *American Economic Review*, 80(5), 1139–1156.
- Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2006). *Sharing the burden of disease: Gender, the household division of labor and the health effects of indoor air pollution in Bangladesh and India*. New Haven, CN: Yale University.
- Plug, E. (2004, March). Estimating the effect of mother's schooling on children's schooling using a sample of adoptees. *American Economic Review*, 94(1), 358–368.
- Plug, E., & Vijverberg, W. P. M. (2003). Schooling, family background and adoption: Is it nature or nurture? *Journal of Political Economy*, 111, 611–641.
- Pollitt, E., & Escamilla, R. (1996). *Developmental timing as moderator in the rehabilitation of malnourished children: The Cali study*. http://www.unicef.org/earlychildhood/index_resources.html.
- Pollitt, E., Gorman, K. S., Engle, P., Martorell, R., & Rivera, J. A. (1993). Early supplementary feeding and cognition: Effects over two decades. *Monographs of the Society for Research in Child Development, Serial No. 235*, 58(7), 1–99.
- Powell, C. (2004). *An evaluation of the roving caregivers programme of the rural family support organization*. Clarendon, Jamaica: May Pen, UNICEF. http://www.unicef.org/earlychildhood/index_resources.html.
- Powell, C., Baker-Henningham, H., Walker, S., Gernay, J., & Grantham-McGregor, S. (2004). Feasibility of integrating early stimulation into primary care for undernourished Jamaican children: Cluster randomised controlled trial. *British Medical Journal*, 329(7457), 89.
- Psacharopoulos, G. (1985). Returns to education: A further international update and implications. *Journal of Human Resources*, 20(4), 583–604.
- Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, 22, 1325–1344.
- Psacharopoulos, G., & Patrinos, H. (2004). Returns to investment in education: A further update. *Educational Economics*, 12, 111–134.
- Quisumbing, A. R., & Maluccio, J. A. (2003). Resources at marriage and intrahousehold allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa. *Oxford Bulletin of Economics and Statistics*, 65(3), 283–328.
- Reinikka, R., & Svensson, J. (2003). *The power of information: Evidence from a campaign to reduce capture*. Washington, DC: World Bank.
- Rivera, J. A., Sotres-Alvarez, D., Habicht, J. P., Shamah, T., & Villalpando, S. (2004). Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children. *Journal of the American Medical Association*, 291, 2563–2570.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rosenzweig, M. R. (1995, May). Why are there returns in schooling? *American Economic Review*, 85(2), 153–158.
- Rosenzweig, M. R. (2006a). Global wage differences and international student flows. In S. M. Collins & C. Graham (Eds.), *Brookings Trade Forum 2006: Global labor markets*. Washington, DC: Brookings Institution Press.
- Rosenzweig, M. R. (2006b). *Higher education and international migration in Asia: Brain circulation*. New Haven, CN: Yale University.
- Rosenzweig, M. R., & Wolpin, K. I. (1986). Evaluating the effects of optimally distributed public programs: Child health and family planning interventions. *American Economic Review*, 76(3), 470–482.
- Rosenzweig, M. R., & Wolpin, K. J. (1988, December). Migration selectivity and the effects of public programs. *Journal of Public Economics*, 37(2), 265–289.
- Rosenzweig, M. R., & Wolpin, K. I. (1995). Sisters, siblings and mothers: The effect of teen-age childbearing on birth outcomes in a dynamic family context. *Econometrica*, 63(2), 303–326.

- Rubalcava, L. N., & Teruel, G. M. (2004, December). The role of maternal cognitive ability on child health. *Economics & Human Biology*, 2(3), 439–455.
- Sakellarisou, C. (2005). *Heterogeneity in the returns to education and experience: Evidence from a high and a low income S.E. Asian country*. Singapore: Nanyang Technological University.
- Sapelli, C., & Vial, B. (2002). The performance of private and public schools in the Chilean voucher system. *Cuadernos de Economía—Latin American Journal of Economics*, 39(118), 423–454.
- Save the Children. (2003) *What's the difference? The impact of early childhood development programmes. A study from Nepal of the effects for children, their families and communities*. Kathmandu, Nepal. http://www.unicef.org/earlychildhood/index_resources.html.
- Save the Children. (2004) *Early childhood care and development: A positive impact*. Myanmar: Save the Children Myanmar Field Office. http://www.unicef.org/earlychildhood/index_resources.html.
- Schultz, T. W. (1975, September). The value of the ability to deal with disequilibria. *Journal of Economic Literature*, 13(3), 827–846.
- Schultz, T. P. (1988). Education investments and returns. In H. Chenery & T. N. Srinivasan (Eds.), *Handbook of development economics* (pp. 543–630). Amsterdam: North-Holland Publishing Company.
- Schultz, P. T. (2004, June). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74(2), 199–250.
- Stark, O., Helmenstein, C., & Prskawetz, S. (1998). Human capital depletion, human capital formation, and migration: A blessing or a ‘curse’? *Economic Letters*, 60(3), 363.
- Stinebrickner, T. (2001). Compensation policies and teacher decisions. *International Economic Review*, 42(3), 751–779.
- Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear IV regression*. NBER Technical Working Paper No. 284. Cambridge, MA: NBER.
- Super, C. M., Herrera, M. G., & Mora, J. O. (1990). Long-term effects of food supplementation and psychosocial intervention on the physical growth of Colombian infants at risk of malnutrition. *Child Development*, 61, 29–49.
- Tan, J. P., Lane, J., & Coustere, P. (1997). Putting inputs to work in elementary schools: What can be done in the Philippines? *Economic Development and Cultural Change*, 45(4), 857–879.
- Thomas, D. (1990, Fall). Intrahousehold resource allocation: An inferential approach. *Journal of Human Resources*, 25(4), 635–664.
- Thomas, D., Frankenberg, E., Friedman, J., Habicht, J. P., Hakimi, M., Jaswadi, , et al. (2003). *Iron deficiency and the well-being of older adults: Early results from a randomized nutrition intervention*. Los Angeles, CA: University of California.
- Todd, P. E., & Wolpin, K. I. (2003, February). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 118, F3–F33.
- Todd, P. E., & Wolpin, K. I. (2006). Using a social experiment to validate a dynamic behavioral model of child schooling and fertility: Assessing the impact of a school subsidy program in Mexico. *American Economic Review*, 96(5), 1384–1417.
- Todd, P. E., & Wolpin, K. I. (2007, Winter). The production of cognitive achievement in children: Home, school and racial test score gaps. *Journal of Human Capital*, 1, 91–136.
- Tokman, A. R. (2002). *Is private education better? Evidence from Chile*. Santiago, Chile: Central Bank of Chile. Central Bank of Chile Working Paper No. 147.
- Townsend, R. M. (1994, May). Risk and insurance in village India. *Econometrica*, 62(3), 539–592.
- United Nations ACC/SCN. (2000). 4th Report on the World Nutrition Situation. New York: United Nations, in collaboration with the International Food Policy Research Institute.
- Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics*, 88(1), 171–177.
- Urquiola, M., & Verhoogen, E. (2006). *Class size and sorting in market equilibrium: Theory and evidence*. New York, NY: Columbia University.
- Urquiola, M., & Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99, 179–215.
- Vazir, S., & Kashinath, K. (1999). Influence of the ICDS on psychosocial development of rural children in Southern India. *Journal of Indian Academy of Applied Psychology*, 25, 11–24.

- Vermeersch, C., & Kremer, M. (2004). School meals, educational attainment, and school competition: Evidence from a randomized evaluation. World Bank Policy Research Working Paper No. WPS3523. Washington, DC: World Bank.
- Victora, C. G., & Barros, F. C. (2006). Cohort profile: The 1982 Pelotas (Brazil) birth cohort study. *International Journal of Epidemiology*, 35, 237–242.
- Victora, M. D., Gomes, V. C., & Barros, F. C. (1990). Cross-cultural differences in development rates: A comparison between British and Brazilian children. *Child: Care, Health and Development*, 16, 151–164.
- Waber, D. P., Vuori-Christiansen, L., Ortiz, N., Clement, J. R., Christiansen, N. E., Mora, J. O., et al. (1981). Nutritional supplementation, maternal education, and cognitive development of infants at risk of malnutrition. *American Journal of Clinical Nutrition*, 34(Suppl. 4), 807–813.
- Watanabe, K., Flores, R., Fujiwara, J., & Tran, L. T. H. (2005). Early childhood development interventions and cognitive development of young children in rural Vietnam. *Journal of Nutrition*, 135(8), 1918–1925.
- Watkins, S. C., Zulu, E., Kohler, H. P., & Behrman, J. R. (Eds.). (2003). Social interactions and HIV/AIDS in rural Africa: Special collection of articles. *Demographic Research*, 1(1), 1–30.
- Welch, F. (1970, January/February). Education in production. *Journal of Political Economy*, 78(1), 35–59.
- White, H. (1980). A heteroscedasticity-consistent covariance matrix and a direct test for heteroscedasticity. *Econometrica*, 48(4), 817–838.
- Wolpin, K. I. (1982). A new test of the permanent income hypothesis: The impact of weather on the income and consumption of farm households in India. *International Economic Review*, 23, 583–594.
- Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economics Review*, 93 (2), 133–138.
- World Bank. (2004). World Development Report 2004: Making services work for poor people. Washington, DC: World Bank.
- World Food Program. (2002). *Global school feeding report 2002*. Rome: World Food Program School Feeding Support Unit.
- Yaqub, S. (2002). Poor children grow into poor adults: Harmful mechanisms or over-deterministic theory. *Journal of International Development*, 14(8), 1081–1093.

This page intentionally left blank

The Place of Nature in Economic Development*

Partha Dasgupta¹

University of Cambridge, Cambridge, United Kingdom
University of Manchester, Manchester, United Kingdom

Contents

1. Nature in Economics	4978
2. Natural Capital and Economic Development	4980
3. Natural Capital and Development Economics	4982
4. Types of Natural Capital	4985
5. Contents	4987
5.1 Externalities	4988
6. Unidirectional Externalities: Exports and Wealth Transfers	4989
6.1 Quantifying externalities	4990
6.2 Internalizing externalities	4990
7. Reciprocal Externalities: Common Property Resources	4992
7.1 Why CPRs?	4992
7.2 The importance of CPRs	4994
7.3 The good news about CPRs	4995
7.4 The bad news about CPRs	4996
7.5 Valuation, evaluation, and sustainable development	5001
8. Inadequacies in GDP	5001
9. Valuing Goods and Services	5002
9.1 Defining shadow prices	5003
9.2 Estimating shadow prices	5003
10. Social Well-Being and Comprehensive Wealth	5006
10.1 A unifying model	5007
10.2 The welfare significance of net domestic product	5008
11. Comprehensive Investment and Sustainable Development	5009
11.1 The basic theory	5009
11.2 Global public goods	5011
11.3 An application	5012
12. Evaluating Policy Reforms	5016
12.1 Social cost-benefit analysis	5016
12.2 Evaluating projects in practice	5017
13. Discounting Climate Change	5021
13.1 Global activism now or later?	5021
13.2 Negative discount rates	5022

13.3 The welfare economics of climate change	5022
13.4 Social aversions to inequality and risk	5024
13.5 Climate treaties	5025
14. Conclusions	5026
Appendix	5029
A.1 Economic perturbation as movement through time: Sustainable development	5029
A.2 Policy Evaluation	5033
End Notes	5034
References	5036

Abstract

In this chapter both theory and empirics are used to show that our picture of the processes of economic development changes radically when nature is introduced as a capital asset. Particular features of institutions that fashion societies' use of the natural-resource base are identified and analyzed. It is also shown that conventional measures of human welfare are inadequate for identifying sustainable development.

A comprehensive measure of an economy's wealth is developed and shown to be the correct index to use both for assessing the sustainability of economic development and for evaluating policies.

JEL classifications: D6, H1, I3, O1, O4, Q2, Q5

Keywords

sustainable development
 wealth
 externalities
 common property resources
 global public goods
 climate change
 social discount rates
 environmental risks
 shadow prices
 natural capital

1. NATURE IN ECONOMICS

Are humanity's dealings with nature sustainable? Should one expect the global economic growth that has been experienced over the past five decades to continue in the foreseeable future? Should we be confident that knowledge and human skills will increase in such ways as to lessen our reliance on nature in relation to humanity's growing numbers and rising economic activity?

Contemporary discussions on these questions are now several decades old. If they have remained alive and continue to be shrill, it is because two opposing empirical perspectives shape them. On the one hand, if we look at specific examples of natural

capital (aquifers, ocean fisheries, tropical forests, the atmosphere as a carbon sink—or *ecosystems*, generally), there is convincing evidence that at the rates at which we currently exploit them they are very likely to change character dramatically for the worse, with little advance notice. Indeed many ecosystems have already collapsed, with short notice.² Carpenter, Pingali, Bennet, and Zurek (2005) and Hassan, Scholes, and Ash (2005), which contain the first two sets of technical reports accompanying the Millennium Ecosystem Assessment (MEA, 2003), found that some 15 out of the 24 major ecosystem services that were examined for the Assessment are either already degraded or currently subject to unsustainable use. On the other hand, if we study historical trends in the prices of marketed resources, or improvements in life expectancy, or growth in recorded incomes in regions that are currently rich and in those that are on the way to becoming rich, resource scarcities would not appear to have bitten. Suppose you were to point to the troubled nations of sub-Saharan Africa and suggest that resource scarcities are acute there today. Those with the former perspective (ecologists generally) will tell you that it is because people in the world's poorest regions face acute resource scarcities relative to their numbers that they are so poor, while those with the latter perspective (economists usually) will inform you that people there experience serious resource scarcities because they are poor. When experts disagree over such a fundamental matter as the direction of causation, there is little to go on.

Those conflicting intuitions are also not unrelated to an intellectual tension between the concerns people share about carbon emissions and acid rains that sweep across regions, nations, and continents, and about declines in firewood and water sources that are specific to the needs and concerns of the poor in small, village communities. That is why “environmental problems” present themselves in different ways to different people. Some identify environmental problems with population growth, while others identify them with wrong sorts of economic growth. There are those who see environmental problems as urban pollution in emerging economies, while others view them through the spectacle of poverty. Some interpret “sustainable development” as sustainable development of the global economy, while others see it in terms of development prospects of a village in sub-Saharan Africa. Each of those visions is correct. There is no single environmental problem; there is a large collection of them (Ehrlich & Ehrlich, 1981, 1990; Dasgupta, 1993, 2001a; Sachs, 2008).

Growth in industrial and agricultural pollutants has accompanied economic development; and in industrialized countries neither preventive nor curative measures have kept pace with their production. That neglect is now prominent in the rapidly growing regions in China, India, Mexico, and Brazil. Moreover, the scale of the human enterprise, both by virtue of unprecedented increases in the size of the world's population and the level of economic activity, has so stretched the capabilities of ecosystems, that humankind is today Earth's dominant species. During the twentieth century world

population grew by a factor of four (to more than 6 billion) and world output by 14, industrial output increased by a multiple of 40 and the use of energy by 16, methane-producing cattle population grew in pace with human population, fish catch increased by a multiple of 35, and carbon and sulfur dioxide emissions by more than 10. The application of nitrogen to the terrestrial environment from the use of fertilizers, fossil fuels, and leguminous crops is now at least as great as that from all natural sources combined. In a notable work Vitousek, Ehrlich, Ehrlich, and Matson (1986) estimated that 40% of the 45–60 billion metric tons of carbon that are harnessed annually by terrestrial photosynthesis (net primary production of the biosphere) is currently being appropriated for human use. To be sure, this is a rough estimate; still, the figure puts the scale of the human presence on the planet in perspective (see also Carpenter et al., 2005; Daily & Ellison, 2002; Ehrlich & Ehrlich, 2008; Hassan et al., 2005; MEA, 2003; Vitousek, Mooney, Lubchenco, & Melillo, 1997).

On the other hand, economic growth itself has brought with it improvements in the quality of a number of environmental resources. The large scale availability of potable water and the increased protection of human populations against both water- and air-borne diseases in industrial countries have come with the economic growth those countries have enjoyed over the past 200 years. Moreover, the physical environment inside the home has improved beyond measure. Cooking in South Asia continues to be a central cause of respiratory illnesses among women. Growth in scientific knowledge, investment in public infrastructure, and universal education in advanced industrial countries have meant that citizens there have far greater knowledge of environmental hazards than their counterparts in poor regions and have the resources to avoid them. Such positive links between economic growth and environmental quality often go unacknowledged by environmentalists in the West.

2. NATURAL CAPITAL AND ECONOMIC DEVELOPMENT

Despite the conflicting intuitions, most economists would appear to be convinced that technological advances and the accumulation of reproducible capital and growth in human capital can overcome diminutions in natural capital. Otherwise it is hard to explain why twentieth-century economics has been so detached from the environmental sciences. Judging by the profession's writings, we economists see nature, when we see it at all, as a backdrop from which resources and services can be drawn in isolation. Macroeconomic forecasts routinely exclude natural capital, and accounting for nature, if it comes into the calculus at all, is an afterthought to the real business of "doing economics." When asked, economists acknowledge nature's existence, but many deny she is worth much. I have heard professional colleagues remark at seminars that the services nature provides amount at best to 2–3% of an economy's output, which is the share of agriculture in the gross domestic product (GDP) of the United States. Why, they ask,

should one incorporate a capital asset of negligible importance in macroeconomic models of growth and distribution?

Typically it is assumed in growth models that nature is a fixed, indestructible factor of production (Ricardo's "land").³ The problem with the assumption is that it is wrong: nature consists of degradable resources. Agricultural land, forests, watersheds, fisheries, fresh water sources, estuaries, the atmosphere—more generally, ecosystems—are resource stocks that are *regenerative*, but suffer from depletion or deterioration when they are overused. (I am excluding oil and natural gas, which are at the limiting end of self-regenerative resources.) Moreover, the environmental sciences tell us that the elasticity of substitution between reproducible capital and human capital, on the one hand, and vital forms of natural capital, on the other, is less than one (Ehrlich & Goulder, 2007). It may have been understandable of economists to assume that nature does not need to be counted at a time when natural resources were abundant relative to the demand that was made of them, but it is not understandable in models of development possibilities open to the world today. The stance taken in modern growth models is questionable also because property rights to natural capital are often either vaguely defined or weakly enforced (Sections 5 and 6), implying that nature's services are underpriced in the market. Official statistics on national income certainly give the impression that natural capital is of small importance; but official statistics are built on market prices, not shadow prices. Studies of local ecosystems suggest that if shadow prices were to be used in economic statistics, the decomposition of national income into its various components would look quite different. For example, Repetto, Magrath, Wells, Beer, & Rossini (1989) and Vincent, Ali, et al. (1997) estimated the decline in forest cover in Indonesia and Malaysia, respectively, and found that when depreciation is included, national accounts there look quite different: net saving rates are some 20–30% lower than recorded rates. In their work on the depreciation of natural resources in Costa Rica, Solorzano et al. (1991) found that the depreciation of three resources (forests, soil, and fisheries) amounted to about 10% of GDP and over a third of capital accumulation.

Distortions in the pricing of primary factors of production filter down to influence research and development. The latter in turn influences the character of technological change. Because nature's services are underpriced in the market, innovators have little reason to economize on their use. We should not be surprised when new technologies are rapacious in the use of natural capital.

Modern growth theories ignore every layer of those resource allocation failures. Recent concerns over global climate change and the growing scarcity of fresh water in the world's poorest regions are perhaps the first acknowledgement among mainstream development economists and international development agencies (e.g., Stern, 2006; UNDP, 2006, 2007/2008) that at the scales at which the world economy has been operating for some time, nature is in many aspects fragile.

In any case we should be sceptical of a theory of economic progress that places such enormous burden on an experience not much more than 250 years old. Extrapolation into the past is a sobering exercise: over the long haul of history (a 5000 years stretch, say, up to about 250 years ago), economic growth even in the currently rich regions was for most of the time not much above zero. The study of possible feedback loops between poverty, population growth, and the character and performance of both human institutions and natural capital is not common currency in modern growth models. That is probably why environmental- and resource economics, or *ecological economics* for short, remains isolated from the main body of contemporary economic thinking.

Evidence of that isolation has been provided strikingly, even if unintentionally, by Kim, Morse, and Zingales (2006), who identified what has mattered to the economics profession since 1970, by surveying 41 of the most prominent refereed economics journals. In his Plenary Lecture in 2007 at the Annual Conference of the European Association of Environmental and Resource Economists, Ehrlich (2008: p. 16) observed that the following words appeared neither in Kim et al. nor in the titles of the top 146 articles (500 cites or more) in those 41 journals: “abatement, aquifer, . . . , biotic, . . . , carrying capacity, climate, . . . , ecosystem, . . . , fertility, forest, . . . , pollution, population, poverty, . . . , soil, . . . , toxic, . . . , warming, and water.”

3. NATURAL CAPITAL AND DEVELOPMENT ECONOMICS

Official development economics reflects the rest of our discipline, in that it too neglects nature’s place in economic development. The neglect looks odd to ecologists, who are trained to study the slow processes that influence long-term development possibilities. A seemingly natural retort to ecologists is that people come first and that, after all, current poverty should matter most. There are two problems with the position. First, the future has a habit of becoming the present. Secondly, extreme poverty is frequently associated with a degraded environment.

It could be that economists neglect nature because the services it provides are judged to be luxury goods, as in the view expressed in prominent newspapers that “economic growth is good for the environment because countries need to put poverty behind them in order to care,” (*The Independent*, 4 December 1999), or that “. . . trade improves the environment, because it raises incomes, and the richer people are, the more willing they are to devote resources to cleaning up their living space,” (*The Economist*, 4 December, 1999). But in the poor world natural capital is not only an amenity, it is also a primary factor of production. Often it is a basic need (see World Resources Institute, 2005 for a global summary). In an early publication, Falconer (1990) recorded the major significance of “minor” forest products among the poor in the humid forest zone of West Africa. She also recorded the many different ways people there have

coped with a decline in their access to those products. Perrings' work (2000) is an exceptional study of the role that biodiversity plays in the lives of the poor in agro-pastoral sub-Saharan Africa. The author also records the decline in biodiversity in recent decades and looks for its causes. In a carefully designed study in the rainforest of Chiapas, Mexico, Lopez-Feldman and Wilen (2008) found that nontimber products there are extracted mainly by the poor. They attribute that in part to the fact that extraction involves search and that the opportunity cost of time (relative to the value of nontimber products to the extractor) is low among the poor.

Hassan et al. (2005) is the most comprehensive study to date on the state of the local natural-resource base in poor regions. The studies confirm that the world's poorest people (some 620 million in number) live in especially fragile natural environments. When wetlands, inland and coastal fisheries, woodlands, forests, estuaries, village ponds, aquifers, and grazing fields are damaged (owing to agricultural encroachment, nitrogen overload, urban extensions, the construction of large dams, resource usurpation by the state, open access, or whatever), it is the rural poor who suffer most. Frequently, there are no alternative sources of livelihood. In contrast, for rich eco-tourists or importers of primary products, there is something else, often somewhere else; which means that there are alternatives.

There may be a second reason why economists have neglected natural capital. Public concerns over environmental problems are often prompted by large-scale disasters, such as nuclear leakage, storms, and floods. The environmental impacts of large undertakings (big dams and irrigation systems) also catch the public eye. This should not surprise. Big impacts of "single" incidents are directly visible. So they provoke public outcry and elicit a response. In contrast, the slow, shifting processes that characterize human-nature interactions are not easy to detect by outsiders; at least, not unless and until a threshold is reached and a catastrophe occurs. Of course, in the poor world small-scale disasters occur all the time, it is only that they are not visible to outsiders. But large numbers of small-scale disasters can over time add up to greater human losses than a small number of large-scale catastrophes. Empirical evidence collected over the past two decades confirms an earlier intuition that rural poverty and environmental degradation in poor countries is in large measure caused by those institutional failures whose deleterious effects accumulate slowly over time (Dasgupta, 1982; Hassan et al., 2005; Repetto, 1988).

Up to now, even growth in the atmospheric concentration of carbon dioxide has affected economic outcomes very slowly. But as the evidence suggests, the process characterizing global climate change is remorseless and in all probability harbors any number of tipping points for the future (Lenton et al., 2008). Although international development agencies (e.g., UNDP, 2007/2008) are now acknowledging the costs that people in the tropics are likely to face owing to climate change, their concerns have been largely about the efficacy of an international "cap-and-trade" system and the

assistance rich nations ought to give poor countries in order to meet climate change. They are legitimate concerns, but it seems to me in order to identify good policies one needs to examine the pathways by which people may get trapped in poverty when development prescriptions ignore natural capital. That requires analysis, not rhetoric.

The systematic neglect of ecology in development economics should be puzzling. 65–75% of people in the world's poorest regions live in rural areas. Mention agricultural land, threshing grounds, grazing fields, village tanks and ponds, woodlands and forests, rivers and streams, coastal fisheries, mangroves, or coral reefs, and the importance of the *local* natural-resource base to the rural poor becomes self-evident. In a pioneering study, the (Indian) Centre for Science and Environment (CSE, 1990) recorded that, of the total number of hours worked by villagers in a microwatershed in the central Himalayas in India, 30% was devoted to cultivation, 20% to fodder collection, and about 25% was spread evenly between fuel collection, animal care, and grazing. Some 20% of time was spent on household chores, of which cooking took up the greatest portion, and the remaining 5% was involved in other activities, such as marketing.

Subsequent studies (e.g., Filmer & Pritchett, 2002) have also recorded the importance of the local natural-resource base in rural life. Nevertheless, apart from agricultural land, *ecological* capital has been absent from the formal models mainstream development economists have used to discuss policy. It is absent too from influential surveys and texts on the economics of development (Banerjee & Duflo, 2005; Dreze & Sen, 1990, 1995; Ray, 1998; Sen, 1999; Stern, 1989).

That neglect has had far-reaching implications for development policy. Tallis, Karava, Marvier, and Chang (2008) report that of the more than 11,000 development projects that have been supported by the World Bank since 1947, only about 15% have included natural capital as a theme. They also report that the number of World Bank projects approved in the period 1993–2007 that had biodiversity as a theme and poverty alleviation and the protection of ecological capital as stated goals was a mere 32, and that only five of those recorded improvements in environmental quality and a reduction in poverty. As all economies suffer from serious distortions in the use of ecological capital (see below), it should be possible to identify policy reforms that both help to reduce poverty and improve ecological services (Repetto, Dower, Jenkins, & Geoghegan, 1992). With that in mind, Tallis et al. have compiled a list of indicators that should prove useful in designing, selecting, and implementing projects that offer the prospect of reaching what the World Bank calls “win-win outcomes.”

Despite the neglect, there is now a growing literature on the links between rural poverty and the local ecology. Because economic theory predated economic empirics in that body of work, empirical studies have frequently been designed by scholars to respond to the theory. In this chapter I draw on that literature to show how economics can be reworked to include nature's services in the study of development processes.⁴ Unfortunately, even now there are few reliable empirical studies. One reason for the

dearth may be that, with the exception of forest cover, government surveys (even the best of them, such as the Indian National Sample Survey) do not include detailed information on ecological capital. In addition to studying household behavior, investigators therefore have to obtain their own data on the state of the local ecology.

Even though the field is nascent, the literature I report below stands somewhat in contrast to the environmental and resource economics that was developed over the decades in the United States. In part influenced by the needs of the US Environmental Protection Agency (EPA), resource economists there mostly studied the economics of timber, oil and natural gas, water resources, and fisheries; while environmental economists focused on environmental amenities and pollution. Moreover, it was customary in the earlier literature to interpret environmental problems as symptoms of market failure.⁵

I shall put some distance between the material I report in this chapter and the themes usually covered in environmental and resource economics. I shall do that partly because, in identifying the place of natural capital in the lives of the rural poor, I want to shift the focus onto ecological services; but partly also because in poor countries, people transact in a wide variety of *nonmarket* institutions. It will be seen below that the inclusion of natural capital in economic reasoning alters not only our assessment of the contemporary development experience, but our understanding of development processes as well. So as to emphasize that the context I am studying here is rural poverty, we will regard ecological capital to be the quintessential example of natural capital. When necessary, though, I shall allude to natural capital, interpreted in an inclusive way.⁶

4. TYPES OF NATURAL CAPITAL

Natural capital is of direct use in consumption (fisheries); of indirect use, as inputs in production (timber); or of use in both (fresh water). The value of a resource is often derived from its usefulness (as a source of food, or as an essential actor in ecosystems—e.g., a keystone species); but there are resources whose value is aesthetic (places of scenic beauty), or intrinsic (primates, blue whales, sacred groves), or a combination of all three (biodiversity). The worth of a natural resource could be based on what is extracted from it (forest products, fisheries), or on its presence as a stock offering service (coral reefs, wetlands, forest cover), or on both (watersheds).

Ecosystems provide innumerable services to us. Among the visible products are food, fibers, fuel, and fresh water, but many remain hidden from view. Among other things, ecosystems maintain a genetic library, preserve and regenerate soil, fix nitrogen and carbon, recycle nutrients, control floods, mitigate droughts, filter pollutants, assimilate waste, pollinate crops, operate the hydrological cycle, and maintain the gaseous composition of the atmosphere. As those services are not visible, it is easy to overlook

them. However, with some ingenuity it is possible to estimate the benefits we enjoy from them. Below I report some of those findings.

Ecosystems offer joint products: wetlands recycle nutrients and produce purified water; mangrove forests protect coastal land from storms and are spawning grounds for fish; and so on. Unhappily, social tensions arise in those many cases where an ecosystem has competing uses (farms versus forests versus urban developments; forests versus agro-ecosystems; coastal fisheries versus aquaculture⁷). As natural capital is a mesh of resources, what one means by an ecosystem is usually influenced by the scope of the problem being studied. A number of ecosystems have a near global reach ("biomes," such as the Savannah), some cover entire regions (river basins), many involve clusters of villages (microwatersheds), while others are confined to the level of a single village (the village pond). Sachs, Gallup, and Mellinger (1998) have traced the location of world poverty in part to the fact that the tropics harbor some of the most fragile environments. Hassan et al. (2005) and Carpenter et al. (2005) contain a classification of ecosystems. They also provide an account of the stresses that are being experienced by both global and local ecosystems.

Environmental pollutants are the reverse of natural resources. In some cases the emission of pollutants amounts directly to the depreciation of reproducible capital (corrosion of material infrastructure). In others it means a degradation of ecosystems (eutrophication of lakes). Roughly speaking "resources" are "goods," many being sinks into which pollutants are discharged (rivers, the atmosphere, and the oceans are among the sinks); while "pollutants" (the degrader of resources) are "bads." Pollution is the other side of conservation.⁸

Ecosystems are driven by interlocking nonlinear processes that run at different speeds and operate at various spatial scales (Steffen et al., 2004). That is why ecosystems harbor multiple basins of attraction. The global climate system is now a well-known example (Bigg, 2003). But small-scale ecosystems also harbor multiple basins of attraction, for similar reasons. So long as phosphorus run-off into a fresh water lake is less than the rate at which the nutrient settles at the bottom, the water column remains clear. But if over a period of time the run-off exceeds that rate, the lake collapses into a eutrophic state.⁹ Usually, of course, the point at which the lake will collapse is unknown. That means the system is driven by nonlinear stochastic processes. So, flips in the capacity of ecosystems to supply useful service to us share three important characteristics: (1) they are frequently irreversible (or at best they take a long time to recover); (2) except in a very limited sense, it is not possible to replace degraded ecosystems by new ones; and (3) ecosystems can collapse abruptly, without much prior warning. Imagine what would happen to a city's inhabitants if the infrastructure connecting it to the outside world was to break down without notice. Vanishing water sources, deteriorating grazing fields, barren slopes, wasting mangroves, and bleached coral reefs are spatially confined instances of a corresponding breakdown among the

rural poor. Ecological collapse, such as the one that has been experienced in recent years in Rwanda, the Horn of Africa, and the Darfur region of Sudan, can also trigger rapid socioeconomic decline (Collier, 2007; Diamond, 2005; Homer-Dixon, 1999).

5. CONTENTS

The plan of this chapter is as follows:

In Sections 5 and 6, I introduce two broad categories of externalities and discuss the kinds of institutions that are likely to be most effective in removing them. The remainder of the chapter discusses ways to value natural capital and develops a method for evaluating economic programs. Two types of evaluation exercises are contrasted: evaluating projects and identifying sustainable development. In Section 7, I show that GDP should not be used as a welfare index in either exercise. Section 8 contains a discussion of a number of methods for valuing ecological capital. In Section 9, I show that a comprehensive measure of wealth is the correct index to use in both types of evaluation exercise. Section 10 demonstrates the reach of the index for identifying sustainable development. Section 11 describes the same for project evaluation.

In many people's minds today, "environmental problems" mean "climate change." One of my aims in this chapter is to dispel that notion. Nevertheless, in the long-run climate change would appear to be one of the most serious environment problems facing poor countries. So in Section 12 I offer an extended discussion of those evaluation methodologies that have fashioned the empirical work on the economics of climate change. I point to weaknesses in the analyses, rather than harp on their strengths. Section 13 summarizes; and Appendix has a formal demonstration of the theoretical propositions put forward in Sections 9–11.

Throughout, I try to mingle theoretical analysis with empirical studies. I do that not only for microstudies (e.g., the value of planting farm forests and investing in fresh water sources; measuring the value of upland forests in stabilizing water flows downstream), but also for macroeconomic identification of sustainable development. I do the latter by putting the theory to work on interpreting the development experience of the world's poorest regions in the period 1970–2000. With regard to the former, it is as well to note that empirical studies of household and village behavior governing local ecosystems differ widely in style. They range from narratives on the lives of people in a single village to econometric studies of data from many villages. Even among the latter, some are based on carefully constructed controls, while others are more casual in their approach to statistics. I believe that I have learnt from each of that heterogeneous body of styles. In any event, empirical studies of human–nature interactions are still so few in number, that it would be foolish to insist on the style we economists have got used to in applied microeconomics. The catholicity of styles will not suit everyone, but in drawing attention to it here I am merely placing my cards on the proverbial table.

5.1 Externalities

An early literature found market failure to be the underlying cause of environmental problems (Baumol & Oates, 1975; Meade, 1973; Pigou, 1920), which is why the phenomenon of market externalities looms large in environmental and resource economics. But there are nonmarket institutions too, and they also can fail. By an *externality* we mean the effects that decisions have on people who have not been party to the decisions. I offer this definition because both theory and contemporary evidence tell us that environmental externalities are a symptom of a more general problem: *institutional failure*. The malfunctioning institution could be the market. But the failure could be that of a group of nations unable to agree on common fisheries policy in the seas; or it could be the state riding roughshod over forest inhabitants; it could be the local community whose norms have collapsed (Section 6); or it could be the household, where the dominant male insists on growing fruit trees (because the fruit can be sold in the market to which the female does not have easy access) rather than trees that would supply the wood-fuel the female is expected to gather from the receding woodlands. The consequences of those malfunctions are resource allocation failures among contemporaries and across the generations. I shall also argue that one of the consequences has been the very high population growth the poor world has experienced in recent decades.

Activities involving ecological capital give rise to externalities because property rights to them are either weakly defined or inadequately enforced. And a common reason for the latter is that key components of natural capital are mobile, as is the case with air and water. By property rights I do not only mean “private” property rights, I include “community” property rights and state property rights. At an extreme end are “global” property rights on global public goods, a concept that is implicit in current discussions on climate change. But neither the idea of global property rights nor of global public goods is new. That humanity has collective responsibility over the state of the world’s oceans used to be explicit in the 1970s, when politicians claimed that the oceans are a “common heritage of mankind.”

In the presence of externalities involving nature’s services, individuals and communities overexploit natural capital, which is another way of saying that ecological services are subsidized. At the global level what is the annual subsidy? One calculation suggested that it is 10% of annual global income (Myers & Kent, 2000). My reading is that the margin of error in that estimate is very large. But it is the only global estimate I have come across. The technical reports accompanying MEA (2003), especially Carpenter et al. (2005) and Hassan et al. (2005), contain quantitative information that could be used to obtain more reliable estimates of nature’s subsidies; which in turn could be used to estimate the shadow prices of various ecological capital assets (Section 8). International organizations such as the World Bank have the resources to undertake that work. But they appear to be reluctant to do so.

Two broad types of externalities may be contrasted: *unidirectional* and *reciprocal* (Dasgupta, 1982). Unidirectional externalities are just that—unidirectional—where one agent (or a set of agents) inflicts or confers an externality on another (or others). The direction of the externality is in part determined by social norms and legal rules (Coase, 1960), a matter to which I return below. Under reciprocal externalities each party inflicts or confers an externality on all others, as in the case of unmanaged common property resources (CPRs). In the following two sections, we study the significance of the two types of externalities.

6. UNIDIRECTIONAL EXTERNALITIES: EXPORTS AND WEALTH TRANSFERS

Exports of primary products often involve unidirectional externalities. There is evidence that, other things being equal, freeing exports from politically motivated restrictions helps economies to grow faster. There is evidence too that the poor as a group enjoy the benefits of faster growth (McCulloch, Winters, & Cirera, 2001). But as the ecological consequences of growth in exports are rarely measured, the case for trade expansion should be qualified beyond the occasional footnote. Here is an example of how an increase in the export of primary products can hurt the poor.¹⁰

An easy way for domestic governments to earn revenue in countries that are rich in forests is to issue timber concessions to private firms. Imagine that concessions are awarded in the upland forests of a watershed. Forests stabilize both soil and water flow. So deforestation gives rise to soil erosion and increases fluctuations in water supply downstream. If the law recognizes the rights of those who suffer damage from deforestation, the timber firm would be required to compensate downstream farmers. But compensation is unlikely when (1) the cause of damage is many miles away, (2) the concession has been awarded by the state,¹¹ and (3) the victims are scattered groups of farmers. Problems are compounded because damages are not uniform across farms: location matters. It can also be that those who are harmed by deforestation do not know the underlying cause of their deteriorating circumstances. As the timber firm is not required to compensate farmers, its operating cost is less than the social cost of deforestation, the latter, as a first approximation, being the firm's logging costs and the damage suffered by all who are adversely affected. So, the export contains an implicit subsidy, paid for by people downstream. And I have not included forest inhabitants, who now live under even more straightened circumstances; or worse, are evicted without compensation. The subsidy is hidden from public scrutiny, but it amounts to a transfer of wealth from the exporting to the importing country. Some of the poorest people in a poor country subsidize the incomes of the average importer in what could well be a rich country. That does not feel right.

6.1 Quantifying externalities

The spatial character of unidirectional externalities is self-evident, but getting a quantitative feel involves hard work. So the literature is sparse. As in other fields of ecology and economics, some of the best advances in ecological economics have been made in studies of “small” problems. There are now several believable estimates of subsidies on the use of natural capital at the local level. Basing their estimate on a formal hydrological model, Pattanayak and Kramer (2001) reported that the drought mitigation benefits farmers enjoy from upstream forests in a group of Indonesian watersheds are 1–10% of average agricultural incomes. In another exemplary work, Pattanayak and Butry (2005) studied the extent to which upstream forests stabilize soil and water flow in Flores, Indonesia (see also Pattanayak, 2004). Downstream benefits were found to be 2–3% of average agricultural incomes. In a study in Costa Rica on pollination services, Ricketts, Daily, Ehrlich, and Michener (2004) discovered that forest-based pollinators increase the annual yield in nearby coffee plantations by as much as 20%. Ricketts et al. (2008) have analyzed the results of some two dozen studies, involving 16 crops in five continents, and discovered that the density of pollinators and the rate at which a site is visited by them declines at rapid exponential rates with the site’s distance from the pollinators’ habitat. At 0.6 km (resp. 1.5 km) from the pollinators’ habitat, for example, the visitation rate (resp. pollinator density) drops to 50% of its maximum.¹²

6.2 Internalizing externalities

How should societies eliminate unidirectional externalities? In the case of the upstream firm and downstream farmers, the state could tax the firm for felling trees (Pigou, 1920). The firm in this case would be the “polluter,” the farmers the “pollutees.” Pigovian taxes therefore invoke the *polluter-pays-principle (PPP)*. The efficient rate of taxation would be the damage suffered by farmers. What the state does with the tax revenue is a distributional matter, to which I shall return presently. Pollution taxes are known today as “green taxes.”

But there is also a “market-friendly” way to eliminate externalities. Lindahl (1958 (1919))—and subsequently Meade (1952, 1973) and Coase (1960)—suggested that the state (or the community; see Section 6) could introduce private property rights on ecological capital, the thought being that markets would emerge to eliminate the externalities. A problem with the proposal, at least as I have presented it here, is that it is not clear who should be awarded property rights. In our example of the upstream firm and downstream farmers, the sense of natural justice might suggest that the rights should be assigned to farmers, who can be regarded as the pollutees. Under a system of “pollutees rights,” the timber firm would be required to compensate farmers for the damage it inflicts on them. Such a property-rights regime also invokes PPP.

Of course, the rights could be awarded to the timber firm instead. In that case it would be the farmers who would have to compensate the firm for not felling trees.

The latter system of property rights invokes the *pollutee-pays-principle* (a reverse PPP, as it were), which to many people would seem repellent. But from the efficiency point of view it is a matter of indifference which system of private property rights is introduced, so long, that is, as the prices that emerge (including those in the markets for externalities) are competitive prices (Starrett, 1972).

In a famous article Arrow (1971) pointed to a problem with Lindahl's proposal. Markets for externalities would be "thin." In our example each market would involve precisely two parties: the timber firm and one farmer. It is hard to imagine that competitive prices could emerge in such circumstances. Nevertheless, markets for externalities have attracted much attention among ecologists and development experts in recent years, under the label *payment for ecosystem services*, or PES (see Pagiola et al., 2002 for a sympathetic review of a market-based PES).

The ethics underlying PES is seemingly attractive. If decision makers in Brazil believe that decimating the Amazon forests is the true path to economic progress there, should not the rest of the world pay Brazil not to raze them to the ground? If the lake on my farm is a sanctuary for migratory birds, should not bird lovers pay me not to drain it for conversion into farm land? Never mind that the market for ecosystem services would be thin, if a system involving PES were put in place, owners of ecological capital and beneficiaries of ecological services would be forced to negotiate. The former group would then have an incentive to conserve their assets (Daily & Ellison, 2002; Goldstein et al., 2006; Pagiola, Landell-Mills, & Bishop, 2002). In a review of current practices, Jack, Kousky, and Sims (2008) note that hundreds of new PES schemes have been established round the globe. China, Costa Rica, and Mexico, for example, have initiated large-scale programs in which landowners receive payment for increasing biodiversity conservation, expanding carbon sequestration, and improving hydrological services.

Although we have no firm empirical understanding of processes in which a single agent is engaged in multiple bilateral negotiations over the supply of what amounts to a "public good," we do have some theoretical understanding. Consider a situation in which the farmers in our example are willing to negotiate with the timber firm. It can be shown that if the parties discount future profits at a very low rate, all but one farmer would free ride and enjoy the benefit (i.e., they would make no payment to the firm). So farmers have no incentives to form a coalition among themselves. Moreover, the (equilibrium) outcome is not unique, meaning that the farmer who negotiates with the firm is not necessarily the one who would be worst affected by deforestation.¹³

Although PES may be good for conservation, one can imagine situations where the system would be bad for poverty reduction and distributive justice. Many of the rural poor in poor countries enjoy nature's services from assets they do not own. Even though they may be willing to participate in a system of property rights in which *they*

are required to pay for ecological services (Pagiola, Rios, & Arcenas, 2008, report in their careful study of a silvo-pastoral project in Nicaragua that they do), it could be that in the world we have come to know, the weaker among the farmers is made to pay a disproportionate amount. Some may even become worse off than they were earlier (Dasgupta & Heal, 1979: chapter 3). One could argue that in those situations the state should pay the resource owner instead, using funds obtained from general taxation. As Reid, Mooney, Cropper, et al. (2005), Bulte, Lipper, Stringer, and Zilberman (2008), and Zilberman, Lipper, and McCarthy (2008) observe, who should pay depends on the context.

A PES system in which the state plays an active role is attractive for wildlife conservation and habitat preservation. In poor countries property rights to grasslands, tropical forests, coastal wetlands, mangroves, and coral reefs are often ambiguous. The state may lay claim to the assets ("public" property being the customary euphemism), but if the terrain is difficult to monitor, inhabitants will continue to reside there and live off its products. Inhabitants are therefore key stakeholders. Without their engagement the ecosystems could not be protected. Meanwhile flocks of tourists visit the sites on a regular basis. An obvious thing for the state to do is to tax tourists and use the revenue to pay local inhabitants for protecting their site from poaching and free-riding. Local inhabitants would then have an incentive to develop rules and regulations to protect the site.

But even if inhabitants as a collective are given an incentive to self-regulate, individual members may not have the incentives to conform to communitarian rules and regulations. The economics of "common property resources" studies ways in which individual incentives can be aligned to communitarian goals. We discuss that next.

7. RECIPROCAL EXTERNALITIES: COMMON PROPERTY RESOURCES

Who owns the local natural-resource base in rural areas? Anthropologists, economists, and political scientists working at the fringes of official development economics have discovered that, barring agricultural land, ecological capital is often neither private nor state property. Nor are they "open-access" resources. They are communal property, which is why they are called *common property resources*, or CPRs. If CPRs are badly managed, households in the community free-ride, and the commons suffer from excessive use. The point of communal management is to restrict use (e.g., by establishing charges or quantity restrictions) and fend off a possible "tragedy of the commons."

7.1 Why CPRs?

As a proportion of total assets, CPRs range widely across ecological zones. In India they are most prominent in arid regions, mountain regions, and unirrigated areas; they are least prominent in humid regions and river valleys (Agarwal & Narain, 1989;

Chopra, Kadekodi, & Murty, 1989). There is a rationale behind this, based on the need to pool risks. Woodlands, for example, are spatially inhomogeneous ecosystems. In some years one group of plants bears fruit in one part of a woodland, in other years some other group in some other part is fecund. Relative to mean output, fluctuations could be presumed to be larger in arid regions, mountain regions, and unirrigated areas. If a woodland were to be divided into private parcels, each household would face greater risks than it would under communal ownership and self-regulation. The reduction in individual household risks may be small, but as average incomes are very low in Indian villages, household benefits from communal ownership could be expected to be large.

Where users are symmetrically placed, distributions would be expected to be symmetric, a subtle matter to devise if the resource is heterogeneous. Rotation of access to the best site is an example of how this can be achieved. It is often practiced in coastal fisheries, fuel reserves in forest land, and fodder sites in the grasslands (Baland & Platteau, 1996). Rotation enables users to get a fair shake.

Of course, it would be possible in principle for the community to parcel out the resource as private property and let households establish a mutual insurance scheme. But that move would jeopardize cooperation in other activities; for at least two reasons. First, cooperation appears to be habit forming (Seabright, 1993); so, dispensing with cooperation in any one activity could lead to a weakening of cooperation in other activities. Secondly, cooperation is more robust when sanctions for opportunism in any one venture include exclusions not only from that venture, but also from other collective ventures. Abandoning cooperation in one field of activity thus reduces the robustness of cooperation in other fields of activity. This fact is an implication of the theory of repeated games. It explains why relationships are so frequently tied to one another in rural communities (Dasgupta, 2007a).

Local ecosystems are frequently CPRs also because their constituents are mobile. Birds and insects fly, fish swim, inorganic materials diffuse in space, and even earthworms are known to travel. Their mobility integrates an ecosystem's various components. Ecosystem dynamics are nonlinear, involving positive feedback in a wide range of states, meaning that the system as a whole is greater than the sum of its spatial parts. Ecosystems therefore have an element of indivisibility to them. If you slice off a significant portion for some other purpose, the productivity (e.g., biomass production) per unit area of what remains is reduced. But even if it were decreed that no portion could be converted for another use, parcelling ecosystems into private bits would be inefficient because of the externalities that are created by the mobile components. Admittedly, private monopoly would avoid the externalities, but it would grant far too much power to one person in the community.

Agricultural land, especially in densely populated areas, is a different matter. Both labor and capital are critical inputs in production. Investment can increase land's productivity enormously. Agricultural land as CPRs would be subject to serious

management problems, including those due to the temptations to free ride on investment costs. The lack of incentives to invest and innovate would lead to stagnation, even decay. The fate of collective farms in what was previously the Soviet Union testifies to that. Those regions of sub-Saharan Africa where land is, or was until recently, held by the kinship were exceptions, but only because land was plentiful in the past and because poor soil quality meant that land had to be kept fallow for extended periods. Of course, it may be that agricultural productivity remained low there *because* land was held by the kinship, not by individuals. As elsewhere in the social sciences, causation typically works in both directions.

7.2 The importance of CPRs

Are CPRs important to rural people? Jodha (1986) reported evidence from a number of dry rural districts in India that the proportion of income among poor families that is based directly on CPRs is 15–25%. Cavendish (2000) has arrived at even larger estimates than Jodha from a study of villages in Zimbabwe: the proportion of income based directly on CPRs is 35%, the figure for the poorest quintile being 40%. At a country level, some 5 billion US dollars of products are drawn from CPRs annually by the Indian rural poor (Beck & Nesmith, 2001). Wood-fuel is prominent among CPR products. It is estimated that some 2.4 billion of the world's poor depend on wood or other biomass fuels for cooking and heating (World Resources Institute, 2005). That is why the finding by Cooke, Kohlin, and Hyde (2008) that community forestry in poor countries continues to ignore the importance of wood-fuel among rural households is ironic.

Marine fisheries are a major source of food and income among the coastal poor. Some 250 million of the world's poor depend on coastal CPRs. Such evidence does not of course prove that CPRs are well managed, but it does show that rural households have strong incentives to devise arrangements whereby they *are* managed.

CPRs not only supply households with a regular flow of ecosystem services and tangible goods (water, fuel-wood, fibers, building material, fruit, honey, and fish), they also offer protection against agricultural risks. In a study of households on the margin of the Tapajos National Forest in the Brazilian Amazon, Pattanayak and Sills (2001) found that households make more trips into the forest for nontimber products when times are hard. CPRs are sometimes the only assets to which the otherwise disenfranchised have access. Hecht, Anderson, and May (1988) have described the importance of *babassu* products among the landless in Maranhão, Brazil. Extraction from those plants offers support to the poorest of the poor, most especially women. The authors reported that *babassu* products are an important source of cash income in the period between agricultural-crop harvests. Economic theory says that where there is a market for labor even the casual wage rate of unskilled laborers would be higher in villages with more abundant CPRs (Dasgupta, 1993). There is evidence of this (Barbier, 2005). That said, I am

not implying that asset-less people featured prominently in community decisions to create the institutions that govern CPRs, I am merely drawing attention to a good feature of CPRs.

7.3 The good news about CPRs

Most often CPRs are not open to outsiders, but only to those having historical rights through kinship ties and community membership. Communal management of local resources makes connection with *social capital*, viewed as a complex of interpersonal networks, and hints at the basis upon which cooperation has traditionally been built. As CPRs have been seats of nonmarket relationships, transactions involving them are often not mediated by market prices. So their fate is frequently unreported in national economic accounts. However, a large empirical literature has confirmed that resource users in many cases cooperate, on occasion through democratic means. The empirical literature on CPRs is now huge (Baland, Bardhan, & Bowles, 2007; Baland & Platteau, 1996, 1999; Bromley, 1992; Chopra, Kadekodi, & Murty, 1989; Feeny, Berkes, McCay, & Acheson, 1990; Ghate, Jodha, & Mukhopadhyay, 2008; Jodha, 1986, 2001; Kadekodi, 2004; National Research Council, 1986, 2002; Netting, 1985; Noronha, 1997; Ostrom, 1990, 1992; Wade, 1988; among many others). The economic *theory* of CPRs was however constructed earlier, in Dasgupta and Heal (1979: chapter 3).

Are CPRs managed communally? Not invariably, but in many cases they are, or have been in the past. Wade (1988) reported his findings from a study of community-based allocation rules over water and the use of grazing land in 41 South Indian villages. He noted that downstream villages had an elaborate set of rules for regulating the use of water from irrigation canals. Fines were imposed on those who violated the rules. Most villages had similar arrangements for the use of grazing land.

In a study on forest conservation in the central Himalayas, Somanathan, Prabhakar, and Mehta (2005) have found that the density of broad-leaf trees was significantly higher in places where the forest was managed by village councils than in areas where neither councils nor the state was involved in forest management.

Of course, cooperation does not appear in vacuum. In the contemporary world there is a potential role of government and nongovernment organizations (NGOs) in helping to build or rebuild local institutions through which communities could get to realize the advantages of collective action. Such help would involve, among other things, devising clearly defined rules concerning the allocation of burdens and benefits, rules whose compliance can be observed (hopefully, verified also) by others involved. In a study in North-West India, Chopra and Gulati (1998) found that distress migration out of villages where NGOs had been at work to create institutions for managing water and pasture land on a communal basis was lower than in villages where there had been little attempt to create such institutions. Significantly, Chopra and

Gulati found that the probability of participation in communal pasture land was higher among villagers who were participating in communal water management schemes than among villagers who were not. That suggests once again that cooperation begets cooperation.

How is cooperation maintained? Universally, studies have found that collective sanctions are imposed on those who misbehave (see, e.g., Baland & Platteau, 1996). Today the underlying mechanism is common knowledge among economists. The theory of repeated games has shown that so long as households do not discount future costs and benefits at too high a rate and so long as behavior is mutually observable, social norms involving the use of sanctions on noncooperative behavior enable cooperation to be maintained. Sanctions range from the punitive and unforgiving (permanent exclusion following a single misdemeanor; widely known in the economic literature as the “grim” norm) to the forgiving (as in graduated sanctions; see below). The grim norm has been found in reciprocal relationships (Czako & Sik, 1988), but appears to be in force only in those environments where the parties have access to formal markets as an alternative. Something like “grim” is needed for preventing people from engaging in noncooperative behavior in an environment where tempting, short-term opportunities appear elsewhere from time to time.

Outside opportunities are often rare in rural communities that are to all intents and purposes enclaves. Communitarian arrangements there are of high value to all and matters are different. Graduated sanctions are in wide use. The first misdemeanor is met by a small punishment, subsequent ones by a stiffer punishment, persistent ones by a punishment which is stiffer still, and so forth (see, e.g., Ostrom, 1992). Where information is imperfect, a small penalty for the first misdemeanor could be a warning that others are watching, or it could be that others signal their acknowledgement that the misdemeanor could have been an error on the part of the offender and that he should try harder next time. And so on.

7.4 The bad news about CPRs

So far, the good report on CPRs. There are, however, two pieces of bad news. The first involves resource allocation in communitarian institutions. Entitlements from CPRs are frequently based on private holdings: richer households enjoy a greater proportion of the benefits. Beteille (1983), for example, drew on examples from India to show that access to CPRs is often restricted to the elite (e.g., caste Hindus). Cavendish (2000) has reported that in absolute terms richer households in his sample of villages took more from CPRs than poor households. In an early review, McKean (1992) noted that benefits from CPRs are frequently captured by the elite. Agarwal and Narain (1996) exposed the same phenomenon in their study of water management practices in the Gangetic plain, as did Bardhan and Dayton-Johnson (2007) in a study of irrigation systems in Mexico and South India.

However, the relative exploitation of CPRs by the poor and the not-so-poor is not uniform across the world. In two large-scale studies of household data in India and Nepal, respectively, Bandopadhyay and Shyamsundar (2004) and Bandopadhyay, Shyamsundar, and Kanel (2006) found that wood-fuel consumption decreases with wealth in India, but increases with wealth in Nepal. Their finding suggests that the availability of cheap substitutes matters. In India, where rural markets are more developed than in Nepal, relatively wealthy households are able to save on labor costs by buying fuel in the market place.

That women are sometimes excluded from CPRs has been recorded in communal forestry (Agarwal, 2001). It is even possible that the elite *exploit* others, in the strong sense that the latter are worse off when the CPR is regulated than they would have been if the CPR was unregulated (Dasgupta, 2000; see also Dasgupta, 2008a). But because cooperation in one activity is usually tied to cooperation in other activities, it would be hard to establish empirically that one group of CPR users is exploiting another group of users.

The second piece of bad news is that CPRs have deteriorated in recent years in many parts of the poor world (Hassan et al., 2005; Jodha, 2001; Perrings, 2000). Why should that have happened in those places where they had been managed in a sustainable manner previously? There are several reasons:

One stems from deteriorating external circumstances, under which both the private and communal profitability of investment in the resource base decline. Political instability is a general cause. It is, of course, a visible cause of resource degradation, as civil disturbance all too frequently expresses itself by a destruction of physical capital. But increased uncertainty in communal property rights is a frequent, often hidden cause. People could worry that the state or warlords will assume authority over the CPRs. If the security of a CPR is uncertain, the returns expected from collective action are low. The influence would run the other way too, with growing resource scarcity contributing to political instability, as rival-groups battle over resources. The feedback could be “positive,” exacerbating the problem for a time, thus reducing expected returns yet further.

The second reason CPRs have deteriorated in many places is rapid population growth. The latter triggers environmental degradation if institutional practices are unable to adapt to the increased pressure on resources. In Cote d’Ivoire, for example, growth in rural population has been accompanied by increased deforestation and reduced fallows. Biomass production has declined, as has agricultural productivity (Lopez, 1998). Leisinger, Schmidt, and Pandya-Loch (2002) offer a wide-ranging discussion of the deleterious effects of recent population growth on food security in poor regions.

However, rapid population growth in the world’s poorest regions in recent decades itself requires explanation. Demographers have argued that a fall in the child mortality

rate (which is otherwise a very good thing) is a major factor (Bhargava, 2008 identifies reasons why even fertility rates respond to declines in child mortality with a lag). Here, I suggest factors that have not been much studied by demographers. Increased economic insecurity, owing to deteriorating institutions, is one: children yield a higher private return in such circumstances than other forms of capital assets (Bledsoe, 1994; Guyer, 1994; Heyser, 1996). Reproductive activity also involves a number of externalities that encourage people in the world's poorest regions to be pronatalists (Dasgupta, 2003). To take an example, consider that when institutions governing the CPRs deteriorate, households free-ride on the resource base. As some of the cost of maintaining a household is passed on to others, the net private benefits of accumulating more "hands" to mine the CPRs can increase. Dasgupta and Mäler (1991) and Nerlove (1991) argued that receding firewood and water sources may increase the household demand for labor, leading to a rise in household size. In an analysis of data from South Africa, Aggarwal, Netanyahu, and Romano (2001) have found a positive link between fertility increase and degradation of CPRs. Filmer and Pritchett (2002) have reported a weak positive link between the two in the Sindh region in Pakistan.

On the other hand, Loughran and Pritchett (1998) found evidence in Nepal that more acute resource scarcity was associated with lower fertility, which suggests that growing scarcity there raised the net cost of having children. Apparently, increasing firewood and water scarcity in the villages in the sample did not have a strong enough effect on the relative productivity of child labor to induce higher fertility. It seems then that the relationship between resource scarcity and fertility can be of either sign.¹⁴ Admittedly, none of the empirical studies just mentioned quite captures what the theory I am alluding to here tells us to study, namely, the link between *desired* household labor and the state of the local natural-resource base; but they come close enough.

The third reason CPRs have deteriorated in some places is that communal rights were overturned by central fiat. To establish its political authority, a number of states in the Sahel, for example, imposed rules that destroyed communal management practices in the forests. Villages ceased to have the authority to enforce sanctions on those who broke communitarian rules. But state officials did not have the expertise to manage the commons, often they were corrupt. Knowledge of the local ecology is held by those who work on the commons. Local participatory democracy offers a mechanism by which that knowledge can inform public policy. Isham, Narayan, and Pritchett (1995) found strong evidence from 121 rural water projects (in 49 countries in Africa, Asia, and Latin America) that participation by beneficiaries is positively correlated with project performance. Relatedly, Thomson, Feeny, and Oakerson (1986), Somanathan (1991), and Baland and Platteau (1996), among others, have identified the many ways by which the exercise of state authority damages local institutions and turn CPRs into open-access resources. In their study of forest management in the central Himalayas, Somanathan et al. (2005) found that crown cover was no less in places that were

governed by village councils than it was in areas managed by the state, but expenditure on governance was an order of magnitude higher in the latter places.

Democratic movements among stakeholders and pressure from international organizations have encouraged a return to communitarian management systems. Shyamsundar's work (2008) is a remarkable synthesis of the findings in nearly 200 articles on the efficacy of devolution of management responsibilities—from the state to local communities—over the local natural-resource base. Her article focuses on wildlife, forestry, and irrigation. The balance of evidence appears to be that devolution leads to better resource management, other things being equal. Shyamsundar of course offers a discussion of what those other things are.

The fourth reason CPRs have deteriorated in many places is that cooperation is fragile, dependent as it is on many factors that have to work simultaneously in its favor. For example, in the face of growing opportunities for private investment in substitute resources, households are more likely to break agreements that involve reciprocity (Campbell et al., 2001; Dasgupta, 1993, 2007a). But when traditional systems of management collapse and are not replaced by adequate institutions, CPRs suffer from neglect. Here are three examples illustrating the phenomenon:

1. Mukhopadhyay's work (2008) is a historical study of the transformation of agrarian land in Goa, India that was earlier owned and regulated by a communitarian institution called the *communidades*. When Goa became a part of India, the government introduced land reforms that gave tenants the right to purchase the land they had worked. Mukhopadhyay does not question the underlying motivation behind land reforms, but notes one unfortunate consequence, which is the breakdown of cooperation among households in maintaining the embankments that had earlier prevented the land from flooding by tidal waters. Over the years deterioration of the embankments has led to an increase in soil salinity.
2. In her study of collectively managed irrigation systems in Nepal, Ostrom (1996) accounted for differences in rights and responsibilities among users (who gets how much water and when, who is responsible for which maintenance task of the canal system, and so forth) in terms of the fact that some farmers are headenders, while others are tailenders. Headenders have a built-in advantage, in that they can prevent tailenders from receiving water. On the other hand, headenders need the tailenders' labor for repair and maintenance of traditional canal systems, which are composed of headworks made of stone, trees, and mud. Ostrom reported that a number of communities in her sample had been given well-meaning aid by donors which installed permanent headworks. What could be better, you may ask. But Ostrom observed that those canal systems that had been improved were frequently in worse repair at the tailend and were delivering less water to tailenders than previously. Ostrom also reported that water allocation was more equitable in traditional farm

management systems than in modern systems managed by external agencies, such as government and foreign donors. She estimated from her sample that agricultural productivity is higher in traditional systems.

Ostrom's explanation for this is that unless it is accompanied by countermeasures, the construction of permanent headworks alters the relative bargaining positions of the head- and tailenders. Headenders do not now need the labor of tailenders to maintain the canal system. So the new sharing scheme involves even less water for tailenders. Headenders gain from having the permanent structures installed, but tailenders lose. This is an example of how well-meaning aid can go wrong if the nature of the institution receiving the aid is not understood by the donor.

- 3.** Village tanks are one of the oldest sources of irrigation in South Asia. In a study of a group of villages in southern India, Balasubramanian (2008) reports that village tanks have deteriorated over the years owing to a decline in collective investment in their maintenance. That decline has taken place as richer households have invested increasingly in private wells. As poor households depend not only on tank water but also on the fuel-wood and fodder that grow round the tanks, construction of private wells has accentuated economic stress among the poor.

And the fifth reason an erosion of CPRs can come in the wake of shifting populations accompanying the development process itself (Dasgupta, 2000). As economic opportunities outside the village improve, those with lesser ties (e.g., young men) are more likely to take advantage of them and make a break with customary obligations. Those with greater attachments (e.g., women) would perceive this and thereby discount at a higher rate the benefits that could be expected from complying with agreements. Either way, norms of reciprocity could be expected to break down, making certain groups of people (women, children, the aged) worse off.

History tells us that CPRs can be expected to decline in importance as economies develop (North & Thomas, 1973). Ensminger's (1992) study of the privatization of common grazing lands among the Orma in north-eastern Kenya established that the transformation took place there with the consent of the elders of the tribe. She attributed the move to cheaper transportation and widening markets, making private ownership of land more profitable. However, as the elders were from the stronger families, privatization accentuated inequality within the tribe. The point is not to lament the decline of the commons; rather, it is to identify those who are likely to get hurt by changes in economic regimes and the accompanying transformations in the use to which the resources are put. That there are winners in the process of economic development is a truism. Much the harder task is to identify the likely losers and have policies in place that act as safety nets for them. In what follows we study policy evaluation and institutional reform when ecological capital is included in the exercise.

7.5 Valuation, evaluation, and sustainable development

In evaluating an economy, there are five questions we can ask: (A) How is the economy doing? (B) How has it performed in recent years? (C) How is it likely to perform under “business as usual”? (D) How is it likely to perform under alternative policies? (E) What policies should be pursued there?

National income accounts offer information relevant for answering question (A), although I argue below that they do so in an unsatisfactory way. Policy evaluation, including *project evaluation*, is a way to answer questions (D) and (E). The idea is to evaluate an economy at a point in time before and after a hypothetical perturbation has been made to it. In contrast, the literature on “sustainable development” answers questions (B) and (C) by evaluating economic change when the perturbation is the passage of time itself.

Question (A) stands apart from questions (B) to (E), at least if conventional practice among national income statisticians is any guide. For it is common practice to summarize the state of an economy by its GDP, or equivalently its (gross) domestic income.

8. INADEQUACIES IN GDP

A good history of the concepts of GDP and national income was provided by Richard Stone in his Nobel Lecture, which is available on the home page of the Nobel Foundation. But even a cursory study tells us that GDP rose to prominence during the 1930s, when industrial nations were suffering from economic depression. At that time there was a need to find an index of aggregate economic activity. GDP filled that need. During the post-War years, although GDP came to be regarded as a welfare index. That interpretation is now so ingrained in us that it has become common practice to use estimates of GDP growth for answering questions (B) to (E). Indeed, if someone talks of economic growth, the listener does not need to ask, “Growth in what?” he will know that the speaker means growth in GDP.

The use of GDP as a welfare index has been much criticized on grounds that it is insensitive to distributional concerns. But the criticism would be met if distributional weights were applied to different income groups (Dasgupta, Marglin, & Sen, 1972). The real weakness of GDP as a welfare index lies elsewhere. Economic growth is no doubt a good thing—it usually purchases a better quality of life—but as we shall see presently, studying movements in GDP is of no use in ascertaining whether economic growth is compatible with “sustainable development.”

A famous report by the World Commission on Environment and Development (1987)—known widely as the Brundtland Report—defined *sustainable development* as “...development that meets the needs of the present without compromising the ability of future generations to meet their own needs.” In their version “sustainable

development” requires that relative to their populations, each generation should bequeath to its successor at least as large a *productive base* as it had itself inherited. Notice that the requirement is derived from a relatively weak notion of intergenerational equity. Sustainable development demands that future generations have no less of the means to meet their needs than we do ourselves. It demands nothing more; it does not, for example, demand that development be *optimal* or *just*. But how is a generation to judge whether it is leaving behind an adequate productive base for its successor?¹⁵

It is clear that tracking GDP will not do. GDP is a linear index of the final goods and services. Even though it has been argued that the weights attached to goods and services ought to be shadow prices (Mirrlees, 1969; Sen, 1976), national income statisticians use market prices, shorn of taxes and subsidies. As large numbers of ecological services in poor countries are transacted in nonmarket institutions, official GDP mis-specifies the level of economic activity. But there is a more fundamental problem with GDP: the index mismanages intertemporal considerations badly. (We confirm below that the United Nations’ Human Development Index suffers from that same weakness.) The point is that GDP ignores the depreciation of capital assets. Among natural resources, that depreciation can range from a full 100% of the services drawn from oil and natural gas, to the degradation of ecosystems through mismanagement. As natural capital is especially vulnerable to overuse, serious criticisms of GDP were first made in the context of environmental and natural-resource problems.¹⁶

We may put the matter another way: GDP measures the aggregate *income* of the *current* generation, whereas in seeking to determine the sustainability of economic development, we should be interested in the *well-being* of current and future generations. So, by *social well-being* we will mean an ethically defendable numerical index of the well-beings of the present and future generations. As the future is uncertain, the numerical index is taken to include an ethically defendable attitude toward that uncertainty. In Section 12 (where I review the welfare theory underlying the economics of climate change) and Appendix (where formal accounts of policy evaluation and sustainable development are illustrated), social well-being is understood to be a generalized form of utilitarianism. In what follows I drop further reference to uncertainty (but see Section 12.4 on the economics of climate change) and assume simply that the index of social well-being reflects it.

9. VALUING GOODS AND SERVICES

Evaluating an economy requires that we value goods and services from a societal point of view. The social values of goods and services are called *shadow prices*, a familiar term in development economics. To keep the analysis neutral across commodities, let social well-being be the *numeraire*.

9.1 Defining shadow prices

In public economics shadow prices are defined as the difference between market prices and optimum taxes or subsidies (Atkinson & Stiglitz, 1980). This is too narrow a base on which to build policy analysis in rural economies, where transactions are frequently based on communitarian relationships. Moreover, the term “optimal” can rarely be applied to policies chosen there. So we revert to the mathematical definition.

Definition 1. *Suppose at date t an economy is awarded an additional unit of some asset free of charge. The asset's shadow price is the resulting change in social well-being.*

Appendix provides the mathematical counterpart of Definition 1 (Eq. A.6). But the definition already tells us that three pieces of information are required for estimating shadow prices:

- (i) A descriptive model of the economy.
- (ii) The size and distribution of the economy's capital assets.
- (iii) A conception of social well-being.

Requirements (i) and (ii) are the basis for estimating the changes that take place in the allocation of resources if an additional unit of the asset is made available free of charge. Requirement (iii) is the basis for placing a value on that change. Dasgupta (1982) used Definition 1 to argue that even in imperfect economies the dynamics of ecosystems have to be taken into account explicitly if we are to estimate the shadow prices of natural resources and environmental pollutants.

At any date an asset's shadow price is a function of the stocks of all assets. Moreover, the price today depends not only on the economy today, but also on the entire future of the economy. Future scarcities of ecological capital are reflected in current shadow prices of all goods and services. So shadow prices are functions of the degree to which various assets are substitutable for one another, not only at the date in question, but also at all subsequent dates as well. Of course, if the conception of social well-being involves the use of high discount rates on the well-being of future generations, the influence on today's shadow prices of future scarcities would be attenuated. Intergenerational ethics plays an important role in the structure of shadow prices, a fact that was displayed in the contrasting recommendations of Cline (1992) and Stern (2006) on the one hand and Nordhaus (1994, 2008) on the other, over how much the world community should spend now to meet the problems of global warming (Section 12.3). In imperfect economies (e.g., those experiencing the tragedy of the commons) an asset's shadow price can be negative even when its market price is positive.

9.2 Estimating shadow prices

How should shadow prices be estimated? Consider the human capital asset we call “health.” Suppose we wish to value a marginal increase in life expectancy. Economists have followed two ways to do this. One is to estimate the (social) cost of bringing about

that increase. The other is to estimate the value of the increase in life expectancy itself. The two would lead to the same estimate at a full optimum, but not in imperfect economies, where the latter is the right way to go about the matter (Definition 1 and Eq. A.6). But as the latter requires us to estimate the value of a statistical life (Viscusi & Aldy, 2003), the method has proved to be controversial when deployed for making cross country comparisons. Arrow et al. (2004) used figures for public health expenditure for the shadow price of health improvements in poor countries, while Arrow, Dasgupta, Goulder, Mumford, and Oleson (2008) have combined age-specific mortality tables with values of statistical life years so as to estimate the shadow price of increases in life expectancy. The latter estimates have been found to be substantially higher than the former, implying that the economies in their sample of countries are highly imperfect.

Using Definition 1 directly is problematic because of the enormous quantity of information demanded by requirements (i)-(iii). So environmental and resource economists have devised two indirect methods (Freeman, 1993; Smith, 1997 are fine expositions of the methods). In one, investigators ask people to place a value on ecological resources. In the other, they study behavior and the consequences of that behavior to infer the value individuals place on those assets. In the latter method, market prices of those goods and services for which there are markets are often taken to be their shadow prices. As an example of the latter, consider an asset that has multiple characteristics (e.g., land). The *hedonic method* uses the market price of a piece of land to uncover the shadow price of one of its characteristic (e.g., the price of its aesthetic qualities). The hedonic method has been much used to value real estate. In their work on inland wetlands in eastern North Carolina (USA), Bin and Polasky (2005) found that, other things being equal, proximity to wetlands *lowered* property values, presumably because of a greater infestation of insects and possibly bad odor.

The valuation methods that have become most popular were devised for environmental amenities, such as places of scenic beauty or cultural significance. The *cost of travel* to a site takes revealed preference to be the basis for valuing the site. Englin and Mendelsohn (1991), for example, is a well-known application of the method for estimating the recreation value of forests. In contrast, in those cases where there is no observed behavior, the *contingent valuation method* (CVM) has proved to be extremely popular (see Carson, 2004 for an extensive bibliography and Smith, 2004 for an excellent history of the method). The idea is to ask people how much they would be willing to pay for the preservation of an environmental amenity (e.g., flood control) or a resource of intrinsic worth (e.g., an animal or bird species).

Each of the above methods is of limited use for valuing the local natural-resource base in the poor world. Moreover, one can question whether requirements (i)-(iii) can be met adequately by studying people's behavior or analyzing their responses even to well-designed questions. One reason for being circumspect about those methods (there are many other reasons) is that people are not often aware of environmental risks. Jalan

and Somanathan (2008) conducted an experiment among residents of a suburb of New Delhi. The aim was to determine the value of information on the health risks that arise from drinking water that contained bacteria of faecal origin. Without purification the piped water in 60% of the households were found to be contaminated. Among households in the sample that had not been purifying their piped water, some were informed by the investigators that their water was possibly contaminated, while the rest were not informed. The authors report that the former group of households was 11% more likely to invest in purification within the following 8 weeks than the latter group. An additional year of schooling of the most educated male in the household was associated with a 3% increase in the probability that its piped water was being treated. The finding is noteworthy because the wealth and education levels of households in the sample were above the national average. If ignorance of environmental risks is pervasive, estimates of the demand for environmental quality that assume full information must be misleading.¹⁷

So we return to requirements (i)-(iii). In their work on the economics of climate change, Cline (1992) and Stern (2006) met (i) and (iii) directly (Section 12.3). Several recent valuation studies have met requirement (i) by estimating the *production function* for nature's service (e.g., pollination as a function of the distance to a forest; primary productivity as a function of biodiversity; net reproduction rate of a species), but have otherwise assumed that market data are more or less sufficient to meet the other requirements.¹⁸ Pattanayak and Kramer (2001) and Pattanayak and Butri (2005), for example, constructed a hydrological model to measure the contribution of upland forests to farm productivity downstream. Hassan (2002) used quantitative models of woody land resources in South Africa to estimate the value to rural inhabitants of (among other resources) the Fynbos Biome, which dominates sandy soils there. Barbier (1994), Barbier and Strand (1998) and Gren et al. (1994) used formal ecological models to compile a catalogue of the various services that are provided by wetlands. In their study of wetlands in northern Nigeria, Acharya (2000) and Acharya and Barbier (2000) applied models of ground water recharge to show that the contribution wetlands make to recharging the basins is some 6% of farm incomes. That is a large figure.

The welfare economics of climate change requires that carbon in the atmosphere is priced. The early literature on the subject did not have a spatial component to that price. A figure of 20 US dollars per ton for carbon's global shadow price was suggested by Fankhauser (1995) and Pearce, Cline, et al. (1996). That figure has been used in the World Bank's work on sustainable development (Section 10.2). But there are likely to be enormous regional variations in the impact of global climate change on economic activity (Dinar, Hassan, Mendelsohn, Benhin, et al., 2008; Mendelsohn, Dinar, & Williams, 2006; Rosenzweig and Hillel, 1998). Agriculture in semiarid tropical countries is expected to suffer from warming, while in temperate regions it will probably benefit. If we apply distributional weights to the losses and gains, the disparity is bigger than the nominal figures that have been suggested, because the former group

of countries is almost all poor while the latter are middle income to rich. Using a range of climate models, Mendelsohn et al. have published estimates of losses and gains in year 2100. The authors aggregated five sectors: agriculture, water, energy, timber, and coasts. Depending on the scenario, they found that the poorest countries (almost all in Africa) are likely to suffer damages from 12% to 23% of their GDP, while the range of impacts on the richest countries (North America and northern Europe) is from damages of 0.1% to a *gain* of 0.9% of their GDP. Dinar et al. fear that with warming, the agricultural income in the semiarid tropics could be more than halved in 2100 from its projected value in the case where there is no warming. But these estimates are based on market prices. If distributional weights are applied to obtain a global shadow price of carbon, it would be a lot higher than if we were merely to add the regional gains and losses. It should also be noted that the effects of climate change on health and the environment (e.g., loss of species) were not included in those estimates.

Determining the shadow prices of ecological capital assets should now be central to the research agenda in development economics. Carpenter et al. (2005) and Hassan et al. (2005) contain the kind of information that would prove useful in such exercises. We would have been far ahead in the development of policies that help to alleviate poverty in poor countries had development economists and policy makers taken natural capital seriously in the past.

What is the point of basing shadow prices solely on one particular use-value when we know that natural capital often possesses other values too? The answer is that the method provides us with *biased* estimates of shadow prices. That can be useful information. For example, in a beautiful paper on the optimal rate of harvest of blue whales, Spence (1974) took the shadow price of whales to be the market value of their flesh, a seemingly absurd and repugnant move. But on estimating the population growth functions of blue whales and the harvest-cost functions, he found that under a wide range of plausible parameter values it would be most profitable commercially for the international whaling industry to agree to a moratorium until the desired long-run population size was reached, and for the industry to subsequently harvest the whales at a rate equal to the population's optimal sustainable yield.¹⁹ In Spence's analysis, preservation was recommended solely on commercial ground. But if preservation is justified when the shadow price of blue whales is estimated from their market price, the recommendation would, obviously, be reinforced if their intrinsic worth were to be added. This was the point of Spence's exercise.

10. SOCIAL WELL-BEING AND COMPREHENSIVE WEALTH

Social well-being is very difficult to estimate directly because it is a nonlinear function of the flow of goods and services. We seek a linear index that moves in unison with social well-being. Shadow prices are essential for the task.

10.1 A unifying model

Imagine that we have estimated shadow prices on the basis of the information covering requirements (i)–(iii). Let us now use those prices as weights to construct an aggregate index of the economy’s comprehensive stock of capital assets. The assets on the list include not only reproducible capital (roads, buildings, machines, instruments) and human capital (health, human talents, education), but also population numbers and natural capital. Moreover, the list contains “knowledge,” including scientific and technological knowledge, and institutional capabilities.

It is typically assumed in economic models, that changes in some of the economic variables are exogenous. Growth or decline in population, and movements in total factor productivity (TFP) and the price of imports are typical examples. Below I show that an obvious way to accommodate exogenous changes in economic variables is to regard *time* also as a capital asset.

Call the aggregate index of the economy’s comprehensive stock of capital assets we have thus constructed, its *comprehensive wealth*. Formally

Definition 2. *An economy’s comprehensive wealth is the (shadow) value of all its capital assets.*

We will find it useful to separate time from all other assets. So let all capital assets excepting for time be indexed by i . Let $K_i(t)$ be the stock of asset i at time $t (\geq 0)$ and $K(t)$ the corresponding vector of assets. We assume that none of the K_s s is a global public good (but see Section 10.2). Let $V(t)$ be social well-being at t . We know that $V(t) = V(K(t), t)$. In Appendix, we study the standard representation of $V(t)$. Let $p_i(t)$ be asset i s (spot) shadow price (i.e., $p_i(t) = \partial V(t)/\partial K_i(t)$; see Appendix) and $r(t)$ the (spot) shadow price of time (i.e., $r(t) = \partial V(t)/\partial t$; see Appendix). Writing $W(t)$ for the economy’s comprehensive wealth at t , we have

$$W(t) = r(t)t + \sum_i p_i(t)K_i(t). \quad (1)$$

Why should we be interested in comprehensive wealth? The reason lies in

Proposition 1. *A small perturbation to an economy increases (resp., decreases) social well-being if, and only if, holding shadow prices constant, it increases (resp., decreases) comprehensive wealth.²⁰*

Proof. Let Δ denote a small perturbation. Assuming V is differentiable, we have

$$\Delta V(t) = [\partial V/\partial t]\Delta t + \sum_i [\partial V/\partial K_i(t)]\Delta K_i(t). \quad (2)$$

As $p_i(t) = \partial V(t)/\partial K_i(t)$, Eq. (2) can be written as

$$\Delta V(t) = r(t)\Delta t + \sum_i p_i(t)\Delta K_i(t). \quad (3)$$

Equation (1) and Proposition 1 suggest that comprehensive wealth (or *wealth*, for short) is a measure of the economy's *productive base*. Moreover, Proposition 1 says that the reason we should be interested in that particular measure of the productive base is that it moves in unison with social well-being. Equation (3) is even stronger than Proposition 1. The equation says that if an economy is perturbed slightly, the change in social well-being accompanying the perturbation *equals* the change in wealth (at constant shadow prices) that is caused by it.

Now $p_i(t)\Delta K_i(t)$ is the shadow value of net investment in asset i , and $\partial V / \partial t$ is the shadow price of time t ($r(t)$ in expression (1)). Write $I_i(t) = p_i(t)\Delta K_i(t)$. Then Eq. (3) can be expressed as

$$\Delta V(t) = r(t)\Delta t + \Sigma[I_i(t)]. \quad (4)$$

Definition 2 says that the expression on the right-hand side of Eq. (4) is the *comprehensive investment* that accompanies the perturbation. This means Proposition 1 can be restated as

Proposition 2. *A small perturbation to an economy increases (resp., decreases) social well-being at t if, and only if, the comprehensive investment at t that accompanies the perturbation is positive (resp. negative).²¹*

Proposition 1 explains why (comprehensive) wealth is the correct measure of social well-being and why it ought to replace GDP, Human Development Index (HDI), and the many other *ad hoc* measures that are currently taken to reflect social well-being. As wealth is a linear index of the stocks of the economy's (comprehensive) list of capital assets, while social well-being is a *nonlinear* function of its determinants (as in various forms of utilitarianism, see Appendix), it is a far more convenient index to use for responding to questions (B) to (E) than social well-being itself.

We could imagine that the typical perturbation considered in Propositions 1 and 2 involves positive investments in some assets (e.g., roads and building, education, and health), but negative investments in other assets (e.g., wetlands and forests). Proposition 2 says that so long as comprehensive investment is positive, social well-being increases. Note though that if certain ecological assets were to become very scarce, their shadow prices would be large, signaling that further declines in their amounts, even when small, would make a significant dent on comprehensive investment.

10.2 The welfare significance of net domestic product

Propositions 1–3 also explain why *net domestic product* (NDP) is not a measure of social well-being. To see why, consider a closed economy. Suppose there are M consumption goods and services (they are all flows), labeled j . Let $C_j(t)$ be the consumption of j at t and $q_j(t)$ its shadow price (Appendix: Eq. A.4). Write NDP at t as $NDP(t)$. Using Eq. (2), we have

$$\text{NDP}(t) = \sum_j q_j(t) C_j(t) + r(t) \Delta t + \sum_i [p_i(t) \Delta K_i(t)],$$

which can be reexpressed as

$$\text{NDP}(t) - \sum_j q_j(t) C_j(t) = r(t) \Delta t + \sum_i [p_i(t) \Delta K_i(t)]. \quad (5)$$

Proposition 2 and Eq. (5) together yield

Proposition 3. *A small perturbation to an economy increases (resp., decreases) social well-being at t if, and only if, aggregate consumption is less than (resp. greater than) net domestic product.*

Proposition 3 uncovers the welfare content of NDP. In a classic work, Lindahl (1933) used what amounts to the obverse of Proposition 3 to define “income” as the maximum consumption that an economy can enjoy without reducing its wealth. But in recent years economists have wanted to claim a lot more for NDP. They have wanted to prove that NDP is proportional to social well-being.²² But it can be shown that $\text{NDP}(t)$ moves in unison with $V(t)$ if, and only if, the conception of social well-being that is being adopted is wholly insensitive to inequality among people, generations, and uncertain contingencies (Dasgupta, 2009; Dasgupta & Mäler, 1991).

Earlier we distinguished a perturbation at a point in time from a perturbation that amounts to the passage of time itself. That distinction tells us that Propositions 1 and 2 can be interpreted in two ways. We study them in the following two sections.

11. COMPREHENSIVE INVESTMENT AND SUSTAINABLE DEVELOPMENT

Consider first the passage of time along an economic forecast. The forecast could be, say, “business as usual.”

11.1 The basic theory

Proposition 1 says that the Brundtland Commission’s notion of “sustainable development” is equivalent to “sustainable social well-being.” For completeness let us define sustainable development in terms of the latter notion.

Definition 3. *An economy enjoys sustainable development at t if $dV(t)/dt \geq 0$.*

Propositions 1 and 2 give us an ethical underpinning to the concept of “sustainable development” in the sense of the Brundtland Commission. They also provide an operational handle on the concept. On using Proposition 2 and Definition 3, we have

Proposition 4. *An economy enjoys sustainable development at t if and only if comprehensive investment at t is nonnegative.*

Notice that Propositions 1–4 are equivalence results. On their own they cannot tell whether an economy is enjoying sustainable development at a point in time. It could be that owing to bad policy choices in the past, the economy is experiencing

unsustainable development even though sustainable development would have been possible had better policies been chosen. Worse still, if substitution possibilities between, say, certain vital ecological assets and other assets are limited and ecological capital has been drawn down ominously owing to past profligacy, it could be that sustainable development simply is not feasible.²³

Propositions 1–4 also say why GDP, NDP, HDI, and other *ad hoc* measures of social well-being will not do. It is certainly possible for an economy’s productive base (i.e., wealth) to grow while GDP, say, increases, which is no doubt a path of economic development we all would like to follow. But it is also possible for an economy’s productive base to *shrink* during a period when GDP (or even NDP) grows.²⁴ The problem is that no one would notice the shrinking if everyone’s eyes were riveted to GDP. If the economy’s productive base continues to shrink, economic growth sooner or later stops and reverses sign. GDP will then decline, as will the standard of living, but no one would suspect that a fall was in store. So, growth in GDP per head can encourage us to think that all is well, when it is not. In that regard, the Human Development Index, or HDI, is not better: it is possible for a country’s HDI to increase even while its productive base shrinks. This means that HDI too can mislead (for illustrations based on international data, see Section 10.3).

Propositions 1, 2, and 4 also imply that the huge empirical literature on the factors influencing economic growth, admirably surveyed by Acemoglu, Johnson, and Robinson (2005), Barro (1997), and Helpman (2004), misdirects. The equations that define cross-country growth regressions have GDP growth on the left-hand side. Proposition 1, on the other hand, says that GDP growth should be replaced by growth in comprehensive wealth. We should want to know, for example, whether, other things being equal, political and civil liberties, economic and legal institutions, or trade liberalization or any one of the many other features of economies people takes interest in promote the accumulation of comprehensive wealth. But for one work, currently in progress (Aidt, 2009), no one would appear to have studied such questions empirically. That there is even today no published study on the subject is no doubt because there are no reliable cross-country estimates of comprehensive wealth. But then there is no reason for any one to create such statistics if economists do not ask for them. And they will certainly not ask for them if they continue not to take natural capital seriously.

The one exception is Aidt (2009) who, in a work-in-progress on a large cross section of countries, has found that indicators of both perceived and actual corruption are negatively correlated with growth in comprehensive wealth *per capita*. The correlation would appear to be stronger and more robust than the correlation between corruption and GDP growth that has been reported by growth economists (e.g., Mauro, 1995). Aidt’s estimates suggest that corrupt countries have been running down their wealth.

That said, a higher growth rate in even comprehensive wealth is not necessarily socially more desirable. A society's objective should be to promote the *optimum* growth in wealth.

If we are to put Propositions 1, 2, and 4 to work on data, we need to find simple formulas for the various shadow prices. In earlier sections I reported on a small recent literature which offers estimates of the shadow prices of a number of natural capital assets. Here we look briefly at simple formulas for the shadow price of time.

Begin with the contribution the "residual" makes to comprehensive investment. In their empirical work on sustainable development, Arrow et al. (2004, 2008) constructed a simple formula for that contribution. $\partial V / \partial t$ was shown to be proportional to the residual. The authors estimated the contribution of the residual on the basis of the countries' macroeconomic data.

Turning to population growth, it could seem intuitive that in place of wealth the correct measure of social well-being is wealth *per capita*. In fact that intuition is generally speaking wrong. Dasgupta (2001a) proved that wealth *per capita* is the correct index only if (a) social well-being takes the form of what may be called "dynamic average utilitarianism" (see Appendix), (b) population grows (or declines) at a constant rate, and (c) each of the equations that represents an economy's accumulation process can be expressed in terms solely of *per capita* capital stocks. Formally if conditions (a)-(c) are satisfied, we have

Proposition 5. *An economy enjoys sustainable development at t if, and only if, at constant shadow prices (comprehensive) wealth per capita does not decline at t.*

11.2 Global public goods

What of public goods? Let $G(t)$ be the stock of a global public good at t , to be interpreted here as carbon concentration in the atmosphere. G is an argument in the V -function of every country. Let $K_m(t)$ be the vector of assets owned by residents of country m . If V_m is social well-being in m , we have,

$$V_m(t) = V_m(K_m(t), G(t), t).$$

As before, let $p_m(t)$ be the vector of shadow prices of all the assets owned by residents of m , and $g_m(t) = \partial V_m(t)/\partial G(t)$. We noted earlier that G may well be an economic "good" for countries in the temperate zone and an economic "bad" in the tropics. For the former, $g_m > 0$; for the latter, $g_m < 0$. Let $E_m(t)$ be the net emission rate from country m and $E(t)$, the net aggregate emission. It follows that

$$dG(t)/dt = {}_m\Sigma(E_m(t)) = E(t). \quad (6)$$

To focus on the accumulation of G , write

$$I_m(t) = {}_i\Sigma p_i(t) dK_i(t)/dt. \quad (7)$$

Using Eq. (7), Eq. (4) generalizes to

$$dV_m(t)/dt = \partial V_m/\partial t + I_m(t) + g_m(t)dG(t)/dt. \quad (8)$$

On using Eq. (6), Eq. (8) becomes

$$dV_m(t)/dt = \partial V_m/\partial t + I_m(t) + g_m(t)[{}_m\Sigma(E_m(t))]. \quad (9)$$

We see next that recent empirical work on sustainable development has made use of Proposition 5. But when valuing climate change, not all authors have made use of Eq. (9).

11.3 An application

Hamilton and Clemens (1999) and World Bank (2006) estimated comprehensive investment in the period 1970–2000 in over 120 countries. The authors added crude estimates of investment in human capital to official estimates of investment in reproducible capital. They then subtracted disinvestments in natural capital from that sum. For investment in human capital, the authors used expenditures on education as proxy. To quantify disinvestments in natural capital, they considered changes in the stocks of commercial forests, oil and minerals, and the concentration of carbon dioxide in the atmosphere. Oil and minerals were valued at their market price minus extraction costs. Forests were valued in terms of their market price minus logging costs. Contributions of forests to ecosystem functions were ignored.

In placing a value to increases in carbon concentration, Hamilton and Clemens (1999) and World Bank (2006) made a curious move. They took the shadow price of global carbon concentration to be -\$20 per ton (${}_m\Sigma g_m(t)$ in our notation) and multiplied that figure to the carbon emission rate of a country to arrive at the damage the country suffers from an increase in carbon concentration. To put it formally, “net benefit” to country m from emissions was identified as $[{}_m\Sigma g_m(t)]E_m(t)$, whereas, as Eq. (9) shows, the correct formula is $g_m(t)[{}_m\Sigma E_m(t)]$. The two expressions are equal only under very special circumstances (e.g., if countries are identical).

Arrow et al. (2008) have interpreted ${}_m\Sigma(g_m(t))$ as the global shadow price of carbon in the atmosphere (-\$20 per ton). Using the estimates in Mendelsohn et al. (2006), they arrived at a figure for $g_m(t)$ for country m by calculating the share of the global damage to m and multiplying $g_m(t)$ by the global emission rate ${}_m\Sigma E_m(t)$. In constructing the first column of figures in Table 1, I have followed that procedure.

The list of natural resources in Hamilton and Clemens (1999) was very incomplete. It did not include water resources, fisheries, air and water pollutants, soil, and other

Table 1 The progress of poor nations

Country/ region	I/Y ^a (%)	% Annual growth rate 1970-2000				
		Population (per head)	TFP ^b	Comprehensive wealth (per head)	GDP (per head)	Δ HDI ^c
Sub-Saharan Africa	-2.1	2.7	0.1	-2.81	-0.1	+
Bangladesh	7.1	2.2	0.7	-0.79	1.9	+
India	9.5	2.0	0.6	-0.45	3.0	+
Nepal	13.3	2.2	0.5	-0.37	1.9	+
Pakistan	8.8	2.7	0.4	-1.42	2.2	+
China	22.7	1.4	3.6	4.47	7.8	+

Adapted from Arrow et al. (2004) and Dasgupta (2001a).

^aComprehensive investment as a share of GDP (average over 1970-2000).

^bTotal factor productivity.

^cChange in HDI between 1970 and 2000.

ecosystems. On the other side of the ledger the authors' notion of human capital was inadequate. Health did not enter the calculus, even though life expectancy increased in most poor countries. Moreover, Hamilton and Clemens had nothing to say about growth in population, nor changes in TFP. In other words, their exercise did not come close to an analysis of sustainable development in the contemporary world, where population growth has been large and TFP growth has been nonnegligible. Furthermore, their estimates of shadow prices were very crude. Nevertheless, one has to start somewhere, and their attempt was a bold, first pass at what is proving to be an enormously difficult research program.

In Dasgupta (2001a) I adapted the Hamilton-Clemens estimates for South Asia, sub-Saharan Africa, and China, by taking into account population growth and public investment in education. The theoretical basis of my estimates was Proposition 5. But I ignored TFP growth in those economies because I did not know how to determine the shadow price of the "residual." Arrow et al. (2004) adapted the estimates in Dasgupta by including the residual. They then went on to ask whether economic development in South Asia, sub-Saharan Africa, and China has been sustainable in recent decades. Table 1 is a refinement of that publication. It remains a *very* crude beginning to the study of sustainable development, but again, it is a start.

The places in question are sub-Saharan Africa, Bangladesh, India, Nepal, Pakistan (all poor countries), and China (a middle income country). The period under study

is 1970–2000. The first column of numbers in Table 1 consists of refinements of the estimates of Hamilton and Clemens of comprehensive investment as a proportion of GDP. I have changed their figures by adding crude estimates of public expenditures on health and education and by estimating the “disinvestment” that corresponds to the increase in carbon concentration ($g_m(t)[_m\Sigma E_m(t)]$ for each m).

The figures in the second column are average annual population growth rates during the period and those in the third column are estimates of annual growth rates of TFP, which we interpret here as the annual percentage rate of change in a combined index of knowledge and institutions. To apply Proposition 5, I have divided the figures in the first column by the ratio of GDP to (comprehensive) wealth (the figures for which are explained later). To them I have simply added the estimates of TFP growth. That computation gives us an approximation to the (percentage) rates of change of (comprehensive) wealth at constant shadow prices. If I now subtract the growth rates in population from them, I arrive at estimates of the annual rate of change in wealth *per capita*. Those estimates are given in the fourth column.

In national income statistics, capital-output ratios are typically in the region 4–5 years. In our numerical computation, however, “capital” should be meant to read the shadow value of all capital assets *excepting for time* (Arrow et al., 2008). We therefore expect capital-GDP ratios to be greater than 4–5 years. Arrow et al. estimated that capital-GDP ratio in China is in the region 10–12 years, most probably it is even higher. In Table 1, which should be taken to be purely illustrative of the kind of empirical investigation in sustainable development that lies ahead of us, I have assumed the capital-GDP ratio to be 10 years.

Before summarizing the findings, it will be useful to get a feel for what Table 1 is showing us. Consider Pakistan. During 1970–2000 comprehensive investment as a proportion of GDP was 8.8%. TFP increased at an annual rate of 0.4%. As both numbers are positive, Pakistan’s productive base was larger in year 2000 than it had been in 1970. But take a look at Pakistan’s population, which grew at 2.7% rate annually. The fourth column shows that Pakistan’s *per capita* wealth *declined* in consequence, at an annual rate of 1.4%, implying that in year 2000 the average Pakistani was a lot poorer than in 1970.

Interestingly, if we were to judge Pakistan’s economic performance in terms of growth in GDP *per capita*, we would obtain a different picture. As the fifth column of Table 1 shows, Pakistan grew at a respectable 2.2% rate a year. If we now look at the sixth column, we find that the United Nations’ HDI for Pakistan improved during the period. Movements in GDP *per capita* and HDI tell us nothing about sustainable development.

The striking message of Table 1 is that during 1970–2000 economic development in *all* the countries on our list other than China (which was a poor country during much of the period under study) was negative. To be sure, sub-Saharan Africa offers no surprise. Comprehensive investment was *negative*, implying that the region *disinvested* in reproducible capital, human capital, and natural capital, taken together, at

an annual rate of 2.1% of GDP. Population grew at 2.7% a year and TFP barely advanced (annual growth rate: 0.1%). Even without performing any calculation, we should suspect that the productive base *per capita* in sub-Saharan Africa declined. The table confirms that it did, at 2.8% each year. If we now look at the fifth column of numbers, we discover that GDP *per capita* in sub-Saharan Africa remained pretty much constant. But the region's HDI showed an improvement—confirming once again that studying movements in HDI enables us to say nothing about sustainable development.

How does one explain the striking difference between movements in GDP and wealth during 1970–2000? There are two factors at work. First, the investment ratios in the first column of numbers in Table 1 are comprehensive investment ratios, and they are lower in each country than the recorded investment ratio. Secondly, and more importantly, the capital-output ratio I have used to convert figures in the first column of numbers is 10 years, not the 4–5 years that appear in national statistics.

Table 1 shows that Pakistan is the worst performer in the Indian subcontinent. But the remaining countries in the region also did not make it. Comprehensive investment in each country (Bangladesh, India, and Nepal) was positive, as was growth in TFP. The two together imply that the productive base expanded there. But population growth was sufficiently high to more than neutralize growth in comprehensive wealth.

Comprehensive investment in China was 22.7% of GDP, which is a large figure. The residual was a high 3.6% annually. Population grew at a relatively low rate: 1.4% per year. We should not be surprised that China's wealth *per capita* expanded—as it happens, at an annual rate of approximately 4.5%. *Per capita* GDP also grew at an annual rate of 7.8%, and HDI improved. In China, GDP *per capita*, HDI, and wealth per head moved parallel to one another.

The figures we have just studied are all *very* rough and ready, but they show how accounting for natural capital can make a substantial difference to our conception of the development process. We should remember that the figures for several shadow prices I used to arrive at Table 1 are conservative. For example, a price of \$20 per ton of carbon in the atmosphere is almost certainly a good deal below its true global social cost. On the other hand, Table 1 underestimates improvements in human capital, at least for India and China. If, instead of using expenditure on health as proxy, we were to value increases in life expectancy in terms of the value of a statistical life, comprehensive investment in those countries would be a lot higher. Nevertheless, with all the above caveats (and more!) in mind, the message we should take away is sobering.

It would be wrong though to think that people in poor countries should have invested more by consuming less. In poor countries the production and distribution of goods and services are highly inefficient, implying that consumption and comprehensive investment there do not compete for a fixed quantity of funds. Better institutions would enable people in the poor world to both consume more and invest more—comprehensively, of course!

12. EVALUATING POLICY REFORMS

Studying sustainable development requires that we compare an economy's wealth as it moves through time. Evaluating policy reforms requires that we estimate the welfare effects of perturbations to an economy at a point in time. Imagine that at date t the government considers making a small change to existing policies. The perturbation could be adjustments to the prevailing structure of taxes, it could be alterations to the existing system of property rights, or it could be a small conservation project. Call any such perturbation a *policy reform*. The government needs a criterion by which to judge whether the reform is socially desirable. As the perturbation is to be introduced at t , we have $\Delta t = 0$ in Eqs. (3) and (4), which therefore reduce, respectively, to

$$\Delta V(t) = \Sigma[I_i(t)] \quad (10)$$

and

$$\Delta V(t) = \Sigma[p_i(t)\Delta K_i(t)]. \quad (11)$$

Obviously, if the reform is not marginal, Eqs. (10) and (11) would be invalid. The traditional method of estimating $\Delta V(t)$ in such a case, which dates back to Dupuit's work in the nineteenth century, is to estimate the social surplus generated by the reform and add that surplus to the right-hand side of Eq. (11). An alternative method would be to estimate an "average" of the shadow prices that prevailed before the reform and those that would prevail if the reform were put in place, and then value changes in the capital assets at those prices. Formally the method is to apply the "mean value theorem." Writing by $p_i^*(t)$ the "average" shadow price of asset i , we have

$$\Delta V(t) = \Sigma[p_i^*(t)\Delta K_i(t)]. \quad (12)$$

12.1 Social cost-benefit analysis

For concreteness we now interpret a "policy reform" to be a *project*. Equations (10) and (11) say that a project's (comprehensive) investment measures its contribution to social well-being. So we have

Proposition 6. *A project should be accepted if and only if it would increase (comprehensive) wealth.*

Proposition 6 says that wealth is the criterion we should use for project evaluation. So project evaluation and the economics of sustainable development involve the same welfare criterion: wealth.²⁵

Call the flow of social well-being social *felicity*. Social well-being is most commonly taken to be the sum of the (discounted) flow of social felicity (Eq. 13 below). Frequently it is assumed too that felicity is a function solely of the flow of consumption (Eq. 13). Under those assumptions Proposition 5 can be restated as

Proposition 7. *Comprehensive investment measures the present discounted value of the resulting changes in consumption.*

Proof: See Appendix.

Proposition 7 is familiar. It is also the basis for social cost-benefit analysis, for we have

Proposition 8. *The criterion that should be used to evaluate projects socially is the present discounted value of the flow of the project's shadow profits.*

Together, Propositions 6 and 8 can be combined as

Proposition 9. *The present discounted value of a project's (net) profits equals its contribution to the economy's (comprehensive) wealth.*

I am unable to say who first proved Proposition 7. It is implicit in Ramsey (1928), who studied an economy at the full optimum. Marglin (1963) proved the proposition for a simple imperfect economy and used it to develop a theory of social cost-benefit analysis (Dasgupta et al., 1972). Our formulation here shows that Propositions 6-9 are very general, covering as they do any kind of institutional imperfection. The propositions form the welfare basis for responding to questions (D) and (E).

12.2 Evaluating projects in practice

Modern social cost-benefit analysis was developed for water resource management (Brown & McGuire, 1967; Eckstein, 1958). What drew the attention of development practitioners to the subject, however, was Little and Mirrlees (1968, 1974), who developed a methodology for use in poor countries. Although the Little-Mirrlees method was meant for imperfect economies, the imperfections it admitted were of a limited, structured kind. Dasgupta et al. (1972) formulated a method for social cost-benefit analysis that is applicable to a wider range of imperfect economies.²⁶ Neither publication, however, had a word to say about ecological capital.

Over the years a number of economists have developed the theory of policy evaluation so as to be applicable to environmental projects in poor countries (e.g., Andersen, 1987; Dasgupta, 1982, 2001a; Duraiappah, 2003; Hufschmidt, James, Meister, Bower, & Dixon, 1983; Newcombe, 1989; Sterner, 2003). Even so, there are very few case studies, and it is worth exploring why.

Although the theory of project evaluation was much discussed within international development agencies following the publications by Little and Mirrlees (see, for example, Squire & Van der Taak, 1975), the subject appears to have been dropped there in the early 1980s. As I understand it, even the World Bank abandoned social cost-benefit analysis.²⁷ And I cannot recall ever reading the annual *Human Development Report* of the

United Nations Development Programme in which a policy that was promoted by its authors had been subjected to serious economic evaluation. Elimination of extreme poverty has been the stated goal among international development agencies in recent years, but it would appear to be believed that the goal can be realized without the help of evaluation.

Those who have taken the logic of social cost-benefit analysis seriously are not the only ones who have complained. Proponents of “random experiments” as a method for evaluating the effectiveness of policy reforms in poor countries have complained too (e.g., Duflo & Kremer, 2005). However, social cost-benefit analysis, in the sense in which I am using the term here, is a different species of evaluation technique from randomized trials. The latter identifies the welfare impacts of a well-defined program, which should be interpreted as a policy reform. Randomized experiments would seem to be most suited to programs that target individuals or local communities in education, health, and sanitation. Kremer and Miguel (2006), for example, is an illuminating exercise. The investigators conducted randomized experiments on a Kenyan program involving the deworming of intestinal parasites. They found that there is no realistic alternative to large long-term external subsidies on deworming drugs. For example, charging a small fee for deworming drugs reduced the demand for treatment by 80%. Intensive school health education intervention had no effect on worm prevention behavior. It would seem that the private value on deworming is very low in Kenya.

The strengths and limitations of random experiments as a policy evaluation tool were brought out in a classic paper by Heckman (1992). In the context of environmental projects, it should be assumed that randomized experiments will have been conducted by ecologists to determine the response ecosystems make to disturbances (e.g., the toxicity of chemicals in water systems). We take it that feasibility reports of environmental projects are based on, among other pieces of information, the findings of such experiments. When a report arrives on a government decision-maker’s desk, however, her job is to evaluate it, period.

It has been said that social cost-benefit analysis is too difficult in practice. Estimating shadow prices is no easy matter. Future uncertainties abound and influence shadow prices. One way to work round that problem is to conduct sensitivity analysis, by varying parameter values in projects and evaluating the project variants. Often the best the evaluator can do is to offer a spread of recommendations: “Accept the project if the range of parameter values is R , reject it if the range is outside it” (see Dasgupta et al., 1972).

One way to make social cost-benefit analysis routine would be to conduct retrospective studies. One imagines that the World Bank, for example, has in store a large pool of project reports. A comparison could be made between a project’s eventual

performance and the expectations that had led to its acceptance. It has been said, for example, that infrastructure costs are almost invariably underestimated at the planning stage. If the claim is true, systematic cost overruns could be used to raise projected costs in current project evaluation. Ecological capital can be subjected to the same kind of exercise.

Duflo and Pande's work (2007) is an example of the value of retrospective evaluation. The authors studied the economic effects of large-scale irrigation dams in India. They found that while downstream districts gained (agricultural productivity increased and volatility in productivity declined), the districts where the dams were built enjoyed only a negligible increase in agricultural productivity even while they experienced a rise in volatility. The latter group could therefore be regarded as having become poorer. That should be useful information when dams are proposed in the future.

If policies are to be evaluated, there is no escaping social cost-benefit analysis. No doubt methodological corners have to be cut in practice, but anchoring evaluation exercises to theory forces practitioners to make clear what corners are being cut. That is good discipline. *Ad hoc* evaluation procedures, usually developed to suit the case in hand, can often be nothing more than a reflection of the evaluator's prior beliefs about the merits of the case. They may even reflect the evaluator's political prejudices.

Project evaluation, for example, is a way to determine whether restoring ecological capital is better than installing substitute forms of capital. In a well-known study Chichilnisky and Heal (1998) compared the costs of restoring the Catskill Watershed in New York State, whose ecological function in the past had been among other things to purify water, to the costs of building a water-purification plant, which would have been 8 billion US dollars. The authors showed the overwhelming economic advantages of restoration over construction. Independent of the many other services the Catskill watershed provides and ignoring the annual running costs of 300 million US dollars for a filtration plant, the capital costs alone showed a more than sixfold advantage for investing in the ecosystem.

The Chichilnisky-Heal study took the social objective (the supply of purified water to New York City) as given. They sought to identify the cost-effective way to achieve an incontrovertible objective. Project evaluation more generally compares the costs with the benefits to determine whether a project should be accepted. Perhaps for those reasons evaluation of ecological projects continues to be rare. Of the most comprehensive case studies I have come across, two were among the earliest in print, while the other pair is of recent vintage.

In a pioneering monograph on a pair of a-forestation projects in northern Nigeria, Andersen (1987) studied the contributions shelterbelts and farm trees, respectively, make

to both household and farm productivity—by supplying building material, fuel-wood, fruit, and fodder; and preserving soil and retaining moisture. The internal rate of return on investment in shelter-beds was found to be 15%; the corresponding figure for farm trees was 19%. It is hard to imagine that social discount rates would be anywhere as high as those figures. We should conclude that both were socially profitable projects.

In an equally interesting exercise, Newcombe (1989) found that population pressure had led to rural deforestation in the regions of Debre Zeit and Debre Berhan, Ethiopia. Subsistence farmers had therefore turned to dung as a source of household fuel. Newcombe showed that a-forestation would enable farmers to switch to wood-fuel as a source of household energy, thereby releasing dung for use as fertilizer. He estimated that the rates of return on such investment were in the astonishingly high range 35–70%, depending on assumptions made on fuel-wood yield and the productivity of dung in agriculture.

Kakujaha-Matundu and Perrings' study (2000) contains an exceptional cost-benefit analysis of the relative social merits of livestock and wildlife in the Nyae Nyae rangeland of Namibia. Livestock offer meat and milk to herders, while wildlife attracts tourists. The rate of return on livestock herding was found to be 10% a year, that on wildlife conservation about 14.5%. The authors showed that if the rangeland were to be optimally partitioned into the two activities, 70% of Nyae Nyae would be committed to wildlife conservation. Earlier, we noted that a PES system could be instituted to provide incentives to the inhabitants of the Nyae Nyae region to conserve wildlife.

The other study, by Whittington, Hanemann, Sadoff, and Jeuland (2009), evaluates alternative methods for clean water to households. The authors observe that the traditional method in rich countries involves centralized water supply sources and wastewater treatment facilities, combined with comprehensive pipe networks for water distribution and sewage collection. But they go on to show that this method may not be the most cost effective in poor countries. They do that first by evaluating a project in rural Africa that invests in the construction of deep boreholes with public hand pumps. The benefits that are considered include reductions in the incidence of diarrhea. At social discount rates of 3–6% a year the project's social profitability was found to be high (the benefit-cost ratio was approximately 3).

Whittington et al. (2009) also evaluate a community-led campaign in Bangladesh to free localities of open defecation. Among the costs is the construction of communal latrines. The authors recognize that the facilities would no be used if they are found to be inconvenient. Even so, at social discount rates of 3–6%, the benefit-cost ratios of the program were found to range from 2.4 to 7.5 depending on the sites chosen. The moral is important: in designing the supply of household water and sanitation facilities, the context matters.

13. DISCOUNTING CLIMATE CHANGE

Perhaps the most vigorous application of social cost-benefit analysis in recent years has been to the economics of climate change. Prominent examples are Cline (1992), Nordhaus (1994, 2008), and Stern (2006). That Cline and Stern reached a very different conclusion on the optimal *timing* of global efforts to curb climate change from the one reached by Nordhaus was noted and discussed in Dasgupta (2007b) and Nordhaus (2007). As intergenerational ethics lies at the heart of the welfare economics of climate change, it is useful to review the issues.

13.1 Global activism now or later?

The current concentration of carbon dioxide in the atmosphere is 385 parts per million (ppm). We are to imagine that there is a world government that can enforce whatever action is required to implement a *world optimum*. Suppose the global objective is to limit concentration to 500 ppm. Cline and Stern argued that the world should spend substantial sums starting today (Stern's figure amounts to 2% of the GDP of rich nations). Nordhaus (1994, 2008) in contrast has argued that it would be more equitable and efficient to invest in reproducible capital and human capital now so as to build up the productive base of economies—including, especially, poor countries. He has proposed an upward sloping climate policy ramp of ever tightening reductions in carbon emissions over the years.²⁸

The difference between the two recommendations can be traced to differences in the authors' conceptions of intergenerational equity. We confirm below that as global warming involves the long run (100 years and more), economic evaluation of climate policies is sensitive to the choice of social discount rates.

The *numeraire* in social cost-benefit analysis is typically a determinant of social felicity. Most commonly the chosen determinant is aggregate consumption (or income) per head. We continue to assume a deterministic world. As before $V(t)$ denote social well-being at t and let $U(t)$ be social felicity *per capita* at t . If $c(t)$ is consumption per head, all three authors assume $U(t) = U(c(t))$. The present value shadow price of $c(t)$ is called the *social discount factor*. The *social discount rate* at t , which we write as $\rho(t)$, is defined as the percentage rate of decline in the social discount factor. Suppose by way of illustration (but see Appendix) that social well-being at $t = 0$ is

$$V(0) = \int_0^\infty [U(c(t))e^{-\delta t}]dt, \quad \text{where } \delta \geq 0. \quad (13)$$

Until Section 12.4 we are to interpret c as *per capita* world consumption and $V(0)$ in Eq. (13) as global well-being. Assume that the integral in expression (13) exists. In Appendix, we confirm that in consumption *numeraire* the social discount factor at t is $[dU(c(t))/dc(t)]e^{-\delta t}$. Let $\alpha(c(t))$ be the elasticity of marginal felicity. As is well known,

social aversion to consumption inequality among individuals requires that $\alpha(c(t)) > 0$. I follow the three authors by assuming that α is independent of $c(t)$. Write the percentage rate of change in $c(t)$ as $g(c(t))$. A simple calculation shows that

$$\rho(t) = \delta + \alpha g(c(t)). \quad (14)$$

13.2 Negative discount rates

We confirm below that Cline, Nordhaus, and Stern have assumed that $\alpha \geq 1$. From Eq. (14), we note that if $g(c(t))$ is forecast to be positive, $\rho(t)$ is positive. Notice though that if $g(c(t))$ is forecast to be negative and δ is small, then $\rho(t)$ is negative. To illustrate, suppose $\delta = 0$, $\eta = 2$, and $g(C_t) = -1\%$ per year. Then Eq. (14) says $\rho_t = -2\%$ per year. I have colleagues in the United States who find illustrations involving negative economic growth to be unrealistic. Recall though from Table 1 that sub-Saharan Africa suffered from negative growth during 1970–2000. What discount rates should government project evaluators there have chosen in 1970 if they had an approximately correct forecast of the shape of things to come?

That negative growth rates in consumption can imply negative social rates discount rates is significant because people in the tropics are likely to suffer greatly from climate change under business as usual. Moreover, the possibility of negative growth in consumption takes an interesting turn when we come to consider uncertainty in future consumption (see Section 12.4).

13.3 The welfare economics of climate change

The most-preferred values of δ and α in Cline (1992), Nordhaus (1994), and Stern (2006), respectively, were as follows:

Cline: $\delta = 0$; $\alpha = 1.5$

Nordhaus: $\delta = 3\%$ a year; $\alpha = 1$

Stern: $\delta = 0.1\%$ a year; $\alpha = 1$

Notice the closeness between the parameter values assumed by Cline and Stern. That explains why their recommendations on the timing of expenditure on climate change were very similar. Unfortunately, despite the closeness, there was no reference to Cline's work in the version of Stern (2006) that was circulated prior to publication. And there was only a perfunctory reference to Cline in Stern and no mention that their recommendations were similar. The custom of having texts screened by independent referees was designed more than three centuries ago to prevent such acts of omission. Priority should matter.

The point estimate of $g(c(t))$ under business as usual in Stern (2006) was 1.3% a year. Using this figure in Eq. (14) yields

$$\begin{aligned}\rho(t) &= 2.05\% \text{ a year for Cline} \\ \rho(t) &= 4.30\% \text{ a year for Nordhaus} \\ \rho(t) &= 1.40\% \text{ a year for Stern}\end{aligned}$$

4.3% a year may not seem very different from 1.4% a year, but is in fact a lot higher when put to work on the economics of the long run. Just how much higher can be seen from the fact that the present value of a given loss in consumption, owing, say, to climate change 100 years from now, if discounted at 4.3% a year is *seventeen* times smaller than the present value of that same consumption loss if the discount rate used is 1.4% a year. The moral is banal: If the time horizon is long, even small differences in social discount rates can mean large differences in the message cost-benefit analysis gives us. The difference between the prescriptions in Cline (1992) and Stern (2006), on the one hand, and Nordhaus (1994), on the other, has little to do with differences in the climate models the authors worked with, they have all to do with differences in the way the authors interpret intergenerational equity. Nordhaus (2007) confirms that by using Stern's specifications for δ and α in the climate-change model he has developed over the past two decades. It should be noted that in our numerical illustration, Nordhaus' choice of $\rho(t) = 4.30\%$ a year would be consistent with the US government's discount rate policy (on the latter, see Viscusi, 2007).

In a recent book Nordhaus (2008) has chosen a figure of 4% a year to discount changes in future consumption. He has justified the figure on grounds that it is approximately the long-run risk-free interest rate in the United States, which amounts to the claim that the risk-free rate is the social rate of return on investment.

I have not understood the underlying logic. The social discount rate on consumption equals the social rate of return on investment in a fully optimum economy. But 4% a year is an observed figure in a world riddled with the externalities carbon emissions give rise to. The atmosphere is even now more or less an open-access global common. For all we know the social rate of return on certain forms of investment (e.g., carbon intensive investment) is today negative. Over three decades ago Brock (1977) noted relatedly that the marginal product of private capital would be biased upward in a market economy due to an overuse of nature's inputs. His argument was that the marginal product of capital increases with nature's inputs and that the latter are underpriced. Of course, it is legitimate to select *investment* as *numeraire* and base the justification of 4% a year on that choice. But if one selects investment as *numeraire* in an imperfect economy, consumption must be revalued at its shadow price (Dasgupta et al., 1972; Little and Mirrlees, 1974; Marglin, 1963). Nordhaus does not do that revaluation.

It should be a requirement of any exercise in social cost-benefit analysis that it be subjected to sensitivity tests. In the economics of climate change, alternative climate scenarios, taken from publications of the Intergovernmental Panel for Climate Change (IPCC) have featured prominently. But social cost-benefit analysis involves the use of

“value” parameters too. In the formulation adopted by Cline, Nordhaus, and Stern, the value parameters are δ and α . Curiously, Stern’s work (2006) contains no simulation in which δ and α were made to assume alternative numerical values. It is almost as though the values Stern chose came from Government House. But that is hardly the way to do welfare economics. We have little prior intuition of what δ and α imply in elaborate models of global climate change, which is all the more reason why project evaluation should include alternative figures for value parameters.

It has been argued by philosophers that, leaving aside the possibility of collective extinction, δ should be taken to be zero (Broome, 1992; Rawls, 1972). On the other hand, Koopmans (1960, 1972) proved that a set of intuitively appealing ethical axioms on infinite felicity streams, when taken together, demand that δ should be positive. In Dasgupta (2008b) I put δ and α to a test by calculating the optimum rate of consumption in a classroom model of consumption and capital accumulation. Assuming $\delta = 0.1\%$ a year and the rate of return on investment to be 4% a year (which is an altogether very generous figure), I found that if $\alpha = 1$, the optimum saving rate is over 95% of GDP. But a 95% saving rate is an absurdly heavy burden on the present generation, in a world where future generations will be increasingly rich owing to the productivity of capital. The calculation suggested, at least to me, that $\alpha = 1$ reflects an ethics that is insensitive to intergenerational inequality.

13.4 Social aversions to inequality and risk

Larger values of α suggest themselves, in the range 1.5–3 (see the exercises in Dasgupta, 2008b). As α is a measure of *inequality aversion*, larger values would recommend *lower* rates of saving in a deterministic world. That would lessen the burden on the current generation. But the future is uncertain. If social well-being under uncertainty is taken to be the expectation of the present discounted value of the sum of social felicities, α is a measure of *risk aversion* as well. It can be shown that if $\alpha > 1$, uncertainty in the rate of return on investment is a reason for saving *more* for the future (Levhari & Srinivasan, 1969). The message is awkward: as α has a dual role in intergenerational ethics, it creates a tension between two opposing considerations.

Let $\tilde{c}(t)$ be the uncertain consumption level per head at t . For mathematical tractability, I assume now that time is discrete ($t = 0, 1, \dots$). Equation (14) says that if E is the expectation operator, expected social well-being can be written as

$$E[V(0)] = E\left\{{}_0\Sigma^\infty [U(\tilde{c}(t))/(1 + \delta)^t]\right\}, \quad \delta \geq 0. \quad (15)$$

Pesaran, Pettenuzzo, and Timmermann (2007) and Weitzman (2007) have shown that if uncertainty over future consumption possibilities is determined econometrically from past observations, it will have a “thick” lower tail (e.g., a power function), implying that there is a positive chance of consumption dropping as close to zero as you could

fear. Weitzman gave color to his analysis by alluding to “catastrophic” events that could come in the wake of climate change. The authors showed that if $\alpha \geq 1$, and the horizon is infinite, and the uncertainty has a thick lower tail, the precautionary motive for saving becomes so great as to render the concept of intergenerational justice embodied in Eq. (15) incoherent. Formally, the integral in Eq. (15) diverges and social cost-benefit analysis becomes meaningless.

The matter is worse, in that a much stronger result has been known for years. Dasgupta (2008b) used the model of capital accumulation in Levhari and Srinivasan (1969) to show that if $\alpha > 1$ and the horizon is infinite and the uncertainty in future consumption possibilities is “large” relative to other parameters, expression (15) does not converge *even if* the probability distribution over $c(t)$ at each date is “thin” tailed, as in the case of the log normal distribution.

How disturbed should we be by this result? Not much, or so it seems to me. The most recent publications of IPCC contain only a finite number of scenarios, none of which considers it to be even a remote possibility that consumption will decline to zero. Moreover, a major artificiality in the models studied by Cline, Nordhaus, and Stern is the assumption that $\alpha \geq 1$. Many years ago Arrow (1965) offered reasons as to why we should work with felicity functions that are bounded at both ends. He argued in favor of U -functions for which $\alpha(c)$ increases monotonically from values less than one to values greater than one. If we invoke such felicity functions, the paradox of infinity generated by large risks would not occur.²⁹

Global project evaluation should involve the use of a more disaggregated version of social well-being than expression (15). Let c_p and c_r , respectively, be average consumption in the poor and rich worlds. And let N_p and N_r be their respective population sizes. An appealing extension of expression (15) would be

$$V(0) = E\left\{ \sum_0^\infty [[L_p(t)U(\tilde{c}_p(t)) + L_r(t)U(\tilde{c}_r(t))] / (1 + \delta)^t] \right\}, \quad \delta \geq 0. \quad (16)$$

If α is taken to be constant in expression (16), it takes on three roles. It is a measure of inequality aversion among contemporaries, across the generations, and across uncertain contingencies. And that may be too much for a single parameter to do adequately. Which is also why nonconstant α s suggest themselves.

13.5 Climate treaties

Maximizing expression (13) (more so, expression (16)) presumes an enlightened world government. So, the social cost-benefit analysis reported in, for example, Stern’s study (2006) is a pure abstraction. In the absence of an overarching authority we can at best hope for agreements among nations that are self-enforcing. Anthropogenic climate change involves reciprocal externalities (Section 6), which is why the formation of coalitions in the face of reciprocal externalities and the policies that are likely to emerge on the basis

of international negotiations has received much attention over the years (Barrett, 1994, 2003; Carraro, 2002; Carraro & Fragnelli, 2004; Dutta & Radner, 2004; Finus & Runshagen, 2003; Uzawa, 2003 are among the prominent contributions). By “self-enforcing,” one means an agreement with the property that if a party expects those others who have signed the agreement to abide by it, then it is in the interest of that party to abide by it too; that is, the agreement is a Nash equilibrium. In negotiations over climate change, the parties are countries. Barrett (1994, 2003) and Dutta and Radner, among others, have argued that the Kyoto Protocol did not lay the groundwork for a self-enforcing treaty. They used that to explain why the Protocol has been a disappointment.

As countries differ in size, wealth, and location, Nash equilibrium is not unique. Among equilibrium outcomes is the “null-treaty,” meaning global noncooperation, commonly referred to as “business as usual.” Treaties would however differ in their efficiency and in the distribution of benefits and burdens. It is a feature of equilibrium outcomes that not all countries would be party to potential treaties; some (among them some small countries) would free ride. Among the choices to be made in designing a treaty are adaptation and mitigation measures. The costs and benefits involving the two kinds of investment would be expected to differ among countries. So, economists who study the political economy of climate change face the problem of having to explain which equilibrium would be selected. Factors outside theoretical models would be particularly relevant here. The power of rich countries could be expected to tilt the selection toward their favor.

In an interesting and important paper, Barrett (2008) observes that although it is frequently claimed that adaptation and mitigation are complements, they are more like substitutes. As countries invest more in the former, they suffer less from climate change and find mitigation less attractive. But mitigation is a global public good (“windmills”), whereas adaptation is a national public good (“dikes”). One can imagine a situation where the globally optimal investment policy obtained from maximizing expression (16) would have every country invest in windmills, but where the noncooperative equilibrium is one where each country constructs only dikes. Imagine that the ideal treaty (with appropriate, credible side payments) sustains a high level of participation and requires so many windmills to be built that no one needs to construct dikes. Barrett constructs examples where, nevertheless, the treaties that are signed are ones under which rich countries construct dikes and pollute the atmosphere, leaving poor countries not so much high and dry, as “low and wet.” Such an ominous possibility cannot yet be ruled out.

14. CONCLUSIONS

In this chapter I have reviewed what seems to me to be the most salient issues in ecological economics when the subject is applied to the field of economic development. My aim here has not been to be scholastic but to examine the lives of the world’s poor

so as to unearth the role of natural capital there. I will have succeeded if the account here of the processes that characterize human-nature interactions reads differently from the accounts in recent surveys of both development economics and environmental and resource economics.

I began with the microfoundations of rural institutions in poor regions and offered a picture of rural poverty in terms of household access to the local natural-resource base. The findings documented bring to date those I reported in Dasgupta (1993). This chapter is a natural extension of the earlier work and has advanced very much the same viewpoint as that did about the processes that shape world poverty.

The concept of sustainable development was explored in a macroeconomic setting and an empirical study was conducted on the character of economic development in the world's poorest regions in the last quarter of the previous century. Natural capital was for the most part seen as "private" goods, in the sense of not being jointly consumable. In the final section there was a discussion of global climate as a "public" good.

What can we conclude from our analysis? It seems to me the following should be noted:

- (1) The socioecological processes that shape extreme poverty in the world's poor regions run at different speeds and operate at different spatial scales. Disasters occur frequently in the poor world, but unlike famines, civil wars, and hurricanes, their occurrence is localized and confined to small groups. That is why it is easy to overlook them.
- (2) The externalities that the use of ecological capital gives rise to are not confined to market failure, they are expressions of institutional failure in its widest sense: failure at the international level, the level of the state, of communities, of households. The locus of failure depends, among other things, on the ecological capital in question. The cause of eutrophication of a village pond in West Bengal is very different from the cause of dead zones in the Gulf of Mexico.
- (3) We should be circumspect about market-friendly solutions to environmental problems. Externalities markets are inevitably thin, meaning that without a sympathetic involvement of the state, the elite would be expected to enjoy the spoils from ecological services.
- (4) The protection and promotion of ecological capital is encouraged in systems where payment is made to owners for the ecological services provided by their capital assets. Whether payment for ecosystem services should be made by the beneficiaries or the state depends on the context.
- (5) The persistence of rural poverty is tied both to the fragile state of the local natural-resource base and the rate of population growth. But the causality is not unidirectional. Each variable would be endogenous in any model that speaks to human-nature interchanges.

- (6) Both economic theory and empirical studies have shown that devolution of management powers over the local natural-resource base is, generally speaking, good for the environment and good for poverty alleviation. But as elsewhere in economics, mixed systems work best. Textbook states would ensure that the local elite do not take a disproportionate amount from the commons. Where the state is weak or corrupt, NGOs can play a major role in keeping the state at bay and the elite from enjoying the bulk of the services from nature.
- (7) As a macroeconomic statistics of social well-being, GDP *per capita* is singularly bad. So is the United Nations' HDI. Among other shortcomings, GDP per head and HDI ignore depreciation of capital. The deficiency can be alarming in the world we live in (Table 1).
- (8) The statistic that moves in unison with social well-being (by the latter we mean an aggregate of the well-beings of the present and all future generations) is a comprehensive measure of wealth, which is the social worth of an economy's entire stock of capital assets, including not only reproducible capital, human capital, and knowledge and institutions, but also natural capital.
- (9) Comprehensive wealth (or wealth for short) can be used both for evaluation exercises and for assessing whether development has been (or is forecast to be) sustainable.
- (10) Although there are still only a few rigorous studies in social cost-benefit analysis of environmental projects, the message we should take away from them is that projects that protect and promote natural capital can be socially very profitable.
- (11) The relative appeal of alternative policies toward mitigating or adapting to climate change is sensitive to the choice of social discount rates. As we still have little intuitive understanding of parameters reflecting ethical values, evaluation exercises should contain sensitivity analysis.
- (12) Statistics (albeit very crude) suggest that in the final quarter of the twentieth century South Asia and sub-Saharan Africa experienced a decline in wealth per head, even though South Asia enjoyed positive growth in GDP per head and improvements in HDI, while sub-Saharan Africa enjoyed an improvement in HDI but experienced a small decline in GDP per head. The data suggest that China in contrast followed a path of sustainable development. It bears emphasis though that the empirical exercise involving Table 1 is so crude and incomplete that the Chinese data misrepresent the situation there. Despite the caveats, the moral is that the macroeconomic history of nations looks very different when nature is included as a capital asset in economic activity.
- (13) The problem of climate change as faced by poor countries can only be addressed at the collective level of nations. But case studies suggest that so far the environmental problems the rural poor face have been caused by institutional failure at the national and community levels. The composition of commodity demands from rich countries can certainly veer poor countries toward unsustainable

resource use. But poor countries usually have choices. Moreover, there is enough inefficiency in poor countries to enable governments there to identify policies that both protect and promote natural capital and alleviate poverty. The idea that the poor world can enjoy sustainable development only when there are significant improvement in the international economic architecture is belied by evidence on village life in poor countries.

The overarching moral that emerges from the studies I have presented here may appear banal, but is salutary:

Development policies that ignore our reliance on ecological capital are seriously harmful—they don't pass the mildest test for equity among contemporaries, nor among people separated by time and uncertain contingencies.

APPENDIX

We develop the simplest version of Propositions 1–4. As in the text, assume that time is continuous. Denote it variously by s and $t(s \geq t \geq 0)$. So as to keep the analysis to essentials, we avoid talking explicitly about future uncertainty; but readers may, if they so choose, interpret social well-being (expression (A.1) below) as *expected* social well-being, and future consumption flows and capital stocks as state contingent vectors.

Because it is not possible to specify a final date for the economy with certainty, the future is taken to be infinitely long (expression (A.1)).

A.1. ECONOMIC PERTURBATION AS MOVEMENT THROUGH TIME: SUSTAINABLE DEVELOPMENT

What follows is a formal account of the theory developed in Section 10.1. Let $\underline{C}(s)$ denote a vector of consumption flows at time s . $\underline{C}(s)$ includes not only standard consumption goods, but also leisure and consumption services supplied by nature. Some consumption goods are marketed, others are not. Consumption goods are indexed by j . Let $K(s)$ denote the quantities of a comprehensive list of capital assets at s . Those quantities are stocks. For simplicity, assume that demographic changes, movements in TFP, and changes in import prices are exogenous. We will incorporate those exogenous changes into the notion of comprehensive investment by regarding time also as a capital asset.³⁰ As in the text, all capital assets other than time are indexed by i .

A.1.1 The formal analysis

To fix ideas, assume population is constant. Let $U(\underline{C}(s), K(s))$ be (social) *felicity* at s . Denote social well-being at t by $V(t)$. $V(t)$ is a stock. For concreteness we assume that

$$V(t) = \int_t^\infty \left[U(\underline{C}(s), \underline{K}(s)) e^{-\delta(s-t)} \right] ds, \quad \delta \geq 0. \quad (\text{A.1})$$

δ is the felicity discount rate, which should not be confused with discount rates to be applied to consumption, or investment, or income (see below). An economic forecast at t is the infinite sequence $\{\underline{C}(s), \underline{K}(s)\}$ for $s \geq t$. We assume that the integral in expression (A.1) converges for the forecast. Definition 2 says that development is sustained at t if

$$dV/dt \geq 0. \quad (\text{A.2})$$

We now prove that (A.2) is satisfied if and only if comprehensive investment at t is nonnegative (Proposition 3).

To save on notation, I avoid writing down an explicit dynamical model of the economy and merely note from Eq. (A.1) that even though sustainability has been defined for one moment of time t (condition (A.2)), the condition requires a forecast of the economy's future beyond t .³¹ That future depends on the economy's stock of assets at t . It also depends on the evolving structure of its technology, people's values and preferences, and institutions beyond t . The stock of assets at any moment s in the future would be determined by the stocks at the "previous" date.³² By proceeding from moment to moment this way, the entire future course of capital stocks would be determined. In short, if $\underline{K}(t)$ were known, $\underline{K}(s)$ and $\underline{C}(s)$, and thereby $U(\underline{C}(s), \underline{K}(s))$, could be determined for all future times $s \geq t$. From Eq. (A.1), $V(t)$ could be determined as well. Therefore we can write

$$V(t) = V(\underline{K}(t), t). \quad (\text{A.3})$$

In Eq. (A.3) V depends explicitly on t because the economy is assumed to undergo changes owing to exogenous factors.

Let $q_j(t)$ denote the shadow price of consumption good j at t . Then

$$q_j(t) = \partial U(\underline{C}(t), \underline{K}(t)) / \partial C_j(t). \quad (\text{A.4})$$

But Proposition 3 says that the shadow prices we need for sustainability analysis are those for the economy's capital assets.

Let us assume without justification that $V(t)$ is differentiable.³³ Differentiating $V(t)$ with respect to t in (A.3) and using (A.2) yields the criterion for sustainable development at t :

$$dV(t)/dt = \partial V/\partial t + \sum [(\partial V(t)/\partial K_i(t))(dK_i(t)/dt)] \geq 0. \quad (\text{A.5})$$

As in Section 9, define

$$p_i(t) = \partial V(t)/\partial K_i(t), \quad \text{for all } i. \quad (\text{A.6})$$

$p_i(t)$ is the (spot) shadow price of the i th asset at t (Definition 1). If i is a factor of production as well as a final consumption good (e.g., a wetland), $p_i(t)$ reflects both. From expressions (A.1), (A.4), and (A.6), we note that the shadow price of consumption goods at all s (the $q_j(s)$ s) are embodied in the shadow price of capital assets at t (the $p_i(t)$ s).

As in Section 9, write $I_i(t) = p_i(t)(dK_i(t)/dt)$ for net investment in asset i . Using Eq. (A.6) in Eq. (A.5) gives us

$$dV(t)/dt = \partial V/\partial t + \sum [p_i(t)(dK_i(t)/dt)] \geq 0 \quad (\text{A.7})$$

or

$$= \partial V/\partial t + \sum [I_i(t)]. \quad (\text{A.8})$$

An easy way to interpret $\partial V/\partial t$ is to regard time itself as a capital asset, say Z . If $Z = t$, we have

$$dZ/dt = 1. \quad (\text{A.9})$$

With Eq. (A.9) defining the additional asset, $\partial V/\partial t$ ($= \partial V/\partial Z$) becomes the shadow price of time and the right-hand side of Eq. (A.8) becomes comprehensive investment at t . This proves Proposition 3.

A.1.2 Accounting for population growth

Population is a capital asset. In macroeconomic models population growth is usually assumed to be exogenous. We make the same assumption here. Let us also assume that population cohorts are identical in their preferences and abilities. Then the size of the population, $P(t)$, is the stock of the demographic asset. It may seem intuitive that the way to tease exogenous growth in population out of $\partial V/\partial t$ is to define comprehensive wealth in *per capita* terms and reexpress Proposition 1 accordingly. Dasgupta (2001a) showed that to be a correct move only under very special circumstances. Let us see why in general the move is illegitimate.

For simplicity of exposition let us revert to the case where there is a single consumption good. Assume as earlier that felicity does not depend on the stock of capital assets. Write $c(t) = C(t)/P(t)$. In his classic work on optimum saving under constant population growth, Koopmans (1965, 1967) assumed social well-being to be the present discounted sum of each generation's *average* felicity, where average felicity is a

function solely of average consumption. If, within each generation, consumption is equally distributed, Koopmans' $V(t)$ assumes the form (Eq. 12 in the text),

$$V(t) = \int_t^\infty [U(c(s))e^{-\delta(s-t)}]ds, \quad \delta \geq 0. \quad (\text{A.10})$$

Meade (1955) had however already drawn attention to a deep problem with expression (A.10): it discriminates against future people if population increases with time.

An alternative (studied in the context of optimum saving, by Arrow and Kurz, 1970; Arrow, Dasgupta, & Mäler, 2003b; Meade, 1955; Mirrlees, 1967; and in the context of optimum saving and population by Dasgupta, 1969) is the present discounted sum of each generation's total felicity:

$$V(t) = \int_t^\infty [P(s)U(c(s))e^{-\delta(s-t)}]ds. \quad (\text{A.11})$$

Arrow et al. (2003b) showed that if expression (A.11) is used for studying sustainable development, we would need to specify the level of consumption, c , at which $U(c) = 0$, which would mean that when specifying U we would have only one degree of freedom (the scale of U). In the problem of optimum saving (as in Arrow & Kurz, 1970) we would not be required to do that, because we are free to choose both the scale and level of U .³⁴

It would be convenient in preparing national accounts if the level of U , not just its scale, could be freely chosen. So consider the following expression for social well-being:

$$V(t) = \frac{\int_t^\infty [P(s)U(c(s))e^{-\delta(s-t)}]ds}{\int_t^\infty [P(s)e^{-\delta(s-t)}]ds}. \quad (\text{A.12})$$

The numerator in expression (A.12) is expression (A.11), whereas the denominator is the present discounted sum of each generation's population. Let us call the ethical theory on which expression (A.12) is based, *dynamic average utilitarianism*.

Notice that the denominator in expression (A.12) would play no role in policy evaluation *at t* (questions (D) and (E); Section 11), because the denominator would simply be a scale factor attached to expression (A.11). But for sustainability analysis (questions (B) and (C); Section 10) the denominator matters, because the evaluation there is undertaken *across time*.

Let $k_i(t) = K_i(t)/P(t)$. Now write $k = (k_1, k_2, \dots, k_i, \dots)$. From expression (A.12), we have

$$V(t) = V(k(t), P(t)). \quad (\text{A.13})$$

Dasgupta (2001a) showed that if P grows (or declines) at a constant rate and if each of the equations that represent the economy's accumulation process can be expressed in terms solely of *per capita* capital stocks, then $\partial V(t)/\partial P(t) = 0$. That proves Proposition 4.

A.2. POLICY EVALUATION

Proposition 6 offered a well-known interpretation of comprehensive investment. To prove it, imagine that the vector of assets at t is not $\underline{K}(t)$ but $\underline{K}(t) + \Delta\underline{K}(t)$, where Δ is an operator denoting a small difference. For simplicity of exposition, suppose (a) there is a single consumption good, (b) population is constant, and (c) felicity depends solely on consumption. In the obvious notation, we then have

$$V(\underline{K}(t) + \Delta\underline{K}(t)) = \int_t^\infty [U'(C(s))\Delta C(s)e^{-\delta(s-t)}] ds. \quad (\text{A.14})$$

Now imagine that at t there a small increase in investment, but only for a brief moment, Δt . We write the change in the vector of capital assets at $t + \Delta t$ consequent upon the brief increase in investment as $\Delta\underline{K}(t)$. So $(\underline{K}(t + \Delta t) + \Delta\underline{K}(t))$ is the resulting vector of capital assets at $t + \Delta t$. Let Δt tend to zero. Equation (A.14) then yields Proposition 6.

Earlier we observed that δ in expression (A.1) is the felicity discount rate. (In the literature on environmental and resource economics δ is frequently called the “pure rate of time preference.”) It is the discount rate to be used if a policy reform (or a project) is described in terms of the changes it brings about to felicities over time. However, policy reforms are typically characterized in terms of the perturbations they cause to *commodity* flows (e.g., consumption flows); which is why it has been customary in social cost-benefit analysis to regard consumption as the *numeraire*. Equation (A.14) says that, viewed from t , the social discount factor at date s is $\partial U/\partial(C(s))e^{-\delta(s-t)}$. Let $\rho(s)$ be the social discount rate at s (the percentage rate of decline of the consumption discount factor). An easy calculation shows that

$$\rho(s) = -[C(s)U''(C(s))/U'(C(s))][dC(s)/ds]/C(s) + \delta. \quad (\text{A.15})$$

But $-C(s)U''(C(s)) / U'(C(s))$ is the elasticity of marginal felicity. If we denote that by $\alpha(C(s))$, Eq. (A.15) reduces to the familiar form

$$\rho(s) = \alpha(C(s))[dC(s)/ds]/C(s) + \delta. \quad (\text{A.16})$$

If equality in consumption is valued, then $U''(C(s)) < 0$; but that means $\alpha(C(s)) > 0$.

It has been common to assume that α is constant. To the best of my knowledge the whole literature on the economics of climate change has been based on that assumption. Equation (A.16) says that even if $\alpha(C(s))$ is taken to be a constant, $\rho(s)$ would be constant only along steady states. The equation also says that $\rho(s)$ is negative if consumption is expected to decline at s (Dasgupta, 2001a). As we noted in Section 11.3, this has far-reaching consequences for the choice of social discount rates when the future is uncertain.

End Notes

- *. In preparing this article I have drawn on my Yan Fu Memorial Lecture at Peking University, 2005; Bingham Lecture at the Research Triangle Institute (Durham, NC), 2006; the Annual Distinguished Speakers Lecture at the Asian Development Bank (Manila), 2007; the Wrigley Lecture at Arizona State University (Tempe), 2007; and the John Kenneth Galbraith Lecture at the American Agricultural Economics Association Annual Conference (Portland, Oregon), 2007. My understanding of ecological and development economics has been much influenced over the years by discussions with Toke Aidt, Kenneth Arrow, Scott Barrett, William ("Buzz") Brock, William Clark, Gretchen Daily, Anantha Duraiappah, Paul Ehrlich, Kirk Hamilton, Rashid Hassan, Sriya Iyer, Robert Kates, Pramila Krishnan, Karl-Göran Mäler, Pranab Mukhopadhyay, Subhrendu Pattanayak, Charles Perrings, Robert Scholes, Ismail Serageldin, Priya Shyamsundar, V. Kerry Smith, E. Somanathan, Jeff Vincent, and the many scholars associated with the Beijer Institute of Ecological Economics and the South Asian Network for Development and Environmental Economics (SANDEE), who are too numerous to name individually. In preparing this chapter I have benefited enormously from the editorial comments of Dani Rodrik.
- 1. The author is the Frank Ramsey Professor of Economics at the University of Cambridge; Fellow of St John's College, Cambridge; and Professor of Environmental and Development Economics at the University of Manchester. Email address: partha.dasgupta@econ.cam.ac.uk.
- 2. An "ecosystem" is a complex of the a-biotic environment and plant, animal, fungi, and microorganism communities, interacting as a functional unit. Humans are an integral part of ecosystems. Brauman, Daily, Ka'eo Duarte, and Mooney (2007), Hassan et al. (2005), MEA (2003), and Reid et al. (2005) offer excellent accounts of the role ecosystems play in economic life.
- 3. Barro and Sala-i-Martin (2003), Helpman (2004), and the survey articles in Aghion and Durlauf (2005) provide excellent examples.
- 4. For many years now I have both grumbled about the continual neglect of the natural-resource base in the study of poverty in poor countries and tried to bring together environmental and development economics (Dasgupta 1982, 1990, 1993, 2001a, 2003, 2004, 2008a, 2008b; Dasgupta & Mäler, 1991). Since its inception in 1996, the journal *Environment and Development Economics* has done much to advance the subject.
- 5. Scholars at *Resources for the Future* have been prominent in developing that excellent body of work. Kneese and Schultze (1975) are a fine representative. Kolstad (1999) and Siebert's study (2008) are excellent texts on the subject, covering both theory and optimal policy.
- 6. The publications that drew attention to the economic importance of ecosystems are the classics by Ehrlich, Ehrlich, and Holdren (1977) and Daily (1997). Brauman et al. (2007) and Hassan et al. (2005) provide a catalogue of those services. For a succinct account of policy measures that are currently being taken across the world with ecological services as the focus, see Turner and Daily (2008).

7. See Tomich, van Noordwijk, and Thomas (2004), Tomich et al. (2005), Palm, Vosti, Sanchez, and Erickson (2005), and Hassan et al. (2005), respectively, on those tensions.
8. This classification was explored in Dasgupta (1982).
9. See Carpenter (2001) and Carpenter, Ludwig, and Brock (1999) for an analysis of the dynamics of (shallow) fresh-water lake systems. The existence of multiple basins of attraction has important implications for the way the lake's shadow price ought to be estimated. Dasgupta and Mäler's study (2003) is a collection of essays that develops the welfare economics of nonlinear ecosystems (see, e.g., the article by Brock & Starrett, 2003).
10. The example is taken from Chichilnisky (1994) and Dasgupta (1990).
In a study of agricultural production in Ghana, Lopez (1997) found that under a common property regime (Section 6), biomass was exploited beyond the level at which aggregate profits would be maximized, implying that reciprocal externalities were at work. The author noted that the underlying externalities would be exacerbated if trade expanded. For further weaknesses of trade expansion under the contemporary world order, see Stiglitz (2002).
Trade does not invariably exacerbate negative externalities. An expansion in international trade in activities with negative externalities for which a country has a comparative *disadvantage* will reduce the externalities. I am grateful to Dani Rodrik for this observation.
11. Colchester (1995) has recounted that political representatives of forest-dwellers in Sarawak, Malaysia, have routinely given logging licenses to members of the state legislature.
12. For a more comprehensive study of ecosystem services produced by avians, see Sekercioglu, Daily, and Ehrlich (2004).
13. See Dasgupta and Sabourian (2008).
14. Dasgupta's work (2003: Appendix) contains a formal model that identifies conditions for each possibility.
15. Pezzey (1992) and Pezzey and Toman (2002) discuss various ways of defining sustainable development. Here I am following the literature that has related the concept to welfare economics.
16. See Aronsson, Johansson, and Lofgren (1997), Dasgupta and Heal (1979), Hartwick (1990, 2000), Lutz (1993), Mäler (1974), Mäler and Wyzga (1976), and Nordhaus and Tobin (1972), among others. Mäler (1991) and Weale (1997) provide outlines of a complete system of national accounts inclusive of environmental natural resources. Lange, Hassan, and Hamilton (2003) and Perrings and Vincent's work (2003) contain applied studies on both valuation and resource accounts.
17. Determining the "willingness to pay" for changes in risk involves additional problems. See Smith and Desvouges (1987).
18. See Barbier (2000), Dasgupta (1982), Dobson, Bradshaw, and Baker (1997), Tilman, Polasky, and Lehman (2005), and Turner et al. (2000) for illustrations of ecosystem production functions and the corresponding dynamics of the socioecological systems.
19. During the moratorium the whale population grows at the fastest possible rate. In Spence's numerical computations, the commercially most-profitable duration of the moratorium was found to be some 10–15 years.
20. This finding has been proved with increasing generality by Arrow, Dasgupta, and Mäler (2003a, 2003b), Dasgupta (2001a), Dasgupta and Mäler (2000), and Hamilton and Clemens (1999). For a more complete account of the theory, see Dasgupta (2008b). But the idea that movements in wealth should be the basis on which "sustainable development" is discussed had been aired informally for several years earlier. See Pearce, Hamilton, et al. (1996) and Serageldin (1995).
21. It may seem odd to regard the first term in Eq. (4) as investment, since no one is doing anything other than wait to see the corresponding asset grow. As waiting is a cost, it seems to me there is nothing

wrong in including $[\partial V/\partial t]\Delta t$ in our conception of comprehensive investment.

There is no settled term yet for the linear index we are calling “comprehensive investment” here. I am borrowing the term from Arrow, Dasgupta, et al. (2008), but it has been called “genuine saving” (World Bank, 2006), and also “inclusive investment” (Dasgupta, 2007a). I am hoping that the term “comprehensive investment” will prevail, because it is vivid.

22. See Weitzman (1976) and an enormous literature that has followed that publication (for references, see Dasgupta, 2008b).
23. Proposition 3 would in this case be saying that it is not feasible to maintain comprehensive wealth and at the same time enjoy positive aggregate consumption.
24. That both are theoretically possible is easy to demonstrate, so I leave it to the reader. Arrow et al. (2008) and Dasgupta (2001a, 2007a) illustrated those possibilities using crude data from contemporary national accounts. I report a version of their findings below.
25. In Dasgupta (2001b) I have more fully developed the role comprehensive wealth plays as a unified criterion in both sustainability analysis and policy evaluation.
26. See Dasgupta (1972) for an account of the similarities and differences between the two methods.
27. Little and Mirrlees (1991) have speculated why.
28. See also Schelling (1999). The Nordhaus-Schelling prescription is widely misconstrued. How can they be so callous, it is asked, as to recommend a policy that could lead to a submersion of the (currently) densely populated coastal region of Bangladesh? In fact their thesis recommends Bangladesh to grow in wealth so as to enable coastal inhabitants to migrate and find employment elsewhere in the country. This may not be a realistic nor aesthetically appealing prescription, but it is not ethically repugnant.
29. There are two other assumptions underlying expression (15) that are surely artifacts: the constant hazard rate (δ) for humanity’s extinction and an infinite horizon. One way to ensure that the ethical framework we invoke does not have contradictions no matter how large the uncertainty would be to abandon the infinite time horizon. But the choice of a terminal date would at best be arbitrary. That is why economists have avoided working with finite time horizon models.
Another possible way out would be to continue to postulate an infinite time horizon, but formalize humanity’s extinction process in terms of a hazard rate that increases in an unbounded fashion over time at a sufficiently high rate. The problem is that we have little intuition on how to formulate that in a way that is scientifically reasonable.
30. But see Arrow et al. (2003b) for the corresponding analysis when demographic changes are endogenous to the model.
31. The sustainability criterion can be extended to cover a period of time by integrating the left-hand side of expression (A.2). For details, see Dasgupta (2008b).
32. We qualify only because in continuous time there is no “previous” date.
33. Dasgupta (2001a: Appendix) offers the justification.
34. In the combined problem of optimum saving and population, expression (A.11) does require of us to specify the value of c at which $U(c) = 0$. On this, see Dasgupta (1969) and Meade (1955).

References

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1A). Amsterdam: North-Holland.
- Acharya, G. (2000). Approaches to valuing the hidden hydrological services of wetland ecosystems. *Ecological Economics*, 35(1), 63–74.

- Acharya, G., & Barbier, E. B. (2000). Valuing ground water recharge through agricultural production in the Hadejia'Jama' are wetlands in northern Nigeria. *Agricultural Economics*, 22, 247–259.
- Agarwal, B. (2001). Participatory exclusions, community forestry, and gender: An analysis for south Asia and a conceptual framework. *World Development*, 29(10), 1623–1648.
- Agarwal, A., & Narain, S. (1989). *Towards green villages: A strategy for environmentally sound and participatory rural development*. New Delhi: Centre for Science and Development.
- Agarwal, A., & Narain, S. (1996). *Dying wisdom: Rise, fall and potential of India's traditional water harvesting systems*. New Delhi: Centre for Science and Development.
- Aggarwal, R., Netanyahu, S., & Romano, C. (2001). Access to natural resources and the fertility decision of women: The case of South Africa. *Environment and Development Economics*, 6(2), 209–236.
- Aghion, P., & Durlauf, S. (Eds.). (2005). *Handbook of growth economics* (Vols. 1A and 1B). Amsterdam: North-Holland.
- Aidt, T. (2009). Corruption, institutions, and development. *Oxford Review of Economic Policy*, 25, 271–291.
- Anderson, D. (1987). *The economics of afforestation: A case study in Africa*. Baltimore: The Johns Hopkins University Press.
- Aronsson, T., Johansson, P.-O., & Lofgren, K.-G. (1997). *Welfare measurement, sustainability and green accounting: A growth theoretic approach*. Cheltenham: Edward Elgar.
- Arrow, K. J. (1965). *Aspects of the theory of risk-bearing*. Helsinki: Yrjö Jahnssonin Säätiö.
- Arrow, K. J. (1971). Political and economic evaluation of social effects of externalities. In M. Intriligator (Ed.), *Frontiers of quantitative economics* (Vol. 1). Amsterdam: North-Holland.
- Arrow, K. J., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P. R., Heal, G. M., et al. (2004). Are we consuming too much? *Journal of Economic Perspectives*, 18(1), 147–172.
- Arrow, K. J., Dasgupta, P., Goulder, L. H., Mumford, K., & Oleson, K. (2008). *China, the U.S., and sustainability: perspectives based on comprehensive wealth*. Discussion Paper, Department of Economics, Stanford University.
- Arrow, K. J., Dasgupta, P., & Mäler, K.-G. (2003a). Evaluating projects and assessing sustainable development in imperfect economies. *Environmental & Resource Economics*, 26(4), 647–685.
- Arrow, K. J., Dasgupta, P., & Mäler, K.-G. (2003b). The genuine savings criterion and the value of population. *Economic Theory*, 21(2), 217–225.
- Arrow, K. J., & Kurz, M. (1970). *Public investment, the rate of return and optimal fiscal policy*. Baltimore: Johns Hopkins University Press.
- Atkinson, A. B., & Stiglitz, J. E. (1980). *Lectures in public economics*. New York: McGraw-Hill.
- Baland, J.-M., Bardhan, P., & Bowles, S. (Eds.). (2007). *Inequality, cooperation, and environmental sustainability*. New York: Russell Sage Foundation.
- Baland, J.-M., & Platteau, J.-P. (1996). *Halting degradation of natural resources: Is there a role for rural communities?*. Oxford: Clarendon Press.
- Baland, J.-M., & Platteau, J.-P. (1999). Inequality and collective action in the commons. *World Development*, 27(5), 773–788.
- Balasubramanian, R. (2008). Community tanks vs private wells: Coping strategies and sustainability issues in south India. In R. Ghate, N. S. Jodha & P. Mukhopadhyay (Eds.), *Promise, trust and evolution: Managing the commons of south Asia*. Oxford: Oxford University Press.
- Bandopadhyay, S., & Shyamsundar, P. (2004). *Fuelwood consumption and participation in community forestry in India*. Policy Research Working Paper 3331. Washington, DC: World Bank.
- Bandopadhyay, S., Shyamsundar, P., & Kanel, K. R. (2006). *Forestry user groups in Nepal: Can institutional change lead to economic development?* Background Paper for the Nepal Poverty Assessment. Washington, DC: World Bank.
- Banerjee, A. V., & Duflo, E. (2005). Growth theory through the lens of development economics. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1A, pp. 474–552). Amsterdam: Elsevier.
- Barbier, E. B. (1994). Valuing environmental functions: Tropical wetlands. *Land Economics*, 70(2), 155–173.
- Barbier, E. B. (2000). Valuing the environment as an input: Review of mangrove-fishery linkages. *Ecological Economics*, 35(1), 47–61.
- Barbier, E. B. (2005). *Natural resources and economic development*. Cambridge: Cambridge University Press.

- Barbier, E. B., & Strand, I. (1998). Valuing mangrove-fishery linkages: A case study of Campeche, Mexico. *Environmental and Resource Economics*, 12(1), 51–66.
- Bardhan, P., & Dayton-Johnson, J. (2007). Inequality and governance of water resources in Mexico and south India. In J.-M. Baland, P. Bardhan & S. Bowles (Eds.), *Inequality, cooperation, and environmental sustainability*. New York: Russell Sage Foundation.
- Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46(3), 878–894.
- Barrett, S. (2003). *Environment & statecraft: The strategy of environmental treaty-making*. New York: Oxford University Press.
- Barrett, S. (2008). *Dikes vs. windmills: Climate treatise and adaptation*. Discussion Paper. The Johns Hopkins University.
- Barro, R. J. (1997). *Determinants of economic growth: A cross-country empirical study*. Cambridge, MA: MIT Press.
- Barro, R. J., & Sala-i-Martin, X. (2003). *Economic growth*. Cambridge, MA: MIT Press.
- Baumol, W. M., & Oates, W. (1975). *The theory of environmental policy*. Englewood Cliffs, NJ: Prentice-Hall.
- Beck, T., & Nesmith, C. (2001). Building on poor people's capacities: The case of common property resources in India and West Africa. *World Development*, 29(1), 119–133.
- Beteille, A. (Ed.). (1983). *Equality and inequality: Theory and practice*. Delhi: Oxford University Press.
- Bhargava, A. (2008). *Food, economics, and health*. Oxford: Oxford University Press.
- Bigg, G. (2003). *The oceans and climate* (2nd ed.). Cambridge: Cambridge University Press.
- Bin, O., & Polasky, S. (2005). Evidence on the amenity value of wetlands in a rural setting. *Journal of Agricultural and Applied Economics*, 37(3), 589–602.
- Bledsoe, C. (1994). 'Children are like young bamboo trees': Potentiality and reproduction in sub-Saharan Africa. In K. Lindahl-Kiessling & H. Landberg (Eds.), *Population, economic development and the environment*. Oxford: Oxford University Press.
- Brauman, K. A., Daily, G. C., Ka'eo Duarte, T., & Mooney, H. A. (2007). The nature and value of ecosystem services: An overview highlighting hydrological services. *Annual Review of Environment and Resources*, 32, 67–98.
- Brock, W. A. (1977). A polluted golden age. In Smith, V. L. (Ed.), *Economics of natural and environmental resources*. New York: Gordon and Breach.
- Brock, W. A., & Starrett, D. A. (2003). Non-convexities in ecological management problems. In P. Dasgupta & K.-G. Mäler (Eds.), *The economics of non-convex ecosystems*. Dordrecht: Kluwer Academic Publishers.
- Bromley, D. W. (Ed.). (1992). *Making the commons work: Theory, practice and policy*. San Francisco: ICS Press.
- Broome, J. (1992). *Counting the cost of global warming*. London: White Horse Press.
- Brown, G., & McGuire, C. B. (1967). A socially optimal pricing policy for a public water agency. *Water Resources Research*, 3(1), 33–43.
- Bulte, E. H., Lipper, L., Stringer, R., & Zilberman, D. (2008). Payments for ecosystem services and poverty reduction: Concepts, issues, and empirical perspectives. *Environment and Development Economics*, 13(3), 245–254.
- Campbell, B., Manando, A., Nemarundwe, N., Sithole, B., De Jong, W., Luckert, M., & Matose, F. (2001). Challenges to proponents of common property resource systems: Despairing voices from the social forests of Zimbabwe. *World Development*, 29(4), 589–600.
- Carpenter, S. R. (2001). Alternate states of ecosystems: Evidence and its implications. In M. C. Press, N. Huntly & S. Levin (Eds.), *Ecology: Achievement and challenge*. London: Basil Blackwell.
- Carpenter, S. R., Ludwig, D., & Brock, W. A. (1999). Management of eutrophication for lakes subject to potentially irreversible change. *Ecological Applications*, 9(4), 751–771.
- Carpenter, S. R., Pingali, P. L., Bennet, E. M., & Zurek, M. B. (Eds.). (2005). *Ecosystems and human well-being, Vol. 2: Scenarios*. Washington, DC: Island Press.
- Carraro, C. (2002). Climate change policy: Models, controversies, and strategies. In T. Tietenberg & H. Folmer (Eds.), *The international yearbook of environmental and resource economics 2002/2003*. Cheltenham: Edward Elgar.

- Carraro, C., & Fragnelli, V. (Eds.). (2004). *Game practice and the environment*. Cheltenham: Edward Elgar.
- Carson, R. T. (2004). *Contingent valuation: A comprehensive bibliography and history*. Northampton, MA: Edward Elgar.
- Cavendish, W. (2000). Empirical regularities in the poverty-environment relationships of rural households: Evidence from Zimbabwe. *World Development*, 28(11), 1979–2003.
- Chichilnisky, G. (1994). North-south trade and the global environment. *American Economic Review*, 84(3), 851–874.
- Chichilnisky, G., & Heal, G. M. (1998). Economic returns from the biosphere. *Nature*, 391, 629–630.
- Chopra, K., & Gulati, S. C. (1998). Environmental degradation, property rights and population movements: Hypotheses and evidence from Rajasthan (India). *Environment and Development Economics*, 3(1), 35–57.
- Chopra, K., Kadekodi, G. K., & Murty, M. N. (1989). *Participatory development: people and common property resources*. New Delhi: Sage Publications.
- Cline, W. R. (1992). *The economics of global warming*. Washington, DC: Institute for International Economics.
- Coase, R. (1960). The problem of social cost. *Journal of Law and Economics*, 3(1), 1–44.
- Colchester, M. (1995). Sustaining the forests: The community-based approach in south and south-east Asia. *Development and Change*, 25(1), 69–100.
- Collier, P. (2007). *The bottom billion*. Oxford: Oxford University Press.
- Cooke, P., Kohlin, G., & Hyde, W. F. (2008). Fuelwood, forests, and community management—Evidence from household studies. *Environment and Development Economics*, 13(1), 103–135.
- CSE—Centre for Science and Environment. (1990). *Human-nature interactions in a central himalayan village: A case study of village Bembri*. New Delhi: Centre for Science and Environment.
- Czako, Å., & Sik, E. (1988). Manager's reciprocal transactions. *Connections*, 11(1), 23–32.
- Daily, G. (Ed.). (1997). *Nature's services: Societal dependence on natural ecosystems*. Washington, DC: Island Press.
- Daily, G., & Ellison, K. (2002). *The new economy of nature: The quest to make conservation profitable*. Washington, DC: Island Press.
- Dasgupta, P. (1969). On the concept of optimum population. *Review of Economic Studies*, 36(3), 295–318.
- Dasgupta, P. (1972). A comparative analysis of the UNIDO guidelines and the OECD manual. *Bulletin of the Oxford University Institute of Economics and Statistics*, 34(1), 33–51.
- Dasgupta, P. (1982). *The control of resources*. Cambridge, MA: Harvard University Press.
- Dasgupta, P. (1990). The environment as a commodity. *Oxford Review of Economic Policy*, 6(1), 51–67.
- Dasgupta, P. (1993). *An inquiry into well-being and destitution*. Oxford: Clarendon Press.
- Dasgupta, P. (2000). Economic progress and the idea of social capital. In P. Dasgupta & I. Serageldin (Eds.), *Social capital: A multifaceted perspective*. Washington, DC: World Bank, pp. 325–424.
- Dasgupta, P. (2001a). *Human well-being and the natural environment* (2nd ed.). Oxford: Oxford University Press. (2004).
- Dasgupta, P. (2001b). Valuing objects and evaluating policies in imperfect economies. *Economic Journal*, 111(Conference Issue), 1–29.
- Dasgupta, P. (2003). Population, poverty, and the natural environment. In K.-G. Mäler & J. Vincent (Eds.), *Handbook of environmental and resource economics* (Vol. 1, pp. 191–247). Amsterdam: North-Holland.
- Dasgupta, P. (2004). World poverty: Causes and pathways. In: F. Bourguignon & B. Pleskovic (Eds.), *Annual world bank conference on development economics 2003: Accelerating development* (pp. 159–196), New York: World Bank and Oxford University Press.
- Dasgupta, P. (2007a). *Economics: A very short introduction*. Oxford: Oxford University Press.
- Dasgupta, P. (2007b). Comments on the stern review's economics of climate change. *National Institute Economic Review*, 199, 4–7.
- Dasgupta, P. (2008a). Common property resources: Economic analytics. In R. Ghate, Jodha, N. S. & P. Mukhopadhyay (Eds.), *Promise, trust and evolution: Managing the commons of south Asia*. Oxford: Oxford University Press.
- Dasgupta, P. (2008b). Discounting climate change. *Journal of Risk and Uncertainty*, 37, 141–169.

- Dasgupta, P. (2009). The welfare economic theory of green national accounts. *Environmental and Resource Economics*, 42, 3–38.
- Dasgupta, P., & Heal, G. (1979). *Economic theory and exhaustible resources*. Cambridge: Cambridge University Press.
- Dasgupta, P., & Mäler, K.-G. (1991). The environment and emerging development issues. In *Proceedings of the annual World Bank conference on development economics 1990* (pp. 101–132), Washington, DC: World Bank.
- Dasgupta, P., & Mäler, K.-G. (2000). Net national product, wealth, and social well-being. *Environment and Development Economics*, 5(1), 69–93.
- Dasgupta, P., & Mäler, K.-G. (Eds.). (2003). *The economics of non-convex ecosystems*. Dordrecht: Kluwer Academic Publishers.
- Dasgupta, P., Marglin, S., & Sen, A. (1972). *Guidelines for project evaluation*. New York: United Nations.
- Dasgupta, P., & Sabourian, H. (2008). *Bilateral negotiations over undirectional externalities*. Mimeo. Faculty of Economics, University of Cambridge.
- Diamond, J. (2005). *Collapse: How societies choose to fail or survive*. London: Allen Lane.
- Dinar, A., Hassan, R., Mendelsohn, R., Benhin, J., et al. (2008). *Climate change and agriculture in Africa: Impact assessment and adaptation strategies*. London: Earthscan.
- Dobson, A. P., Bradshaw, A. D., & Baker, A. J. M. (1997). Hope for the future: Restoration ecology and conservation biology. *Science*, 277, 515–522.
- Dreze, J., & Sen, A. (1990). *Hunger and public action*. Oxford: Clarendon Press.
- Dreze, J., & Sen, A. (1995). *India: Economic development and social opportunities*. Oxford: Oxford University Press.
- Duflo, E., & Kremer, M. (2005). Use of randomization in the evaluation of development effectiveness. In G. Pitman, O. Feinstein & G. Ingram (Eds.), *Development effectiveness*. New Brunswick, NJ: Transaction Publisher.
- Duflo, E., & Pandey, R. (2007). Dams. *Quarterly Journal of Economics*, 122(2), 601–646.
- Duraiappah, A. K. (2003). *Computational models in the economics of environment and development*. Dordrecht: Kluwer.
- Dutta, P. K., & Radner, R. (2004). Self-enforcing climate change treatise. *Proceedings of the National Academy of Sciences*, 101(14), 5174–5179.
- Eckstein, O. (1958). *Water resources development: The economics of project evaluation*. Cambridge, MA: Harvard University Press.
- Ehrlich, P. R. (2008). Key issues for attention from ecological economists. *Environment and Development Economics*, 13(1), 1–20.
- Ehrlich, P. R., & Ehrlich, A. H. (1981). *Extinction: The causes and consequences of the disappearance of species*. New York, NY: Random House.
- Ehrlich, P. R., & Ehrlich, A. H. (1990). *The population explosion*. New York: Simon and Schuster.
- Ehrlich, P. R., & Ehrlich, A. H. (2008). *The dominant animal: Human evolution and the environment*. Washington, DC: Island Press.
- Ehrlich, P. R., & Goulder, L. H. (2007). Is current consumption excessive? A general framework and some indications for the United States. *Conservation Biology*, 21(5), 1145–1154.
- Ehrlich, P. R., Ehrlich, A. H., & Holdren, J. P. (1977). *EcoScience: Population, resources, environment*. San Francisco, CA: W.H. Freeman.
- Englin, J., & Mendelsohn, R. (1991). A hedonic travel cost analysis for valuation of multiple components of site quality: The recreation value of forest management. *Journal of Environmental Economics and Management*, 21(2), 275–290.
- Ensminger, J. (1992). *Making a market: The institutional transformation of an African society*. New York: Cambridge University Press.
- Falconer, J. (1990). *The major significance of minor forest products: The local use and value of forests in the West African humid forest zone*. Rome: Food and Agriculture Organization.
- Fankhauser, S. (1995). *Valuing climate change: The economics of greenhouse*. London: Earthscan.
- Feeny, D., Berkes, F., McCay, B. J., & Acheson, J. M. (1990). The tragedy of the commons: Twenty-two years later. *Human Ecology*, 18(1), 1–19.

- Filmer, D., & Pritchett, L. (2002). Environmental degradation and the demand for children: Searching for the vicious circle in pakistan. *Environment and Development Economics*, 7(1), 123–146.
- Finus, M., & Rundshagen, B. (2003). Endogenous coalition formation in global pollution control: A partition function approach. In C. Carraro (Ed.), *The endogenous formation of economic coalitions*. Cheltenham: Edward Elgar.
- Freeman, A. M., III. (1993). *The measurement of environmental and resource values: Theory and methods*. Washington, DC: Resources for the Future.
- Ghate, R., Jodha, N. S., & Mukhopadhyay, P. (Eds.). (2008). *Promise, trust and evolution: Managing the commons of south Asia*. Oxford: Oxford University Press.
- Goldstein, J., Daily, G. C., Friday, J. B., Matson, P. A., Naylor, R. L., & Vitousek, P. M. (2006). Business strategies for conservation on private lands: Koa forestry as a case study. *Proceedings of the National Academy of Sciences*, 103(26), 10140–10145.
- Gren, I. M., Folke, C., Turner, R. K., & Bateman, I. (1994). Primary and secondary values of wetland ecosystems. *Environmental and Resources Economics*, 4(1), 55–74.
- Guyer, J. L. (1994). Lineal identities and lateral networks: The logic of polyandrous motherhood. In C. Bledsoe & G. Pison (Eds.), *Nuptiality in sub-Saharan Africa: Contemporary anthropological and demographic perspectives*. Oxford: Clarendon Press.
- Hamilton, K., & Clemens, M. (1999). Genuine savings rates in developing countries. *World Bank Economic Review*, 13(2), 333–356.
- Hartwick, J. M. (1990). Natural resources, national accounting, and economic depreciation. *Journal of Public Economics*, 43(3), 291–304.
- Hartwick, J. (2000). *National accounting and capital*. Cheltenham: Edward Elgar.
- Hassan, R. (Ed.). (2002). *Accounting for stock and flow values of woody land resources*. Pretoria: University of Pretoria.
- Hassan, R., Scholes, R., & Ash, N. (Eds.). (2005). *Ecosystems and human well-being, Vol. 1: State and trends*. Washington, DC: Island Press.
- Hecht, S., Anderson, A. B., & May, P. (1988). The subsidy from nature: Shifting cultivation, successional palm forests and rural development. *Human Organization*, 47(1), 25–35.
- Heckman, J. J. (1992). Randomization and social policy evaluation. In C. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs*. Cambridge, MA: Harvard University Press.
- Helpman, E. (2004). *The mystery of economic growth*. Cambridge, MA: Belknap.
- Heyser, N. (1996). *Gender, population and environment in the context of deforestation: A Malaysian case study*. Geneva: United Nations Research Institute for Social Development.
- Homer-Dixon, T. E. (1999). *Environment, scarcity, and violence*. Princeton, NJ: Princeton University Press.
- Hufschmidt, M. M., James, D. E., Meister, A. D., Bower, B. T., & Dixon, J. A. (1983). *Environment, natural systems, and development*. Baltimore, MD: The Johns Hopkins University Press.
- Isham, J., Narayan, D., & Pritchett, L. H. (1995). Does participation improve performance? Establishing causality with subjective data. *World Bank Economic Review*, 9(2), 175–200.
- Jack, B. K., Kousky, C., & Sims, K. R. E. (2008). Designing payment for ecosystem services: Lessons from previous experience with incentive-based mechanisms. *Proceedings of the National Academy of Sciences*, 105(28), 9465–9470.
- Jalan, J., & Somanathan, E. (2008). The importance of being informed: Experimental evidence on demand for environmental quality. *Journal of Development Economics*, 87(1), 14–28.
- Jodha, N. S. (1986). Common property resources and the rural poor. *Economic and Political Weekly*, 21, 1169–1181.
- Jodha, N. S. (2001). *Living on the edge: Sustaining agriculture and community resources in fragile environments*. Delhi: Oxford University Press.
- Kadekodi, G. K. (2004). *Common property resource management: Reflections on theory and the Indian experience*. Delhi: Oxford University Press.
- Kim, E. H., Morse, A., & Zingales, L. (2006). What has mattered to economics since 1970. *Journal of Economic Perspectives*, 20(4), 189–202.
- Kneese, A. V., & Schultze, C. L. (1975). *Pollution, prices, and public policy*. Washington, DC: Brookings Institution.

- Kolstad, C. (1999). *Environmental economics*. New York: Oxford University Press.
- Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica*, 28(2), 287–309.
- Koopmans, T. C. (1965). *On the concept of optimal economic growth. Pontificiae Academiae Scientiarum Scripta Varia*, 28. Amsterdam: North-Holland (Reprinted in The Econometric Approach to Development Planning, 1966).
- Koopmans, T. C. (1967). Objectives, constraints and outcomes in optimal growth models. *Econometrica*, 35(1), 1–15.
- Koopmans, T. C. (1972). Representation of preference orderings over time. In C. B. McGuire & Roy Radner (Eds.), *Decision and organization*. Amsterdam: North-Holland.
- Kremer, M., & Miguel, E. (2006). *The illusion of sustainability*. Working Paper No. 091. Cambridge, MA: Bureau for Research and Economic Analysis of Development, Center for International Development, Harvard University.
- Lange, G.-M., Hassan, R., & Hamilton, K. (2003). *Environmental accounting in action: Case studies from south Africa*. Cheltenham: Edward Elgar.
- Leisinger, K. M., Schmidt, K. M., & Pandya-Loch, R. (2002). *Six billion and counting: Population and food security in the 21st century*. Washington, DC: International Food Policy Research Institute.
- Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., et al. (2008). Tipping elements in the earth's climate system. *Proceedings of the National Academy of Sciences*, 105(6), 1786–1793.
- Levhari, D., & Srinivasan, T. N. (1969). Optimal savings under uncertainty. *Review of Economic Studies*, 36(2), 153–163.
- Lindahl, E. R. (1933). The concept of income. In G. Bagge (Ed.), *Economic essays in honour of gustav cassel*. London: Allen & Unwin.
- Lindahl, E. R. (1958). Just taxation—A positive solution. In R. A. Musgrave & A. T. Peacock (Eds.), *Classics in the theory of public finance*. London: MacMillan. Translated from the original article in Swedish, published in 1919.
- Little, I. M. D., & Mirrlees, J. A. (1968). *Manual of industrial project analysis in developing countries: Social cost benefit analysis*. Paris: OECD.
- Little, I. M. D., & Mirrlees, J. A. (1974). *Project appraisal and planning for developing countries*. London: Heinemann.
- Little, I. M. D., & Mirrlees, J. A. (1991). Project appraisal and planning twenty years on. *Proceedings of the annual world bank conference on development economics 1990* (pp. 351–382), Washington, DC: World Bank.
- Lopez, R. (1997). Environmental externalities in traditional agriculture and the impact of trade liberalization: The case of Ghana. *Journal of Development Economics*, 53(1), 17–39.
- Lopez, R. (1998). The tragedy of the commons in Cote d'Ivoire agriculture: Empirical evidence and implications for evaluating trade policies. *World Bank Economic Review*, 12(1), 105–132.
- Lopez-Feldman, A., & Wilen, J. E. (2008). Poverty and spatial dimensions of non-timber forest extraction. *Environment and Development Economics*, 13(5), 621–642.
- Loughran, D., & Pritchett, L. (1998). *Environmental scarcity, resource collection, and the demand for children in Nepal*. Discussion Paper. Washington, DC: World Bank.
- Lutz, E. (Ed.). (1993). *Toward improved accounting for the environment*. Washington, DC: World Bank.
- Mäler, K.-G. (1974). *Environmental economics: A theoretical enquiry*. Baltimore, MD: Johns Hopkins University Press.
- Mäler, K.-G. (1991). National accounting and environmental resources. *Environmental and Resource Economics*, 1(1), 1–15.
- Mäler, K.-G., & Wyzga, R. E. (1976). *Economic measurement of environmental Damage*. Paris: OECD.
- Marglin, S. A. (1963). The opportunity costs of public investment. *Quarterly Journal of Economics*, 77(2), 274–289.
- Mauro, P. (1995). Corruption and economic growth. *Quarterly Journal of Economics*, 110(3), 681–712.
- McCulloch, N., Winters, A. A., & Cirera, X. (2001). *Trade liberalization and poverty: A handbook*. London: Department for International Development.
- McKean, M. (1992). Success on the commons: A comparative examination of institutions for common property resource management. *Journal of Theoretical Politics*, 4(2), 256–268.

- Meade, J. E. (1952). External economies and diseconomies in a competitive situation. *Economic Journal*, 62(1), 54–67.
- Meade, J. E. (1955). *Trade and welfare*. Oxford: Oxford University Press.
- Meade, J. E. (1973). *The theory of externalities*. Geneva: Institute Universitaire de Hautes Etudes Internationales.
- MEA—Millennium Ecosystem Assessment. (2003). *Ecosystems and human well-being*. Washington, DC: Island Press.
- Mendelsohn, R., Dinar, A., & Williams, L. (2006). The distributional impact of climate change in rich and poor countries. *Environment and Development Economics*, 11(2), 159–178.
- Mirrlees, J. A. (1967). Optimum growth when technology is changing. *Review of Economic Studies*, 34(1), 95–124.
- Mirrlees, J. A. (1969). The evaluation of national income in an imperfect economy. *Pakistan Development Review*, 9(1), 1–13.
- Mukhopadhyay, P. (2008). Heterogeneity, commons, and privatization: Agrarian institutional change in Goa. In R. Ghate, Jodha, N. S. & P. Mukhopadhyay (Eds.), *Promise, trust and evolution: Managing the commons of south Asia*. Oxford: Oxford University Press.
- Myers, N., & Kent, J. (2000). *Perverse subsidies: How tax dollars undercut our environment and our economies*. Washington, DC: Island Press.
- National Research Council. (1986). *Proceedings of a conference on common property resource management*, Washington, DC: National Academy Press.
- National Research Council. (2002). *The drama of the commons*. Washington, DC: National Academy Press.
- Nerlove, M. (1991). Population and the environment: A parable of firewood and other tales. *American Journal of Agricultural Economics*, 75(1), 59–71.
- Netting, R. (1985). *Hill farmers of Nigeria: Cultural ecology of the Kofyar of the Jos Plateau*. Seattle: University of Washington Press.
- Newcombe, K. J. (1989). An economic justification of rural afforestation: The case of ethiopia. In G. Schramm & Warford, J. J. (Eds.), *Environmental management and economic development*. Baltimore: The Johns Hopkins University Press.
- Nordhaus, W. D. (1994). *Managing the global commons: The economics of climate change*. Cambridge, MA: MIT Press.
- Nordhaus, W. D. (2007). The stern review on the economics of climate change. *Journal of Economic Literature*, 45(3), 686–702.
- Nordhaus, W. D. (2008). *A question of balance: Weighing the options on global warming policies*. New Haven, CT: Yale University Press.
- Nordhaus, W. D., & Tobin, J. (1972). Is economic growth obsolete? In *Economic growth, 5th Anniversary colloquium of the NBER*. New York: Columbia University Press.
- Norronha, R. (1997). Common property resource management in traditional societies. In P. Dasgupta & K.-G. Mäler (Eds.), *The environment and emerging development issues*. Oxford: Clarendon Press.
- North, D., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge: Cambridge University Press.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.
- Ostrom, E. (1992). *Crafting institutions for self-governing irrigation systems*. San Francisco, CA: International Center for Self-Governance Press.
- Ostrom, E. (1996). Incentives, rules of the game, and development. In *Proceedings of the Annual World Bank Conference on Development Economics, 1995* (pp. 207–234). Supplement to the World Bank Economic Review and the World Bank Research Observer.
- Pagiola, S., Landell-Mills, N., & Bishop, J. (2002). Making market-based mechanisms work for forests and people. In S. Pagiola, J. Bishop & N. Landell-Mills (Eds.), *Selling forest environmental services: Market-based mechanisms for conservation and development*. London: Earthscan.
- Pagiola, S., Rios, A. R., & Arcenas, A. (2008). Can the poor participate in payments for environmental services? Lessons from the silvopastoral project in Nicaragua. *Environment and Development Economics*, 13(3), 299–326.

- Palm, C. A., Vosti, S. A., Sanchez, P. A., & Erickson, P. J. (Eds.), (2005). *Slash-and-burn agriculture: The search for alternatives*. New York: Columbia University Press.
- Pattanayak, S. K. (2004). Valuing watershed services: Concepts and empirics from Southeast Asia. *Agriculture, Ecosystems & Environment*, 104(1), 171–184.
- Pattanayak, S. K., & Butry, D. T. (2005). Spatial complementarity of forests and farms: Accounting for ecosystem services. *American Journal of Agricultural Economics*, 87(4), 995–1008.
- Pattanayak, S. K., & Kramer, R. A. (2001). Worth of watersheds: A producer surplus approach for valuing drought mitigation in eastern Indonesia. *Environment and Development Economics*, 6(1), 123–146.
- Pattanayak, S. K., & Sills, E. O. (2001). Do tropical forests provide natural insurance? The microeconomics of non-timber forest collection in the Brazilian Amazon. *Land Economics*, 77(4), 595–612.
- Pearce, D., Cline, W., Achanta, A., Fankhauser, S., Pachauri, R., Tol, R., et al. (1996). The social cost of climate change: Greenhouse damage and the benefits of control. In J. Bruce, H. Lee & E. Haites (Eds.), *Climate change 1995: Economic and social dimensions of climate change*. Cambridge: Cambridge University Press.
- Pearce, D., Hamilton, K., & Atkinson, G. (1996). Measuring sustainable development: Progress on indicators. *Environment and Development Economics*, 1(1), 85–101.
- Perrings, C. (Ed.). (2000). *The economics of biodiversity conservation in sub-Saharan Africa*. Cheltenham: Edward Elgar.
- Perrings, C., & Vincent, J. R. (Eds.), (2003). *Natural resource accounting and economic development: Theory and practice*. Cheltenham: Edward Elgar.
- Pesaran, H., Pettenuzzo, D., & Timmermann, A. (2007). Learning, structural instability, and present value calculations. *Econometric Reviews*, 6(2–4), 253–288.
- Pezzey, J. C. V. (1992). *Sustainable development concepts: An economic analysis*. World Bank Environment Paper No. 2. Washington, DC: World Bank.
- Pezzey, J. C. V., & Toman, M. A. (2002). Progress and problems in the economics of sustainability. In T. Tietenberg & H. Folmer (Eds.), *The international yearbook of environmental and resource economics 2002/2003*. Cheltenham, UK: Edward Elgar.
- Pigou, A. C. (1920). *The economics of welfare*. London: Macmillan.
- Ramsey, F. P. (1928). A mathematical theory of saving. *Economic Journal*, 38(4), 543–549.
- Rawls, J. (1972). *A theory of justice*. Oxford: Oxford University Press.
- Ray, D. (1998). *Development economics*. Princeton, NJ: Princeton University Press.
- Reid, W. V., Mooney, H. A., Cropper, A., et al. (2005). *Ecosystems and human-wellbeing: Synthesis*. Washington, DC: World Resources Institute.
- Repetto, R. (1988). *Economic policy reform for natural resource conservation*. Environment Department Working Paper No. 4. Washington, DC: World Bank.
- Repetto, R., Dower, R. C., Jenkins, R., & Geoghegan, J. (1992). *Green fees: How a tax shift can work for the environment and the economy*. Washington, DC: World Resources Institute.
- Repetto, R., Magrath, W., Wells, M., Beer, C., & Rossini, F. (1989). *Wasting assets: Natural resources and the national income accounts*. Washington, DC: World Resources Institute.
- Ricketts, T. H., Daily, G. C., Ehrlich, P. R., & Michener, C. (2004). Economic value of tropical forests in coffee production. *Proceedings of the National Academy of Sciences*, 101, 12579–12582.
- Ricketts, T. H., Regetz, J., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., Bogdanski, A., et al. (2008). Landscape effects on crop pollination services: Are there general patterns? *Ecology Letters*, 11, 499–515.
- Rosenzweig, C., & Hillel, D. (1998). *Climate change and the global harvest: Potential impacts of the greenhouse effect on agriculture*. New York: Oxford University Press.
- Sachs, J. D. (2008). *Common wealth: Economics for a crowded planet*. New York: Penguin Press.
- Sachs, J. D., Gallup, J. L., & Mellinger, A. (1998). Geography and economic development. In B. Pleskovic & J. E. Stiglitz (Eds.), *Annual world bank conference on development economics*. Washington, DC: World Bank.
- Schelling, T. C. (1999). Intergenerational discounting. In P. R. Portney & J. P. Weyant (Eds.), *Discounting and intergenerational equity*. Washington, DC: Resources for the Future.

- Seabright, P. (1993). Managing local commons: Theoretical issues in incentive design. *Journal of Economic Perspectives*, 7(1), 13–34.
- Sekercioglu, C. H., Daily, C. G., & Ehrlich, P. R. (2004). Ecosystem consequences of bird declines. *Proceedings of the National Academy of Sciences*, 101(52), 18042–18047.
- Sen, A. (1976). Real national income. *Review of Economic Studies*, 43(1), 19–39.
- Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.
- Serageldin, I. (1995). Are we saving enough for the future? In *Monitoring environmental progress, report on work in progress*. Environmentally Sustainable Development, World Bank, Washington, DC.
- Shyamsundar, P. (2008). Decentralization, devolution, and collective action. In R. Ghate, N. S. Jodha, & P. Mukhopadhyay (Eds.), *Promise, trust and evolution: Managing the commons of south Asia*. Oxford: Oxford University Press.
- Siebert, H. (2008). *Economics of the environment: Theory and policy* (7th ed.). Berlin: Springer-Verlag.
- Smith, V. K. (1997). Pricing what is priceless: A status report on non-market valuation of environmental resources. In H. Folmer & T. Tietenberg (Eds.), *The international yearbook of environmental and resource economics 1997/1998*. Cheltenham: Edward Elgar.
- Smith, V. K. (2004). Fifty years of contingent valuation. In T. Tietenberg & H. Folmer (Eds.), *The international yearbook of environmental and resource economics 2004/2005*. Cheltenham: Edward Elgar.
- Smith, V. K., & Desvouges, W. H. (1987). An empirical analysis of the economic value of risk changes. *Journal of Political Economy*, 95(1), 89–114.
- Solorzano, R., de Camino, R., Woodward, R., Tosi, J., Watson, V., Vasquez, A., et al. (1991). *Accounts overdue: Natural resource depreciation in Costa Rica*. Washington, DC: World Resources Institute.
- Somanathan, E. (1991). Deforestation, property rights and incentives in central Himalaya. *Economic and Political Weekly*, 26(Special Issue: January 26), PE37–46.
- Somanathan, E., Prabhakar, R., & Mehta, B. S. (2005). *Does decentralization work? Forest conservation in the Himalayas*. Discussion Paper 05–4. Delhi: Indian Statistical Institute.
- Spence, A. M. (1974). Blue whales and optimal control theory. In H. Göttinger (Ed.), *Systems approaches and environmental problems*. Göttingen: Vandenhoeck and Ruprecht.
- Squire, L., & Van der Taak, H. (1975). *Economic analysis of projects*. Baltimore: Johns Hopkins University Press.
- Starrett, D. A. (1972). Fundamental non-convexities in the theory of externalities. *Journal of Economic Theory*, 4(1), 180–199.
- Steffen, W., Sanderson, A., Tyson, P. D., Jäger, J., Matson, P. A., Moore, B. III, et al. (2004). *Global change and the earth system*. Berlin: Springer.
- Stern, N. H. (1989). The economics of development: A survey. *Economic Journal*, 99(2), 597–685.
- Stern, N. H. (2006). *The stern review of the economics of climate change*. Cambridge: Cambridge University Press.
- Sterner, T. (2003). *Policy instruments for environmental and natural resource management*. Washington, DC: Resources for the Future.
- Stiglitz, J. E. (2002). *Globalization and its discontents*. New York: W.W. Norton.
- Tallis, H., Kareiva, P., Marvier, M., & Chang, A. (2008). An ecosystem services framework to support both practical conservation and economic development. *Proceedings of the National Academy of Sciences*, 105(28), 9457–9464.
- Thomson, J. T., Feeny, D. H., & Oakerson, R. J. (1986). Institutional dynamics: The evolution and dissolution of common property resource management. In *National Research Council, Proceedings of a Conference on Common Property Resource Management*. Washington, DC: US National Academy of Sciences.
- Tilman, D., Polasky, S., & Lehman, C. (2005). Diversity, productivity and temporal stability in the economics of humans and nature. *Journal of Environmental Economics and Management*, 49(3), 405–426.
- Tomich, T. P., Palm, C. A., Velarde, S. J., Geist, H., Gillison, A. N., Lebel, L., et al. (2005). *Forest and agroecosystem tradeoffs in the humid tropics*. Nairobi: Alternatives to Slash-and-Burn Programme.
- Tomich, T. P., van Noordwijk, M., & Thomas, D. E. (Eds.). (2004). Environmental services and land use change: Bridging the gap between policy and research in Southeast Asia. *Agriculture Ecosystems & Environment*, 104(1).

- Turner, R. K., & Daily, G. C. (2008). The ecosystem services framework for natural capital conservation. *Environment and Resource Economics*, 39(1), 25–35.
- Turner, R. K., van den Bergh, C. J. M., Soderqvist, T., Barendregt, A., Van der Straaten, J., Maltby, E., et al. (2000). Ecological-economics of wetlands: Scientific integration for management and policy. *Ecological Economics*, 35(1), 7–23.
- UNDP—United Nations Development Programme. (2006, 2007/2008). *Human development report*. New York: Oxford University Press.
- Uzawa, H. (2003). *Economic theory and global warming*. Cambridge: Cambridge University Press.
- Vincent, J. R., Ali, R. M., et al. (1997). *Environment and development in a resource-rich economy: Malaysia under the new economic policy*. Cambridge, MA: Harvard Institute for International Development.
- Viscusi, W. K. (2007). Rational discounting for regulatory analysis. *University of Chicago Law Review*, 74(1), 209–246.
- Viscusi, W. K., & Aldy, J. E. (2003). The value of a statistical life: A critical review of market estimates throughout the world. *The Journal of Risk and Uncertainty*, 27(1), 5–76.
- Vitousek, P. M., Ehrlich, P. R., Ehrlich, A. H., & Matson, P. (1986). Human appropriation of the product of photosynthesis. *BioScience*, 36, 368–373.
- Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of earth's ecosystem. *Science*, 277, 494–499.
- Wade, R. (1988). *Village republics: Economic conditions for collective action in south India*. Cambridge: Cambridge University Press.
- Weale, M. (1997). Environment and national accounts. In P. Dasgupta & K.-G. Mäler (Eds.), *The environment and emerging development issues* (Vol. 1). Oxford: Clarendon Press.
- Weitzman, M. L. (1976). On the welfare significance of national product in a dynamic economy. *Quarterly Journal of Economics*, 90(1), 156–162.
- Weitzman, M. L. (2007). *The role of uncertainty in the economics of catastrophic climate change*. Discussion Paper. Department of Economics, Harvard University. *Review of Economics and Statistics* (forthcoming).
- Whittington, D., Hanemann, W. M., Sadoff, C., & Jeuland, M. (2009). Water and sanitation challenge paper. In B. Lomborg (Ed.), *Global crises, global solutions*. (2nd ed.). Cambridge: Cambridge University Press.
- World Bank. (2006). *Where is the wealth of nations?* Washington, DC: World Bank.
- World Commission on Environment and Development. (1987). *Our common future*. New York: Oxford University Press.
- World Resources Institute. (2005). *World resources 2005: The wealth of the poor*. Washington, DC: World Resources Institute.
- Zilberman, D., Lipper, L., & McCarthy, N. (2008). When could payments for environmental services benefit the poor? *Environment and Development Economics*, 13(3), 255–278.

INDEX

A

- Adult mortality
 - and BMI, 4832
 - decline in, 4812–4813
 - and growth, 4816, 4866
- African Growth and Opportunity Act (AGOA), 4497
- Agglomeration, 4043–4044, 4067
- Argentina
 - decentralization and local community participation, 4940
 - poverty, 4687
 - recovery from economic meltdown, 4684
 - trade protection, 4063
- Authoritarianism, 4638, 4646

B

- Bad governance, 4598–4599, 4604–4605
- Bangladesh
 - family planning and reproductive health, social experiment, 4855–4857
 - financial access, gaps in, 4711
 - Grameen Bank, 4704, 4706–4707, 4712, 4728, 4749, 4769, 4772
 - Rural Advancement Committee, 4431
- Bank Rakyat Indonesia, 4717, 4722
- Behavioral economics, 4707
- Body mass index (BMI), 4829, 4831–4832, 4866
- Brain drain, 4375–4377

C

- Campaign for Child Survival, 4418
- Capital account liberalization, 4286, 4288
 - adverse effects, 4310
 - growth benefits, 4317
- Capital controls
 - implementation and measurement issues, 4337–4338
 - macroeconomic implications, 4338–4339
 - microeconomic implications, 4339–4341
- Capital flows and economic growth
 - capital controls, 4337–4341
 - developing countries, investment-constraint, 4336
 - empirical evidence, 4332–4334
 - interpretation of findings, 4334–4336
 - policy implications, 4337
- Capital markets and educational investments, 4902–4904
- Capital shortage and neoclassical growth models, 4435–4437
- Central bank independence (CBI), case study, 4631–4633
- Childhood nutrition and education, 4912–4913
- Children's human capital, parents investment, 4805–4810
 - complementarities between health and education inputs, 4809
 - disabilities among children, 4809–4810
 - health endowments at birth, 4806–4807
 - investment in twins, 4807–4808
 - wage returns to schooling, 4806
- Chilean voucher system, 4907–4908
- China
 - domestic content requirements, 4106
 - exports, 4071
 - FDI, 4104
 - financial globalization, 4285
 - housing reform, 4558
 - income inequality, 4683
 - joint venture activity, 4103
 - migration, 4394
 - national pension system, 4684
 - performance-linked wages, 4683
 - strengthening labor institutions, 4684
 - tariff protection, 4080
- Clientelism, 4612
- Climate change, discounting
 - climate treaties, 5025–5026
 - global activism, 5021–5022
 - inequality and risk aversions, 5024–5025
 - negative discount rates, 5022
 - welfare economics, 5022–5024
- Common property resources (CPRs)
 - bad features
 - communal rights, 4998–4999
 - communitarian institutions, 4996
 - fragile cooperation, 4999–5000
 - migration, 5000
 - rapid population growth, 4997–4998
 - resource degradation, 4997
 - good features
 - communal management, 4995–4996
 - cooperation maintenance, 4996
 - importance, 4994–4995
 - need for, 4992–4994
 - agricultural land, 4993–4994
 - jeopardizing cooperation, 4993
 - local ecosystems, 4993
- Communal property rights, 4550–4552
- Computer-assisted learning (CAL) program, 4936
- Conditional cash transfer (CCT) programs, 4905

- Conditionality**
- achievements and failures, 4471–4472
 - commitment problems, 4470–4471
 - country ownership, 4478
 - "cash on delivery" aid, 4482
 - floating tranche conditionality, 4481
 - future of, 4483
 - meaning of, 4479
 - performance-based conditionality, 4481
 - Poverty Reduction Strategy Papers, 4479–4480
 - social participation, 4480–4481
 - critics of, 4468
 - governance conditionality, 4474–4475
 - versus* ownership, 4472–4474
 - ownership, meaning, 4479
 - policy conditionality, 4469–4470
 - selective aid allocation, 4475
 - conditions for aid effectiveness, 4476
 - developing rules for, 4477
 - dynamic aid allocation, 4477–4478
 - redirecting aid, 4475
- Constrained maximization and dynamic
- reduced-form demand relations
 - data limitations, 4900–4901
 - human capital investments, 4901–4902
 - income and capital markets, 4902–4904
 - parental schooling, 4904–4905
 - preschool experiences and subsequent education
 - intent-to-treat model, 4913
 - malnutrition, 4912
 - summary and implications, 4914
- private educational cost reduction
- conditional cash transfers, 4905–4907
 - educational vouchers, 4907–4910
 - subsidies, 4910–4911
- reduced-form demand relation estimates, 4901
- welfare function, 4899
- Corruption**
- BEEPS II Interactive Dataset: Enterprise Survey in Transition, 2002, 4607
 - budget promptness, 4607
 - bureaucratic responsiveness, 4608
 - control of, 4606
 - on legislative innovation, 4607–4608
 - misallocation of funds, 4608
 - Transparency International view of, 4607
- Corruption and aid, 4459–4461
- Country ownership, conditionality, 4478**
- "cash on delivery" aid, 4482
- floating tranche conditionality, 4481
- future of, 4483
- meaning of, 4479
- performance-based conditionality, 4481
- Poverty Reduction Strategy Papers, 4479–4480
- social participation, 4480–4481
- Credit bureaus, 4734–4735**
- Credit market failure, 4723–4724**
- impacts from solving market failures
 - impact of financial access to rural poor, 4738
 - loans impact on microentrepreneurs, 4738
 - naïve estimate, 4737
 - nonprofit group-lending microfinance program, 4739
 - program placement bias, 4736
 - selection bias, 4736
 - interventions and mechanisms
 - credit scoring and credit bureaus, 4734–4735
 - dynamic incentives, 4735–4736
 - group lending and related mechanisms in practice, 4731–4732
 - group lending in theory, 4729–4731
 - repayment schedules, 4732–4734
 - nature of frictions and policy examples, 4724–4729
 - adverse and advantageous selection, 4726–4727
 - ex ante moral hazard, 4727–4728
 - ex post moral hazard, 4728–4729
- Credit saving, 4706, 4749**
- Credit scoring, 4734**
- D**
- De facto power, governance and development, 4633–4637*
- general lessons, 4637
 - U.S. South, case study, 4635–4637
- Demand-side educational policies, 4886**
- budget constraints, 4889–4891
 - constrained maximization, dynamic reduced-form demand relations, 4899–4914
 - objective functions, 4886–4889
 - production functions, 4891–4899
- Democracy and aid, 4464–4465**
- as necessary condition, 4465–4466
 - as sufficient condition, 4466–4467
- Development Assistance Committee (DAC), 4425**
- Development Economics in Action (Killick), 4599*
- Distortionism, 4664–4665**
- Diversification and industrial policy**
- higher productivity and wages, 4059

- rate of discovery, 4060
- self-discovery, 4059
- Donor policies, 4483
 - evaluation, 4489–4490
 - learning and accountability, 4485–4486
 - proliferation of projects and programs, 4484–4485
 - randomized trials and evidence-based policies, 4486–4489
- E**
- Educational cost reduction
 - conditional cash transfers, 4905–4907
 - educational vouchers
 - Chilean voucher system, 4907–4908
 - community-level data, 4907–4908
 - Enuesta Protección Social (EPS) surveys, 4909
 - Programa de Ampliación de Cobertura de la Educación Secundaria (PACES) program, 4909
 - subsidies, 4910–4911
- Educational investment
 - data types for investigation
 - cross-sectional household data, 4955–4957
 - cross-sectional surveys, 4559
 - educational service providers, 4957
 - longitudinal/panel data, 4957–4559
 - measurement errors, 4954
 - power, sample size, and sample design, 4953–4954
 - qualitative data sources, 4559
 - relevant variables, 4954
 - representativeness, 4953
 - timing and duration, 4954
 - demand side, 4886
 - budget constraints, 4889–4891
 - constrained maximization, dynamic reduced-form demand relations, 4899–4914
 - objective functions, 4886–4889
 - production functions, 4891–4899
 - econometric approaches
 - construction of standard errors, 4950–4951
 - fixed-effects (FE) estimates, 4949–4950
 - instrumental variable (IV) or two-stage least squares (2SLS) estimates, 4948
 - propensity score matching (PSM) estimates, 4949
 - sample selection, 4951
 - framework analysis
 - budget constraints, 4889–4891
 - estimates and policy implications, 4899–4914
- objective functions, 4886–4889
- production functions, 4891–4899
- products and inputs markets
 - educational expenditures and age distributions, 4945–4596
 - sorting in educational markets, 4943–4944
 - supplies and sectoral choices, 4944–4945
- strengths and limitations
 - data quality, 4953–4954
 - education and development, 4955–4959
- supply side
 - estimates and policy implications, 4920–4943
 - modeling, 4915–4920
- Educational vouchers, 4907–4910
- EDUCO decentralization program, 4940
- Emigration impact, on sending countries, 4386–4396
 - information and flow of ideas, 4394–4395
 - labor markets and fiscal accounts, 4386–4390
 - human-capital accumulation, 4389
 - labor supply changes, 4387–4388
 - skilled and unskilled labor, 4386–4387
 - skilled emigration, of taxation, 4388–4389
 - US-Mexico labor-market, 4387–4388
 - wage elasticities, 4387
 - remittances and return migration, 4390–4394
 - empirical identification strategy, 4393–4394
 - in GDP, 4390–4391
 - household expenditure, 4392–4393
 - return on investments, 4391
 - types, 4390–4391
- Employment protection legislation (EPL), 4672
- Employment regulations, 4672–4673
- Endogenous property rights
 - expropriation, 4559–4583
 - commitment, 4563–4567
 - cross-sectional empirical regularities, 4581–4583
 - exit option, 4571–4572
 - framework, 4562–4563
 - heterogeneous producers, 4578–4579
 - no commitment, 4567–4568
 - public ownership, 4574–4576
 - reputation, 4568–4571
 - secrecy, 4572–4574
 - and taxation, 4579–4581
 - voice, 4576–4578
- state effectiveness
 - benchmark, 4584–4585
 - empirical regularities, 4586–4588
 - extensions, 4585–4586

- Enuesta Protección Social (EPS) surveys, 4909
- Equity and governance**
- equality multiplier, 4643–4644
 - general lessons, 4646–4647
 - political economic equilibrium, 4644–4646
 - Scandinavian model of capitalism, 4641–4643
- Equity market liberalization**, 4298, 4316, 4346
- Exchange rate policies**
- classification
 - De facto* exchange rate regime classification, 4227–4228
 - exchange rate changes in fixers and floaters, 4224–4225
 - reserve changes in fixers and floaters, 4224, 4226
 - volatility, 4229
 - inside nonfloat group
 - conventional pegs, 4233
 - fear of floating, 4233, 4236
 - fixed regimes, 4233
 - hard pegs, 4233
 - trends in post-Bretton Woods era, 4230–4233
- Expropriation, endogenous property rights**
- commitment
 - coercive authority, 4564–4566
 - Nash equilibrium, 4564
 - timing structure, 4563–4564
 - Zamindari system, 4565
 - countries expropriation, 4560–4562
 - cross-sectional empirical regularities
 - ICRG expropriation, 4581
 - settler mortality, 4582
 - standard deviation, 4582–4583
 - Cuban revolution, 4562
 - exit option, 4571–4572
 - framework, 4562–4563
 - Glorious Revolution, 4559
 - heterogeneous producers, 4578–4579
 - no commitment, 4567–4568
 - public ownership
 - coercive authority, 4574
 - private ownership economies, 4575–4576
 - reputation
 - coercive authorities, 4568–4572
 - natural resources, 4571
 - optimal credible expropriation rate, 4570
 - secrecy
 - coercive authority, 4572–4574
 - incentive constraints, 4573
 - and taxation
 - coercive authority, 4581
- fundamental transgressions, 4580
- ICRG expropriation, 4579, 4581
- tax systems, 4580
- voice**
- coercive authority, 4576–4577
 - dynamic model, 4577
 - ex ante* optimal expropriation rate, 4578
 - objective function, 4576
- Externalities**
- definition, 4988
 - ecological capital, 4988
 - nature's services, 4988
 - reciprocal
 - bad features, 4996–5000
 - good features, 4995–4996
 - importance, 4994–4995
 - need for CPRs, 4992–4994
 - valuation, evaluation, and sustainable development, 5001
 - types, 4989
 - unidirectional
 - internalizing, 4990–4992
 - quantifying, 4990
 - revenue earning, 4989
- Extractive Industries Transparency Initiative**, 4497
- F**
- Family planning policies and fertility**, 4850
- behavioral demand for births, 4851
 - culture, customs, or religion, 4851
 - family planning effort (FPE) effect on, 4851–4853
- Fertility**
- determinants and policy consequences, 4847
 - Bangladesh, family planning and reproductive health, 4855–4857
 - cross-effects estimation of fertility variation, 4853–4855
 - family planning policies and fertility, 4850–4853
 - income and price effects, 4848–4850
- and development**
- age composition and saving rates, 4845
 - labor force participation, 4845
 - life-cycle pattern of consumption and savings, 4844–4845
 - population explosion, 4844
 - schooling and vocational training, 4846
- and family planning**, 4850–4853
- health and development, stylized facts

- human health capital, determinants and consequences, 4799–4800
- incomplete death registration, 4797–4798
- Fetal origins hypothesis, 4840
- Financial access
 - borrowings, 4707–4708
 - credit market innovations, 4723
 - credit market failure, 4723–4724
 - impacts from solving market failures, 4736–4739
 - interventions and mechanisms, 4729–4736
 - nature of frictions and policy examples, 4724–4729
- financial intermediation and economic growth, 4712–4715
- financial development and poor, 4714
- financial expansion and economic growth, 4713
- interest rate caps, 4713
- financial needs, 4705
- global access, 4710–4712
 - foreign capital investment, 4711
 - gaps in access, 4710–4711
 - microfinance outreach, 4711–4712
- governments, businesses, and nonprofit institutions, 4770–4774
- household-level decision making, 4709
- market structure, 4709
- mechanisms, 4706–4707
- microcredit, 4704–4705
- productive potential of capital, 4705
- randomized controlled trials, 4709
- returns to capital, 4715–4723
 - average returns, 4720–4721
 - estimating profit functions, 4717–4719
 - field experiments
 - evidence from, 4719–4720
 - final caveat, 4722–4723
 - interest rates, sensitivity to, 4721–4722
 - market tests, 4717
 - poor entrepreneurs and high-priced credit, 4716
 - weak prospects of loans, 4716
- risk management and insurance, 4759
 - failure of insurance markets, 4762–4766
 - health insurance, 4769–4770
 - partnership models and index-based insurance, 4766–4769
 - persistence of shocks, 4760–4761
- saving, economics of, 4739
 - basic models of saving, 4742–4744
 - concepts of poverty-destitution, problems with, 4741–4742
 - constraints, 4745–4757
 - impacts, 4757–4759
 - revisiting “the poor can't save” assumption, 4740–4741
 - savings, 4707–4708
- Financial crises
 - banking crises, 4306
 - manifestations, 4306
 - short-term debt, 4303
- Financial globalization and economic policies
 - capital, 4287
 - capital controls as policy tool
 - implementation and measurement issues, 4337–4338
 - macroeconomic implications, 4338–4339
 - microeconomic implications, 4339–4341
 - capital flows and economic growth
 - developing countries, investment-constraint, 4336
 - empirical evidence, 4332–4334
 - interpretation of findings, 4334–4336
 - policy implications, 4337
 - economic policies and growth outcomes, 4306–4309
 - evolution, 4289
 - composition of foreign assets and liabilities, 4292
 - de jure* integration, 4291
 - gross international financial assets and liabilities, 4290
 - factors driving, 4292
 - financial integration, measurement
 - capital controls, 4288–4289
 - financial flows/stocks, 4289
 - macroeconomic implications, 4292
 - empirical evidence on direct channels, 4296–4299
 - empirical evidence on indirect channels, 4299–4302
 - theory of economic growth, 4294–4296
 - macroeconomic policies
 - evidence, 4330–4332
 - theory, 4330
 - macroeconomic volatility and patterns of international risk-sharing
 - crises as special case of volatility, 4304–4306
 - theory, 4303
 - merits, 4286
 - policy relevance, 4285

- Financial globalization and economic policies
(Continued)
- research on, 4285
 - structural and policy-related factors, 4287
 - structural characteristics and economic policies
 - financial sector development, 4310–4316
 - institutional quality, 4316–4323
 - trade openness, 4323–4329
- Financial integration
- approaches, 4341–4342
 - and consumption volatility, 4303
 - macroeconomic policies, 4301–4302
 - measurement
 - capital controls, 4288–4289
 - financial flows/stocks, 4289
 - and productivity growth, 4302
- Financial intermediation and economic growth, 4712–4715
- financial development and poor, 4714
 - financial expansion and economic growth, 4713
 - interest rate caps, 4713
- Financial literacy, 4749–4751
- Financial repression, 4713
- Fixed-effects (FE) estimates, 4949–4950
- Floating and Inflation targeting (FIT) model, 4239–4240
- Floating tranche conditionality, 4481
- Foreign aid
- donor policies, 4483
 - evaluation, 4489–4490
 - learning and accountability, 4485–4486
 - proliferation of projects and programs, 4484–4485
 - randomized trials and evidence-based policies, 4486–4489
 - effectiveness, 4418–4420
 - case studies, 4444–4445
 - cross-country evidence, 4445–4447
 - project evaluations, 4443–4444
 - and social indicators, 4447–4448
 - fiscal policy and volatility, 4452–4454
 - to fragile states, 4462–4463
 - global poverty, 4417
 - and governance, 4454
 - aid dependence and growth, 4455
 - and common-pool problem, 4458–4459
 - and corruption, 4459–4461
 - history and the state, 4456–4457
 - political leadership, 4457–4458
 - and growth in theory
 - binding constraints, 4439–4441
- capital shortage and neoclassical growth models, 4435–4437
- complementary inputs, 4441–4442
- growth *versus* basic needs, 4442
- poverty-trap models, 4437–4438
- virtuous circles and amplification effects, 4439
- history of aid, 4424
- meeting basic needs, 4490
- agriculture and rural development, 4492–4493
 - public service delivery, 4491–4492
- The Millennium Development Goals, 4423
- operational approach, 4421
- partnership model, 4463
 - aid for democratic governments, 4464–4465
 - democracy as necessary condition, 4465–4466
 - democracy as sufficient condition, 4466–4467
- partnership model for donors, 4420
- political instability and war, 4461–4462
- prospects for scaling-up aid, 4421–4422
- recent trends, 4425–4431
 - allocation of aid, 4427–4428
 - average aid intensity, 4430–4431
 - debt relief, 4433–4435
- Development Assistance Committee (DAC), 4425
- forms, 4431–4432
- fungibility, 4432–4433
- generosity of aid from developed countries, 4429
- official aid, 4431
- overseas development assistance (ODA), 4425
- share in gross national income, 4429
- and trade theory
- Dutch disease, 4450–4452
 - transfer problem, 4449–4450
- for twenty-first century, 4493–4497
- volatility, 4453–4454
- Foreign direct investment (FDI) and industrial policy, 4099
- and host-country factor markets, 4108–4110
- and productivity, 4103–4108
- alliances with foreign firms, 4104
 - fiscal incentives for foreign enterprise, 4107–4108
 - horizontal spillovers, 4105–4106
 - joint ventures, 4104
- sectors targeted for FDI inflows, 4100–4102
- and technology spillovers, 4185–4195
- Framework analysis, educational investment budget constraints, 4889–4891

- constrained maximization, dynamic reduced-form demand relations
 data limitations, 4900–4901
 human capital investments, 4901–4902
 income and capital markets, 4902–4904
 parental schooling, 4904–4905
 preschool experiences and subsequent education, 4911–4914
 private educational cost reduction, 4905–4911
 reduced-form demand relation estimates, 4901
 welfare function, 4899
 objective functions, 4886–4889
 production functions
 education as endogenous, 4897–4898
 empirical estimates, 4893, 4895–4896
 international migration, 4898
 knowledge and skills production functions, 4891–4892
 marginal productivity, 4893–4894
 summary and implications, 4898–4899
 wage production function estimates, 4896–4897
- G**
 Gains from trade, identifying mechanisms, 4095–4099
 consumption of increased variety, 4095
 importing goods that embody new technology, 4095–4096
 increasing competition, 4096
 learning by doing through exporting, 4098–4099
 reallocation of market shares to most productive firms, 4096–4097
 General equilibrium analyses and foreign aid and Dutch disease, 4450–4452
 resource transfer problem, 4449–4450
 Global public goods, 5011–5012
 Glorious Revolution of 1688, 4610
 Governance and development
 bad governance, 4598–4599, 4604–4605
 consequences and causes of variation in correlates of governance, 4615–4620
 first stage, 4620–4625
 origins of bad governance in Sierra Leone, 4625–4629
 definition, 4602–4603
 improving governance, 4629
 governance and equity, 4641–4647
 Iron Law of Oligarchy, 4637–4640
 persistence of *de facto* power, 4633–4637
 pitfalls of reform, 4630
 promoting reform, 4647–4648
 See-Saw effect, 4631–4633
 successful reform, 4640–4641
 measurement
 political institutions, 4603
 state capacity and effectiveness, 4603–4608
 policy view, 4601
 private governance, 4600
 promoting reform, 4647–4648
 public governance, 4600
 Scandinavian model of governance and equity
 equality multiplier, 4643–4644
 general lessons, 4646–4647
 political equilibrium, 4644–4646
 rise of, 4641–4643
 theoretical mechanisms, 4608
 collective choices of society, 4609
 de jure and *de facto* political power, 4609–4610
 Sierra Leone, political economy of governance, 4610–4615
 Governance conditionality, 4474–4475
 Government effectiveness, 4606
 meritocratic bureaucracy, 4624
 and weberianness, 4624–4625
 Grameen Bank, 4704, 4706–4707, 4712, 4728, 4749, 4769, 4772
 Group lending
 and related mechanisms in practice, 4731–4732
 in theory, 4729–4731
 Guard labor, 4530–4534
 binding labor-endowment constraint, 4532–4533
 Lagrangian multiplier, 4532
 and productive labor, 4531–4532
 theoretical predictions, 4534
- H**
 Handbook of Development Economics (Gersovitz), 4706
 Harris-Todaro two-sector model, 4671
 Health and population policies, 4786
 death and birth rates, historical transition
 age-specific pattern of mortality, 4789
 childhood mortality reduction, 4790
 international population policy assistance, 4792–4797
 reduction in fertility, 4790–4792

- Health and population policies (*Continued*)**
- determinants and consequences, framework for studying
 - estimation without bias health production functions, 4802–4804
 - life expectancy at birth and nation income, 4801
 - marginal product of health human capital, estimation, 4804–4805
 - mortality transition phases, 4812–4813
 - parents' investment in children, 4805–4810
 - spillovers from health human capital, 4810–4812
 - determinants and economic growth,
 - macroeconomic evidence
 - development and disease environment, contemporary relationship, 4814
 - health affecting development, assessment, 4814–4815
 - health status and economic consequences, 4818–4820
 - impact of income on health, 4816–4818
 - and development, macroeconomics of
 - fetal and early child development to adult productive life span, 4840–4843
 - heterogeneity of height, 4828
 - indicators of health stocks, 4822–4828
 - fertility and development, stylized facts
 - human health capital, determinants and consequences, 4799–4800
 - incomplete death registration, 4797–4798
 - macroeconomics of, 4844–4847
 - fertility determinants and policy consequences
 - Bangladesh, family planning and reproductive health, 4855–4857
 - cross-effects estimation of fertility variation, 4853–4855
 - family planning policies and fertility, 4850–4853
 - income and price effects, 4848–4850
 - Health human capital
 - adult height
 - and schooling, 4838–4839
 - and wages, 4835–4837
 - birth weight, 4833–4834
 - body mass index (BMI), 4831–4832
 - business cycle conditions and fetal development, 4843
 - determinants and consequences of accumulating, 4799–4800
 - heterogeneity of height as measure of, 4828
 - and hookworm infection eradication, 4843
 - human capital inputs, indicators, 4837–4838
 - marginal product estimation, 4804–4805
 - spillovers, 4810–4812
 - Health inequality, 4812–4813
 - Health insurance, 4769–4770
 - Health production function estimation, 4802–4804
 - Height, 4822–4823
 - season-of-birth effects, 4841
 - and wage returns, 4829–4830
 - Heterogeneous schooling attainment impacts, 4897
 - Higher Impact Adjustment Lending (HIAL) initiative, 4481
 - Human capital investments, 4901–4902
 - Human Development Index (HDI), 5008, 5010

I

- IMF Government Financial Statistics, 4793
- Immigration policy regimes
 - Doha Development Agenda Mode IV, 4396
 - policy design, 4398–4402
 - cross-country differences, 4399
 - ex ante* selection criteria, 4399
 - ex post* selection criterion, 4399
 - illegal immigration, 4399
 - legal and illegal immigration, 4400–4402
 - legal immigration regulation, 4398–4399
 - political economy, 4396
 - fiscal consequences, 4397–4398
 - international factor mobility, 4397
 - pension systems, 4398
 - voters pressure, 4397
- incomplete death registration, 4797–4798
- India
 - brain drain from, 4389
 - common property resources (CPRs), 4992–4993
 - domestic content requirements, 4106
 - financial globalization, 4285
 - labor laws, 4673
 - labor migration, 4364
 - nonprofit group-lending microfinance program, 4739
 - "Operation Blackboard," 4934
 - public service delivery, 4491
 - tariff protection, 4080
 - tax consequences of skilled emigration, 4389
 - Zamindari system, 4565–4566
- Industrial policy (IP), 4041
 - foreign direct investment (FDI), 4099
 - and host-country factor markets, 4108–4110

- and productivity, 4103–4108
- sectors targeted for FDI inflows, 4100–4102
- and technology spillovers, 4185–4195
- infant-industry protection, empirical
 - evidence, 4041
 - cross-country studies, 4067–4069
 - cross-industry studies, 4065–4067
 - export subsidies, 4070–4071
 - Marshallian externalities, 4062
 - Mill and Bastable tests, 4062
 - single-industry studies, 4063–4065
 - tariffs, 4063
 - trade protection, 4063
- Marshallian externalities, 4043–4044
- multiple equilibria and latent comparative
 - advantage
 - discussion, 4051–4052
 - dynamic externalities, 4049–4050
 - exogenous international prices, 4045–4047
 - international prices determined in north, 4047–4049
- national and global gains, 4060–4062
- as sector-specific collective action, 4054–4059
 - relation of the model with recent contributions, 4057–4059
 - simple model, 4055–4057
- self-discovery and diversification, 4059–4060
- trade and economic development, 4071
 - mechanisms for gains from trade, identification, 4095–4099
 - trade policies, stylized facts, 4072–4080
 - trade policies, trade volumes productivity and growth, 4080–4091
 - trade volume vs. trade policy, 4091–4092
- without latent comparative advantage
 - interindustry externalities, 4053–4054
 - sector-level rents, 4052–4053
- Infant-industry protection, empirical evidence, 4041
 - cross-country studies
 - capital accumulation, 4069
 - grouping of, 4067
 - import tariffs and economic growth, 4068
 - pattern of protection, 4068–4069
 - cross-industry studies
 - fundamental problems, 4066
 - in granting import licenses, 4066–4067
 - removal of protection, 4065
 - tariff protection for declining industries, 4066
 - trade protection and productivity growth, 4065
 - export subsidies, 4070–4071
- Marshallian externalities, 4062
- Mill and Bastable tests, 4062
- single-industry studies, 4063–4065
 - Airbus's entry impact on US and European welfare, 4064
 - Japanese semiconductor industry protection, 4063–4064
 - microcomputer industry protection, 4065
 - production subsidies in Denmark, 4064
 - tariff protection effects in steel rail industry, 4064
 - tariff protection effects in tinplate industry, 4064
 - tariffs, 4063
 - trade protection, 4063
- Infant mortality, 4827–4828, 4866, 4869
- Inflation targeting
 - demand and supply shocks, 4262
 - as monetary policy, 4236–4240
- Informal labor market
 - employment data, 4686
 - rapid growth, 4685
 - sectoral mobility, 4685
 - traditional view, 4684
 - wages comparison, 4687
- Institutional quality
 - empirical evidence, 4317–4323
 - theory, 4316–4317
- Instrumental variable (IV)/two-stage least squares (2SLS) estimates, 4948–4949
- Insurance and risk management
 - failure of insurance markets
 - adverse selection and moral hazard, 4762
 - economic costs of illness, 4764
 - gifts and transfers, 4764
 - informal village insurance, 4763
 - lack of full regional insurance, 4765
 - limited commitment, 4764
 - mutual insurance system, 4763–4764
 - regional risk-sharing, 4763
 - risk-bearing and income distribution, 4762
 - health insurance, 4769–4770
 - partnership models and index-based insurance, 4766–4769
 - persistence of shocks, 4760–4761
 - poverty and vulnerability, 4759
- Inter-American Development Bank, 4693
- Internal migration, 4857, 4859
- International country risk guide (ICRG), 4554

- International migration and developing world
 cognitive dissonance, 4365
 dimensions, 4366–4386
 data sources, 4367–4368
 networks and migration costs, 4383–4385
 patterns, 4368–4375
 skilled emigration, 4375–4383
 and education, 4898
 emigration impact, on sending countries,
 4386–4396
 information and flow of ideas, 4394–4395
 labor markets and fiscal accounts, 4386–4390
 remittances and return migration, 4390–4394
 ideas transmission, 4364
 immigration policy regimes, 4396–4402
 policy design, 4398–4402
 political economy, 4396–4398
 labor flows, 4364
 patterns of
 bilateral migration flow, 4373–4376
 foreign-born population, 4369, 4371–4372
 migrants source, 4368–4369
 migration and income, 4373
 OECD immigrants share, 4368–4370
 propensity to emigrate, 4372–4373
- International population policy assistance, 4792
 expenditures on health and family planning,
 4793–4794
 fertility control in terms of contraception, 4797
 trends and differentials in population policies,
 4795–4796
- International risk-sharing and macroeconomic
 volatility
 crises as special case of volatility, 4304–4306
 theory, 4303
- Iron Law of Oligarchy, 4637–4640
 Bolivian revolution, case study, 4638–4639
 general lessons, 4639–4340
- L**
- Labor institutions
 country cases, 4682
 Argentina, 4684
 China, 4683–4684
 cross-country evidence, 4680–4682
 labor subindex of the Fraser Institute,
 4681–4682, 4693
 measures of rigidity, 4681
 effect on outcomes, 4660
 employment regulations, 4672–4673
 income distribution, 4659, 4662, 4666
- increasing wages, 4659
 informal sector, 4684–4687
 mandated benefits, 4674
 measures of institutional orientation,
 4661–4663
 minimum wages, microevidence on, 4667–4670
 productivity, 4659, 4665, 4675
 responses of employment to output and wages
 Labor institutions, 4659–4660
 theoretical perspectives
 communications for better decisions, 4665
 distortionism, 4664–4665
 impact on outcome, 4666–4667
 institutions as effective bargaining, 4665
 union effects, 4674–4680
 wage curve, 4670–4671
- Labor regulations, 4660–4661, 4673, 4681, 4693
 Lender's payoff function, 4541
 Life-cycle hypothesis (LCH), 4742–4743
 Loan demand and interest rates, 4721–4722
 Loss aversion, 4752–4753
- M**
- Macroeconomic volatility and patterns of
 international risk-sharing
 crises as special case of volatility, 4304–4306
 theory, 4303
- Malnutrition, in children, 4912
- Marshallian externalities, 4043–4044, 4061, 4067
 dynamic externalities, 4049–4050
 and multiple equilibria, 4046–4047
- Mexican Household's Income and Expenditure
 surveys, 4758
- Mexican National Survey of Microenterprises, 4718
- Microcredit, 4704–4705
 lending models and outreach, 4712
 limiting financial access, 4711
 poor entrepreneurs and high-priced
 credit, 4716
 returns to financial access, 4715
 unbanked and underbanked adults, 4710
- Microcredit Summit Campaign Report of 2007,
 4710
- Microfinance, 4708
 foreign capital investment, 4711
 outreach of, 4711–4712
- Microfinance Information Exchange (MIX), 4712
- Microinsurance, 4708, 4756, 4760
- Migration, 4659
- The Millennium Development Goals, 4423, 4429,
 4481

- M**
- Minimum wages
 - in developing countries, 4668–4669
 - elasticity of employment, 4667
 - employment effects, 4667
 - lighthouse effect, 4670
 - and poverty, 4670
 - Monetary aggregates, measurement drawbacks, 4236–4238
 - Monetary and exchange rate policies (MERP)
 - classical dichotomy, 4217
 - classical view of, 4217
 - on economic integration, 4249–4251
 - exchange rate anchors in 1980s, 4254–4258
 - exchange rate and growth, link between, 4241–4246
 - depreciation, 4245
 - exchange rate regimes, effects, 4243–4244
 - overvaluation, 4245–4246
 - undervaluation, 4245
 - exchange rate regimes and monetary policy frameworks, 4220
 - exchange rate anchors to inflation targeting, 4236–4240
 - exchange rate policies, classification, 4224–4230
 - exchange rate policy trends in post-Bretton Woods era, 4230–4233
 - inside nonfloat group, 4233–4236
 - mercantilist motive, 4265–4267
 - prudential motive, 4263–4265
 - and financial development, 4251–4253
 - financial integration and crises in 1990s, 4258–4260
 - fear of floating, 4259
 - ineffectiveness of monetary policy, 4258
 - unilateral dollarization, 4259
 - Floating and Inflation targeting (FIT), 4261–4263
 - implications, 4216
 - on income distribution, 4248–4249
 - nominal variables and development, 4217–4220
 - exchange rate flexibility, 4219
 - real models, 4218
 - and output volatility, 4246–4247
 - on price stability, 4247–4248
- Mortality**
- mortality variation by month of birth, 4841
 - reason for decline
 - calories consumption, 4821
 - nutrition literature, 4821–4822
 - rising standards of living, 4821
 - transition phases, 4812–4813
- N**
- Nash bargaining formula, 4547
 - Nash equilibrium, 4564
 - Natural capital
 - and development economics, 4982–4985
 - and economic development, 4980–4982
 - ecosystems, 4981
 - macroeconomic, 4980–4981
 - types, 4985–4987
 - competing uses, 4986
 - nonlinear stochastic processes, 4986–4987
 - visible and invisible products, 4985–4986
 - Nature, in economic development
 - comprehensive investment and sustainable development
 - application, 5012–5015
 - basic theory, 5009–5011
 - global public goods, 5011–5012
 - discounting climate change, 5021
 - climate treaties, 5025–5026
 - global activism, 5021–5022
 - inequality and risk aversions, 5024–5025
 - negative discount rates, 5022
 - welfare economics, 5022–5024
 - economic perturbation
 - formal analysis, 5029–5031
 - population growth accounting, 5031–5033
 - empirical perspectives, 4978–4979
 - environmental problems, 4979
 - environmental resources, in improvements, 4980
 - evaluating policy reforms
 - projects evaluation, 5017–5020
 - social cost-benefit analysis, 5016–5017
 - externalities
 - definition, 4988
 - ecological capital, 4988
 - nature's services, 4988
 - types, 4989
 - GDP, inadequacies, 5001–5002
 - industrial and agricultural pollutants, 4979–4980
 - natural capital
 - and development economics, 4982–4985
 - and economic development, 4980–4982
 - types, 4985–4987
 - overview, 4987–4989
 - policy evaluation, 5033–5034
 - reciprocal externalities
 - bad features, 4996–5000
 - good features, 4995–4996

Nature, in economic development (*Continued*)
 importance, 4994–4995
 need for CPRs, 4992–4994
 valuation, evaluation, and sustainable development, 5001
 social well-being and comprehensive wealth
 net domestic product, 5008–5009
 unifying model, 5007–5008
 unidirectional externalities, 4989
 internalizing, 4990–4992
 quantifying, 4990
 valuing goods and services, 5002–5006
 Net domestic product (NDP), 5008–5009
 New Immigrant Survey (NIS), 4380, 4386
 Nonwage compensation, 4674

O

Openness
 and endogeneity problems, 4085–4088
 and growth, 4089, 4118
 measuring, 4081–4085
 need for openness to trade, 4090
Oportunidades program, 4905
 Optimal assignment, property rights
 asset ownership
 Nash bargaining formula, 4547
 optimal allocation theory, 4548
 property rights approach, 4546
 symmetric expressions, 4548
 communal property rights, 4550–4552
 tenancy role, 4548–4550
 Output volatility, 4303–4304
 Overseas development assistance (ODA), 4425

P

Parental preference functions, 4888
 Parental schooling, 4904–4905
 Patrimonialism, 4612–4614, 4625–4628
 Payment for ecosystem services (PES), 4991
 Performance-based conditionality, 4481
 Policy conditionality, 4469–4470
 Policy neutrality, 4041
 Policy reforms evaluation, 5016–5020
 projects evaluation
 cost benefit analysis, 5020
 ecological capital restoration, 5019
 Little–Mirrlees method, 5017–5018
 population pressure, 5020
 random experiments, 5018
 retrospective evaluation, 5019
 social cost-benefit analysis, 5015–5017

Policy-related educational supply-side changes
 decentralization and local community participation
 EDUCO decentralization program, 4940
 natural experiment, 4941
 summary and implications, 4942–4943
 quality changes
 postschooling education, 4936–4937
 preschool education, 4932
 school-age education, 4932–4936
 suppliers quantity increases
 ECD programs, 4924–4925
 preschool education, 4924–4925
 school-age education, 4925–4932
 supply-side providers and employees incentives
 monitoring and financial incentive program, 4938
 school programs, 4937–4938
 Polluter-pays-principle (PPP), 4990
 Population and health policies, 4786
 death and birth rates, historical transition
 age-specific pattern of mortality, 4789
 childhood mortality reduction, 4790
 international population policy assistance, 4792–4797
 reduction in fertility, 4790–4792
 fertility and development, macroeconomics of, 4844–4847
 fertility determinants and policy consequences, 4847
 Bangladesh, family planning and reproductive health, 4855–4857
 cross-effects estimation of fertility variation, 4853–4855
 family planning policies and fertility, 4850–4853
 income and price effects, 4848–4850
 health and development, macroeconomics of, 4820
 fetal and early child development to adult productive life span, 4840–4843
 heterogeneity of height, 4828
 indicators of health stocks, 4822–4828
 health determinants and consequences, framework for studying, 4800
 estimation without bias health production functions, 4802–4804
 life expectancy at birth and nation income, 4801
 marginal product of health human capital, estimation, 4804–4805

- mortality transition phases, 4812–4813
 - parents' investment in children, 4805–4810
 - spillovers from health human capital, 4810–4812
 - health determinants and economic growth, macroeconomic evidence, 4813
 - development and disease environment, 4814
 - health affecting development, assessment, 4814–4815
 - health status and economic consequences, 4818–4820
 - impact of income on health, 4816–4818
 - health, fertility and development, stylized facts
 - human health capital, determinants and consequences, 4799–4800
 - incomplete death registration, 4797–4798
 - internal migration, demographic transition and policy evaluation, 4857–4860
 - life expectancy, 4787
 - policy directions, 4861–4864
 - population policy research, agenda for, 4788
 - Postschool years, 4887
 - Poverty-destitution, problems with
 - borrowings, 4742
 - caloric requirements, 4741–4742
 - international poverty line, 4741
 - Poverty Reduction Strategy Papers (PRSPs), 4479–4480
 - Poverty-trap models, 4437–4438
 - Preschool years, 4887
 - Production functions, 4891–4899
 - Products and inputs markets, educational
 - investment
 - educational expenditures and age distributions, 4945–4956
 - educational markets, 4943–4944
 - supplies and sectoral choices, 4944–4945
 - Programa de Ampliación de Cobertura de la Educación Secundaria (PACES) program, 4909
 - PROGRESA program, 4888, 4905–4907
 - Propensity score matching (PSM) estimates, 4949
 - Property Rights and Economic Development and contracts, 4527
 - definition, 4526
 - endogenous property rights
 - expropriation, 4559–4583
 - state effectiveness, 4583–4588
 - microeconomic approach, 4528
 - resource allocation and property rights, 4558
 - evidence of, 4552–4558
 - limiting expropriation, 4529–4534
 - optimal assignment, 4545–4552
 - trade barrier, 4534–4545
 - Public sector educational service
 - decentralization, 4920
 - incentives and effort, 4919–4920
- R**
- Reference currency, 4229
 - Regulatory quality, 4606
 - Remedial education program, 4935
 - Reproducible health human capital, 4811–4812, 4828
 - Resource allocation and property rights
 - evidence of
 - Field and Torero study, 4558
 - Galiani and Schargrodsy's study, 4557
 - International Country Risk Guide (ICRG), 4554
 - private sector banks, 4558
 - theoretical parameter, 4553
 - trade channels, 4552
 - World Bank, 4554
 - limiting expropriation
 - basic model, 4529–4530
 - guard labor, 4530–4534
 - optimal assignment
 - communal property rights, 4550–4552
 - optimal asset ownership, 4545–4548
 - tenancy role, 4548–4550
 - as trade barrier
 - collateralizability of assets, 4537–4545
 - trade in assets, 4534–4537
 - Returns to capital, approaches for measuring, 4718
 - Risk-sharing and macroeconomic volatility
 - crises as special case of volatility, 4304–4306
 - theory, 4303
 - Rotating savings and credit associations (ROSCAs), 4733, 4740, 4745, 4747
 - Rule of law, defined, 4606
 - Rural savings behavior, 4740
- S**
- Save More Tomorrow™ (SMarT), 4753–4754
 - Saving, economics of, 4739
 - basic models, 4742–4744
 - concepts of poverty-destitution, problems with, 4741–4742
 - constraints, 4745–4757
 - default options, 4753–4754
 - loss aversion and mental accounting, 4752–4753

- Saving, economics of (*Continued*)
 - marketing, 4754
 - planning and financial literacy, 4749–4751
 - pricing, 4751–4752
 - self and spousal control, 4745–4749
 - social networks and peer effects, 4754–4755
 - technology, 4755–4757
 - impacts, 4757–4759
 - revisiting “the poor can’t save” assumption, 4740–4741
- Saving for Change program, 4755
- Scandinavian model of capitalism, case study, 4641–4643
- School-age years, 4887
- Season-of-birth effects, 4841
- See-Saw effect, governance and development
 - central bank independence (CBI), case study, 4631–4633
 - general lessons, 4633
- Shadow prices, 5002–5006
 - definition, 5003
 - estimation
 - ecological capital assets, 5006
 - economics of climate change, 5005–5006
 - hedonic and contingent valuation method (CVM), 5004
 - limitation, 5004–5005
 - methods, 5003–5004
- Sierra Leone
 - bad governance, origins of
 - patrimonialism, general ideas, 4625–4628
 - in Sierra Leonean context, 4628–4629
 - political economy of governance, 4610–4615
- Skilled emigration
 - brain drain
 - DM data, 4375
 - emigration and human-capital relation, 4377
 - opportunity, 4376
 - skill acquisition, 4375–4376
 - inequality and selection, 4377–4383
 - migration cost, 4378, 4381–4382
 - Mincer wage regressions, 4380
 - New Immigrant Survey (NIS) report, 4380–4381
 - OECD immigrants by destination region and education level, 4382–4383
 - positive and negative-selection hypothesis, 4378–4380
 - relative base wages estimation, 4380–4381
 - reward to skill, 4379
- Supply side educational policies
 - educational service providers, 4915–4918
 - demand function, 4915
 - first-order conditions, 4916–4918
 - production function technologies, 4918
 - knowledge and skill production functions
 - educational inputs, 4921
 - educational production functions, 4922
 - marginal products, 4923
 - peer effects, 4923–4924
 - policy-related educational supply-side changes
 - decentralization and local community participation, 4939–4943
 - quality changes, 4932–4937
 - suppliers quantity Increases, 4924–4932
 - supply-side providers and employees
 - incentives, 4937–4939
 - public sector educational service
 - decentralization, 4920
 - incentives and effort, 4919–4920
 - objective functions, 4918–4919
- Sustainable development
 - application, 5012–5015
 - comprehensive investment, 5012
 - poor nations progress, 5012–5015
 - basic theory, 5009–5011
 - corruption, 5010
 - economy's productive base, 5010
 - population growth, 5011
 - wealth, 5010
 - economic perturbation
 - formal analysis, 5029–5031
 - population growth accounting, 5031–5033
 - global public goods, 5011–5012
- T**
- Trade and economic development, 4071
 - mechanisms for gains from trade, identification consumption of increased variety, 4095
 - importing goods that embody new technology, 4095–4096
 - increasing competition, 4096
 - learning by doing through exporting, 4098–4099
 - reallocation of market shares to most productive firms, 4096–4097
 - trade policies, stylized facts, 4072–4080
 - trade policies, trade volumes productivity and growth, 4080
 - endogeneity problems, 4085–4088
 - measuring openness, 4081–4085

- selection of variables, 4088–4091
- trade volume vs. trade policy, 4091–4092
- Trade barrier, property rights as
 - autarky option, 4536–4537
 - binding constraint, 4542
 - de Soto effect, 4543
 - formal and informal sector, 4543–4545
 - lender's payoff function, 4541
 - optimal debt contract function, 4540
 - resource allocation, 4542–4543
- Trade integration
 - empirical evidence
 - growth and volatility, 4326–4329
 - structural factors and growth, 4324–4325
 - trade openness, 4326
 - theory, 4323
- Trade policies 1980–2004
 - evolution of country trade policies, 4072–4080
 - export shares in developing and industrial countries, 4072
- Trade theory and aid
 - Dutch disease, 4450–4452
 - transfer problem, 4449–4450
- Twins and fertility, 4853–4855

- U**
- Unidirectional externalities
 - internalizing
 - payment for ecosystem services, 4991–4992
 - private property rights, 4990–4991

- quantifying, 4990
- revenue earning, 4989
- Union effects in developing countries, 4674–4680
- Unionism and economic outcomes, 4675–4677
- U.S. South before and after Civil War, case study, 4635–4637

- V**
- Vitamin D deficiency and child development, 4843, 4870
- Voluntary default, 4728–4729

- W**
- Wage compression policy, 4645
- Wage curve, 4670–4671
- Wage earning functions, 4894
- Wage production function estimates
 - cognitive skills, 4896–4897
 - school quality and quantity, 4896
- World Bank, 4554
 - Country Policy and Institutional Assessment (CPIA), 4476
 - Database of Labor Market Indicators Across Countries, 4680
 - 1990 proclamation, 4660–4661
- World Development Indicators, 4425
- World Economic Forum's Global Competitiveness Report, 4679
- Woytinsky Lecture, 4903

- Z**
- Zamindari system, 4565