

Second Edition

ADVANCED INTERNATIONAL TRADE

Theory and Evidence



ROBERT C. FEENSTRA

ADVANCED INTERNATIONAL TRADE

Second Edition



Advanced International Trade

THEORY AND EVIDENCE

Second Edition

Robert C. Feenstra

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Copyright © 2016 by Princeton University Press

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 6 Oxford Street, Woodstock, Oxfordshire
OX20 1TW

press.princeton.edu

Jacket image: *Wards Island Bridge*, Bascove © 2001, private collection

All Rights Reserved

ISBN 978-0-691-16164-8

Library of Congress Control Number 2015946714

British Library Cataloging-in-Publication Data is available

This book has been composed in Minion Pro and Myriad

Printed on acid-free paper. ∞

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

To Heather and Evan

CONTENTS

Acknowledgments ix
Foreword to the Second Edition xi

Chapter 1

Preliminaries: Two-Sector Models 1

Chapter 2

The Heckscher-Ohlin Model 25

Chapter 3

Many Goods and Factors 51

Chapter 4

Trade in Intermediate Inputs and Wages 83

Chapter 5

Monopolistic Competition and the Gravity Equation I 119

Chapter 6

Monopolistic Competition and the Gravity Equation II 155

Chapter 7

Gains from Trade and Regional Agreements 186

Chapter 8

Import Tariffs and Dumping 214

Chapter 9

Import Quotas and Export Subsidies 256

Chapter 10

Political Economy of Trade Policy 299

Chapter 11

Trade and Endogenous Growth 331

Chapter 12

Multinationals and Organization of the Firm 360

Appendix A

Price, Productivity, and Terms of Trade Indexes 403

Appendix B

Discrete Choice Models 419

References 431

Index 465

ACKNOWLEDGMENTS

A book like this would not be possible without the assistance of many people; indeed, without a career full of teachers and colleagues who have shaped the field and my own views of it. I am fortunate to have had many of these who were all generous with their time and insights. As an undergraduate at the University of British Columbia, I first learned international trade from the original edition of Richard Caves and Ronald Jones, *World Trade and Payments* (Little, Brown, 1973), and then from the “dual” point of view with Alan Woodland. His continuing influence will be clear in this book. I also took several graduate courses in duality theory from Erwin Diewert, and while it took longer for these ideas to affect my research, their impact has been quite profound.

In 1977 I began my graduate work at MIT, where I learned international trade from Jagdish Bhagwati and T. N. Srinivasan, both leaders and collaborators in the field. These courses greatly expanded my knowledge of all aspects of international trade and stimulated me to become a trade economist. The book I have written probably does not do justice to the great breadth of research topics I learned there. These topics are well presented in the textbook by Jagdish Bhagwati, Arvind Panagariya, and T. N. Srinivasan, *Lectures in International Trade* (MIT Press, 1998), to which the reader is referred.

In 1981, I joined Jagdish Bhagwati as a colleague at Columbia University, and for the next five years enjoyed the company of Ronald Findlay, Guillermo Calvo, Robert Mundell, Maury Obstfeld, Stanislaw Wellisz, and other regular attendees at the Wednesday afternoon international seminar. Richard Brecher was also a visitor at Columbia during this period. I am much indebted to these colleagues for the stimulating seminars and conversations. My early work in trade did not stray too far from the familiar two-sector model, but in 1982 I was invited by Robert Baldwin to participate in a conference of the National Bureau of Economic Research (NBER) focusing on the empirical assessment of U.S. trade policies. For this conference I prepared a paper on the voluntary export restraint with Japan in automobiles, and thus my empirically oriented research in trade began.

Following this, I participated in more conferences of the NBER, and since 1992 have directed the International Trade and Investment group there. These conferences have been enormously influential to my research. During the 1980s, there was a large amount of research on trade and trade policy under imperfect competition. Some of this work was done for NBER conferences, and I was able to witness all of it. In the 1990s the focus shifted very substantially to “endogenous” growth models, and this research was again supported by a working group at the NBER and many other researchers. I was fortunate to spend a sabbatical leave in 1989 at the Institute for Advanced Study of the Hebrew University of Jerusalem, along with Wilfred Ethier, Gene Grossman, Paul Krugman, and James Markusen, and under the organization of Elhanan Helpman and Assaf Razin, during which time these growth models were being developed. Furthermore, throughout the 1980s and 1990s there has been an increased awareness of empirical issues in international

trade, with fine work by many new colleagues. And during the past decade a great deal of attention has been given to the role of heterogeneous firms in shaping trade patterns and in the gains from trade. So in the span of three decades since beginning my career, I have seen four great waves of research in international trade, and I attempt to summarize all of them in this second edition of the book.

My own research, much of it included here, would not have been possible without generous funding from the National Science Foundation, the Ford Foundation, and the Sloan Foundation, as well as the assistance of many graduate students over the years. Some of their research appears in this book, and many others have labored faithfully to construct datasets that have been widely distributed. Let me thank in particular Dorsati Madani, Maria Yang, and Shunli Yao for their past work on datasets, along with David Yue, Roger Butters, Seungjoon Lee, Songhua Lin, and Alyson Ma for their help with the first edition of this textbook. My colleagues and students at UC Davis have been very helpful in discussing this second edition, and I especially thank Vladimir Tyazhelnikov, who provided invaluable research assistance. Finally, I would like to thank a number of other colleagues who contributed to this book: Lee Branstetter provided a number of the empirical exercises, with the STATA programs written by Kaoru Nabeshima; Bin Xu read nearly all the chapters at an early stage and provided detailed comments; and Bruce Blonigen, Donald Davis, Earl Grinols, Doug Irwin, Jiandong Ju, James Levinsohn, Nina Pavcnik, Larry Qiu, James Rauch, and Daniel Trefler provided comments or datasets for specific chapters. I am grateful to these individuals and many others whose input has improved this book.

FOREWORD TO THE SECOND EDITION

This book is intended for a graduate course in international trade. I assume that all readers have completed graduate courses in microeconomics and econometrics. My goal is to bring the reader from that common point up to the most recent research in international trade, in both theory and empirical work. This book is not intended to be difficult, and the mathematics used should be accessible to any graduate student. The material covered will give the reader the skills needed to understand the latest articles and working papers in the field.

The first edition of this textbook was published in 2004, and I am gratified that it has been used in many graduate courses across the United States and around the world. The goals for this second edition are the same as for the first: to provide an accessible treatment of the theory and empirical applications in international trade, thereby bringing the reader up to the frontier of research. But this second edition has another goal, too, and that is to incorporate the most important new material that has appeared in the ten years since the first edition. Foremost among that research are the monopolistic competition model of international trade with heterogeneous firms, due to Melitz (2003), and the Ricardian model of international trade with heterogeneous productivities across countries, due to Eaton and Kortum (2002). While these articles were published prior to the 2004 first edition of this book, they did not appear there other than as references for further reading. That shortcoming is corrected in this second edition.

Because the work of Eaton and Kortum (2002) builds on the Ricardian model with a continuum of goods, we present that model at the end of chapter 3 (drawing upon the presentation in Matsuyama 2008). Then the new chapter 6 has been added, dealing with monopolistic competition and the gravity equation with heterogeneous firms. The presentation in that chapter owes a great deal to articles by Melitz and Redding (2014a) and by Head and Mayer (2014) in the new *Handbook of International Economics, Volume 4*. Melitz and Redding introduce some simplifications to the original Melitz (2003) model and a method of solving it that I rely on in chapter 6, while Head and Mayer show how the gravity equation can be obtained quite generally from a wide range of models, including the Eaton-Kortum model, which is also introduced in chapter 6.

Besides the new chapter 6, other chapters have been rewritten extensively. Chapter 4, dealing with trade in intermediate inputs and wages, now draws on the presentation in my Ohlin Lecture (Feenstra 2010a) and incorporates the work by Grossman and Rossi-Hansberg (2008) on trade in tasks. Chapter 5, focusing on monopolistic competition under homogeneous firms, now includes material on measuring the gains from product variety drawing on Feenstra (1994, 2010b). Chapter 8, which discusses import tariffs and dumping, includes new material on the pass-through of tariffs (or exchange rates) using functional forms that lead to exact expressions and also recent empirical work in this area. Chapter 9, an inquiry into import quotas and export subsidies, naturally lends itself to a fuller treatment of the quality of traded goods, drawing on Feenstra and Romalis (2014). Chapter

10, on the political economy of trade policy, now includes the work of Ossa (2011). Chapter 11, dealing with endogenous growth, includes a brief exposition of growth with heterogeneous firms due to Sampson (2016). And chapter 12, an exploration of multinational corporations and the organization of the firm, has also been rewritten extensively, drawing upon Antràs (2003) and Antràs and Helpman (2004). Much of that material is covered in more depth in the volume by Antràs (2015). Throughout this second edition, I have benefited greatly from researchers providing the surveys and articles that I have incorporated here, for which I offer my thanks.

As before, an instructor's manual that accompanies this book provides solutions to the problems at the end of the chapters.¹ In addition, I have included empirical exercises that replicate the results in some chapters. Completing all of these could be the topic for a second course, but even in a first course there will be a payoff to trying some of the exercises. The data and programs for these can be found on my home page at www.robertfeenstra.info.

The notational conventions from the first edition have been retained in the second edition. I consistently use *subscripts* to refer to goods or factors, whereas *superscripts* refer to consumers or countries. In general, then, *subscripts refer to commodities and superscripts refer to agents*. The index used (h, i, j, k, l, m , or n) will depend on the context. The symbol “ c ” is used for both costs and consumption, though in some chapters I instead use “ $d(p)$ ” for consumption to avoid confusion. The output of firms is consistently denoted by “ y ” and exports are denoted by “ x ”, though in some cases “ x ” denotes inputs. Uppercase letters are used in some cases to denote vectors or matrixes, and in other cases to denote the number of goods (N), factors (M), households (H), or countries (C), and sporadically elsewhere. The symbols α and β are used generically for intercept and slope coefficients, including fixed and marginal labor costs, except that with heterogeneous firms we use the symbol φ for the output produced by one unit of labor and $1/\varphi$ for marginal labor costs.

The contents of several chapters included here have been previously published. Chapter 4 draws on material from my book *Offshoring in the Global Economy: Theory and Evidence* (MIT Press, 2010a) and chapter 5 draws upon the *Scottish Journal of Political Economy* (Feenstra 2002). Some material from chapters 8–10 has appeared in articles published in the *Journal of International Economics* and the *Quarterly Journal of Economics*, and material from chapter 11 has appeared in the *Journal of Development Economics* and in the *American Economic Review*.

¹ Faculty wishing to obtain the instructors manual should contact Princeton University Press.

ADVANCED INTERNATIONAL TRADE

Preliminaries: Two-Sector Models

We begin our study of international trade with the classic Ricardian model, which has two goods and one factor (labor). The Ricardian model introduces us to the idea that technological differences across countries matter. In comparison, the Heckscher-Ohlin model dispenses with the notion of technological differences and instead shows how *factor endowments* form the basis for trade. While this may be fine in theory, the model performs very poorly in practice: as we show in the next chapter, the Heckscher-Ohlin model is hopelessly inadequate as an explanation for historical or modern trade patterns unless we allow for technological differences across countries. For this reason, the Ricardian model is as relevant today as it has always been. Our treatment of it in this chapter is a simple review of undergraduate material, but we will present a more sophisticated version of the Ricardian model (with a continuum of goods) in chapter 3.

After reviewing the Ricardian model, we turn to the two-good, two-factor model that occupies most of this chapter and forms the basis of the Heckscher-Ohlin model. We shall suppose that the two goods are traded on international markets, but do not allow for any movements of factors across borders. This reflects the fact that the movement of labor and capital across countries is often subject to controls at the border and is generally much less free than the movement of goods. Our goal in the next chapter will be to determine the pattern of international trade between countries. In this chapter, we simplify things by focusing primarily on *one* country, treating world prices as given, and examine the properties of this two-by-two model. The student who understands all the properties of this model has already come a long way in his or her study of international trade.

RICARDIAN MODEL

Indexing goods by the subscript i , let a_i denote the labor needed per unit of production of each good at home, while a_i^* is the labor need per unit of production in the foreign country, $i = 1, 2$. The total labor force at home is L and abroad is L^* . Labor is perfectly mobile between the industries in each country, but immobile across countries. This means that both goods are produced in the home country only if the wages earned in the two industries are the same. Since the marginal product of labor in each industry is $1/a_i$, and workers are paid the value of their marginal

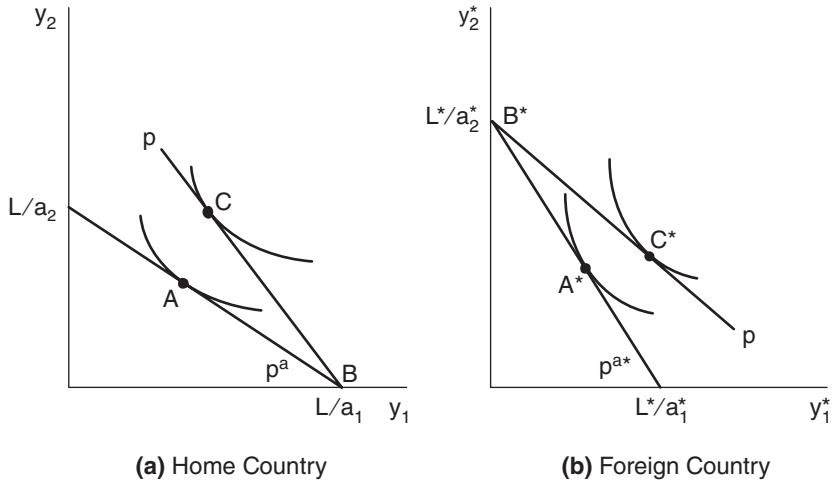


Figure 1.1

products, wages are equalized across industries if and only if $p_1/a_1 = p_2/a_2$, where p_i is the price in each industry. Letting $p = p_1/p_2$ denote the *relative price* of good 1 (using good 2 as the numeraire), this condition is $p = a_1/a_2$.

These results are illustrated in figure 1.1(a) and (b), where we graph the production possibility frontiers (PPFs) for the home and foreign countries. With all labor devoted to good i at home, it can produce L/a_i units, $i=1, 2$, so this establishes the intercepts of the PPF, and similarly for the foreign country. The slope of the PPF in each country (ignoring the negative sign) is then a_1/a_2 and a_1^*/a_2^* . Under autarky (i.e., no international trade), the equilibrium relative prices p^a and p^{a*} must equal these slopes in order to have both goods produced in both countries, as argued above. Thus, the autarky equilibrium at home and abroad might occur at points A and A^* . Suppose that the home country has a *comparative advantage* in producing good 1, meaning that $a_1/a_2 < a_1^*/a_2^*$. This implies that the home autarky relative price of good 1 is *lower* than that abroad.

Now letting the two countries engage in international trade, what is the equilibrium price p at which world demand equals world supply? To answer this, it is helpful to graph the world relative supply and demand curves, as illustrated in figure 1.2. For the relative price satisfying $p < p^a = a_1/a_2$ and $p < p^{a*} = a_1^*/a_2^*$ both countries are fully specialized in good 2 (since wages earned in that sector are higher), so the world relative supply of good 1 is zero. For $p^a < p < p^{a*}$, the home country is fully specialized in good 1 whereas the foreign country is still specialized in good 2, so that the world relative supply is $(L/a_1)/(L^*/a_2^*)$, as labeled in figure 1.2. Finally, for $p > p^a$ and $p > p^{a*}$, both countries are specialized in good 1. So we see that the world relative supply curve has a “stair-step” shape, which reflects the linearity of the PPFs.

To obtain world relative demand, let us make the simplifying assumption that tastes are identical and homothetic across the countries. Then demand will be independent of the distribution of income *across* the countries. Demand being homothetic means that *relative* demand d_1/d_2 in either country is a downward-sloping function of the relative price p , as illustrated in figure 1.2. In the case we have shown, relative demand intersects relative supply at the world price p that lies *between* p^a and p^{a*} , but this does

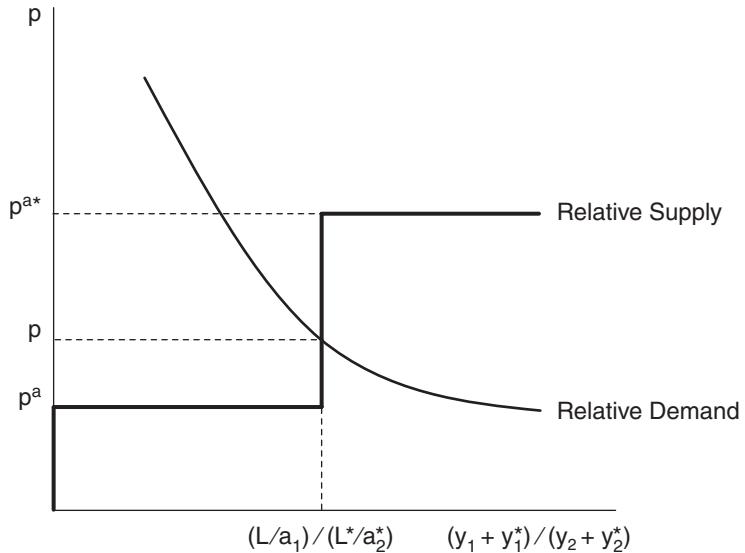


Figure 1.2

not need to occur: instead, we can have relative demand intersect one of the flat segments of relative supply, so that the equilibrium price with trade *equals* the autarky price in one country.¹

Focusing on the case where $p^a < p < p^{a*}$, we can go back to the PPF of each country and graph the production and consumption points with free trade. Since $p > p^a$, the home country is fully specialized in good 1 at point B , as illustrated in figure 1.1(a), and then trades at the relative price p to obtain consumption at point C . Conversely, since $p < p^{a*}$, the foreign country is fully specialized in the production of good 2 at point B^* in figure 1.1(b), and then trades at the relative price p to obtain consumption at point C^* . Clearly, *both* countries are better off under free trade than they were in autarky: trade has allowed them to obtain a consumption point that is above the PPF.

Notice that the home country exports good 1, which is in keeping with its comparative advantage in the production of that good, $a_1/a_2 < a_1^*/a_2^*$. Thus, *trade patterns are determined by comparative advantage*, which is a deep insight from the Ricardian model. This occurs even if one country has an *absolute disadvantage* in both goods, such as $a_1 > a_1^*$ and $a_2 > a_2^*$, so that more labor is needed per unit of production of *either* good at home than abroad. The reason that it is still possible for the home country to export is that its *wages* will adjust to reflect its productivities: under free trade, its wages are lower than those abroad.² Thus, while trade patterns in the Ricardian model are determined by *comparative advantage*, the level of wages across countries is determined by *absolute advantage*.

¹This occurs if one country is very large. Use figures 1.1 and 1.2 to show that if the home country is very large, then $p = p^a$ and the home country does not gain from trade.

²The home country exports good 1, so wages earned with free trade are $w = p/a_1$. Conversely, the foreign country exports good 2 (the numeraire), and so wages earned there are $w^* = 1/a_2^* > p/a_1^*$, where the inequality follows since $p < a_1^*/a_2^*$ in the equilibrium being considered. Then using $a_1 > a_1^*$, we obtain $w = p/a_1 < p/a_1^* < w^*$.

TWO-GOOD, TWO-FACTOR MODEL

While the Ricardian model focuses on technology, the Heckscher-Ohlin model, which we study in the next chapter, focuses on factors of production. So we now assume that there are two factor inputs—labor and capital. Restricting our attention to a single country, we will suppose that it produces two goods with the production functions $y_i = f_i(L_i, K_i)$, $i = 1, 2$, where y_i is the output produced using labor L_i and capital K_i . These production functions are assumed to be increasing, concave, and homogeneous of degree one in the inputs (L_i, K_i) .³ The last assumption means that there are *constant returns to scale* in the production of each good. This will be a maintained assumption for the next several chapters, but we should be point out that it is rather restrictive. It has long been thought that *increasing returns to scale* might be an important reason to have trade between countries: if a firm with increasing returns is able to sell in a foreign market, this expansion of output will bring a reduction in its average costs of production, which is an indication of greater efficiency. Indeed, this was a principal reason why Canada entered into a free-trade agreement with the United States in 1989: to give its firms free access to the large American market. We will return to these interesting issues in chapter 5, but for now, ignore increasing returns to scale.

We will assume that labor and capital are *fully mobile* between the two industries, so we are taking a “long run” point of view. Of course, the amount of factors employed in each industry is constrained by the endowments found in the economy. These resource constraints are stated as

$$\begin{aligned} L_1 + L_2 &\leq L, \\ K_1 + K_2 &\leq K, \end{aligned} \tag{1.1}$$

where the endowments L and K are fixed. Maximizing the amount of good 2, $y_2 = f_2(L_2, K_2)$, subject to a given amount of good 1, $y_1 = f_1(L_1, K_1)$, and the resource constraints in (1.1) give us $y_2 = h(y_1, L, K)$. The graph of y_2 as a function of y_1 is shown as the PPF in figure 1.3. As drawn, y_2 is a *concave* function of y_1 , $\partial^2 h(y_1, L, K)/\partial y_1^2 < 0$. This familiar result follows from the fact that the production functions $f_i(L_i, K_i)$ are assumed to be concave. Another way to express this is to consider all points $S = (y_1, y_2)$ that are feasible to produce given the resource constraints in (1.1). This production possibilities set S is *convex*, meaning that if $y^a = (y_1^a, y_2^a)$ and $y^b = (y_1^b, y_2^b)$ are both elements of S , then any point between them $\lambda y^a + (1 - \lambda) y^b$ is also in S , for $0 \leq \lambda \leq 1$.⁴

The production possibilities frontier summarizes the technology of the economy, but in order to determine where the economy produces on the PPF we need to add some assumptions about the market structure. We will assume perfect competition in the product markets and factor markets. Furthermore, we will suppose that product prices are given *exogenously*: we can think of these prices as established on world markets, and outside the control of the “small” country being considered.

³Students not familiar with these terms are referred to problems 1.1 and 1.2.

⁴See problems 1.1 and 1.3 to prove the convexity of the production possibilities set, and to establish its slope.

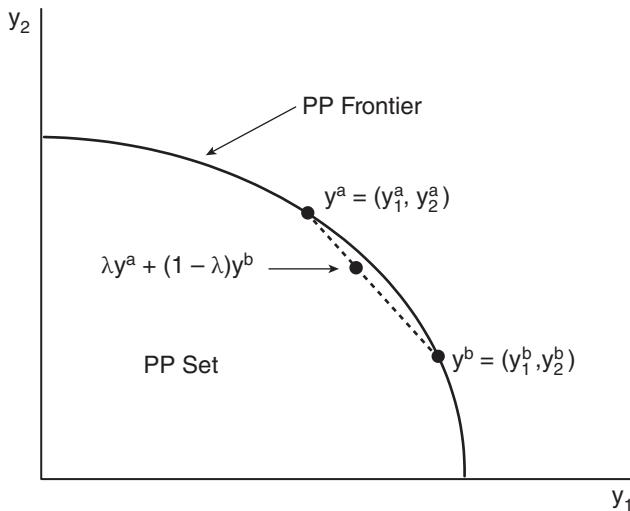


Figure 1.3

GDP FUNCTION

With the assumption of perfect competition, the amounts produced in each industry will maximize gross domestic product (GDP) for the economy: this is Adam Smith's "invisible hand" in action. That is, the industry outputs of the competitive economy will be chosen to maximize GDP:

$$G(p_1, p_2, L, K) = \max_{y_1, y_2} p_1 y_1 + p_2 y_2 \quad \text{s.t. } y_2 = h(y_1, L, K). \quad (1.2)$$

To solve this problem, we can substitute the constraint into the objective function and write it as choosing y_1 to maximize $p_1 y_1 + p_2 h(y_1, L, K)$. The first-order condition for this problem is $p_1 + p_2 (\partial h / \partial y_1) = 0$, or,

$$p = \frac{p_1}{p_2} = -\frac{\partial h}{\partial y_1} = -\frac{\partial y_2}{\partial y_1}. \quad (1.3)$$

Thus, the economy will produce where the relative price of good 1, $p = p_1/p_2$, is equal to the slope of the production possibilities frontier.⁵ This is illustrated by the point A in figure 1.4, where the line tangent through point A has the slope of (negative) p . An increase in this price will raise the slope of this line, leading to a new tangency at point B. As illustrated, then, the economy will produce more of good 1 and less of good 2.

The GDP function introduced in (1.2) has many convenient properties, and we will make use of it throughout this book. To show just one property, suppose that we differentiate the GDP function with respect to the price of good i , obtaining

⁵Notice that the slope of the price line tangent to the PPF (in absolute value) equals the relative price of the good on the horizontal axis, or good 1 in figure 1.4.

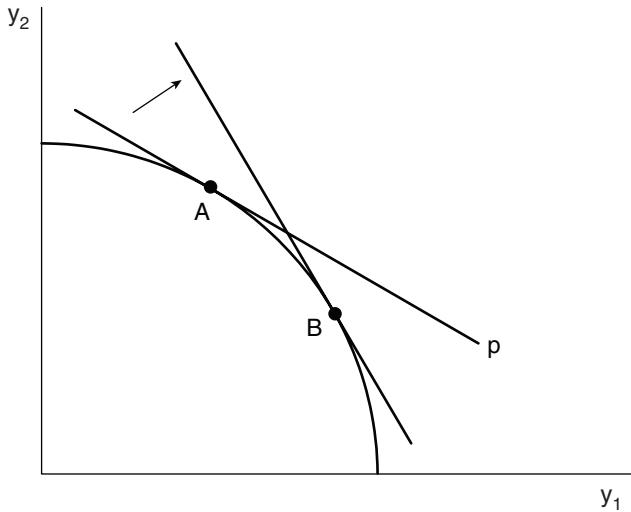


Figure 1.4

$$\frac{\partial G}{\partial p_i} = y_i + \left(p_1 \frac{\partial y_1}{\partial p_i} + p_2 \frac{\partial y_2}{\partial p_i} \right). \quad (1.4)$$

It turns out that the terms in parentheses on the right of (1.4) sum to *zero*, so that $\partial G/\partial p_i = y_i$. In other words, the derivative of the GDP function with respect to *prices* equals the *outputs* of the economy.

The fact that the terms in parentheses sum to zero is an application of the “envelope theorem,” which states that when we differentiate a function that has been maximized (such as GDP) with respect to an exogenous variable (such as p_i), then we can *ignore* the changes in the endogenous variables (y_1 and y_2) in this derivative. To prove that these terms sum to zero, totally differentiate $y_2 = h(y_1, L, K)$ with respect to y_1 and y_2 and use (1.3) to obtain $p_1 dy_1 = -p_2 dy_2$, or $p_1 dy_1 + p_2 dy_2 = 0$. This equality must hold for any small movement in y_1 and y_2 around the PPF, and in particular, for the small movement in outputs *induced* by the change in p_i . In other words, $p_1(\partial y_1/\partial p_i) + p_2(\partial y_2/\partial p_i) = 0$, so the terms in parentheses on the right of (1.4) vanish and it follows that $\partial G/\partial p_i = y_i$.⁶

EQUILIBRIUM CONDITIONS

We now want to state succinctly the equilibrium conditions to determine factor prices and outputs. It will be convenient to work with the *unit-cost functions* that are dual to the production functions $f_i(L_i, K_i)$. These are defined by

$$c_i(w, r) = \min_{L_i, K_i \geq 0} \{wL_i + rK_i \mid f_i(L_i, K_i) \geq 1\}. \quad (1.5)$$

⁶Other convenient properties of the GDP function are explored in problem 1.4.

In words, $c_i(w, r)$ is the minimum cost to produce one unit of output. Because of our assumption of constant returns to scale, these unit-costs are equal to both marginal costs and average costs. It is easily demonstrated that the unit-cost functions $c_i(w, r)$ are non-decreasing and concave in (w, r) . We will write the *solution* to the minimization in (1.5) as $c_i(w, r) = wa_{il} + ra_{ik}$, where a_{il} is optimal choice for L_i , and a_{ik} is optimal choice for K_i . It should be stressed that these optimal choices for labor and capital *depend* on the factor prices, so that they should be written in full as $a_{il}(w, r)$ and $a_{ik}(w, r)$. However, we will usually not make these arguments explicit.

Differentiating the unit-cost function with respect to the wage, we obtain

$$\frac{\partial c_i}{\partial w} = a_{il} + \left(w \frac{\partial a_{il}}{\partial w} + r \frac{\partial a_{ik}}{\partial w} \right). \quad (1.6)$$

As we found with differentiating the GDP function, it turns out that the terms in parentheses on the right of (1.6) sum to zero, which is again an application of the “envelope theorem.” It follows that the derivative of the unit-costs with respect to the wage equals the labor needed for one unit of production, $\partial c_i / \partial w = a_{il}$. Similarly, $\partial c_i / \partial r = a_{ik}$.

To prove this result, notice that the constraint in the cost-minimization problem can be written as the isoquant $f_i(a_{il}, a_{ik}) = 1$. Totally differentiate this to obtain $f_{il} da_{il} + f_{ik} da_{ik} = 0$, where $f_{il} \equiv \partial f_i / \partial L_i$ and $f_{ik} \equiv \partial f_i / \partial K_i$. This equality must hold for any small movement of labor da_{il} and capital da_{ik} around the isoquant, and in particular, for the change in labor and capital *induced* by a change in wages. Therefore, $f_{il}(\partial a_{il} / \partial w) + f_{ik}(\partial a_{ik} / \partial w) = 0$. Now multiply this through by the product price p_i , noting that $p_i f_{il} = w$ and $p_i f_{ik} = r$ from the profit-maximization conditions for a competitive firm. Then we see that the terms in parentheses on the right of (1.6) sum to zero.

The first set of equilibrium conditions for the two-by-two economy is that *profits equal zero*. This follows from free entry under perfect competition. The zero-profit conditions are stated as

$$\begin{aligned} p_1 &= c_1(w, r), \\ p_2 &= c_2(w, r). \end{aligned} \quad (1.7)$$

The second set of equilibrium conditions is full employment of both resources. These are the same as the resource constraints (1.1), except that now we express them as equalities. In addition, we will rewrite the labor and capital used in each industry in terms of the derivatives of the unit-cost function. Since $\partial c_i / \partial w = a_{il}$ is the labor used for *one unit* of production, it follows that the total labor used in $L_i = y_i a_{il}$, and similarly the total capital used is $K_i = y_i a_{ik}$. Substituting these into (1.1), the full-employment conditions for the economy are written as

$$\begin{aligned} \underbrace{\frac{a_{1L}y_1}{L_1} + \frac{a_{2L}y_2}{L_2}}_{L} &= L, \\ \underbrace{\frac{a_{1K}y_1}{K_1} + \frac{a_{2K}y_2}{K_2}}_K &= K. \end{aligned} \quad (1.8)$$

Notice that (1.7) and (1.8) together are *four* equations in *four* unknowns, namely, (w, r) and (y_1, y_2) . The parameters of these equations, p_1 , p_2 , L , and K , are given exogenously. Because the unit-cost functions are nonlinear, however, it is not enough to just count equations and unknowns: we need to study these equations in detail to

understand whether the solutions are unique and strictly positive, or not. Our task for the rest of this chapter will be to understand the properties of these equations and their solutions.

To guide us in this investigation, there are three key questions that we can ask: (1) what is the solution for factor prices? (2) if prices change, how do factor prices change? (3) if endowments change, how do outputs change? Each of these questions is taken up in the sections that follow. The methods we shall use follow the “dual” approach of Woodland (1977, 1982), Mussa (1979), and Dixit and Norman (1980).

DETERMINATION OF FACTOR PRICES

Notice that our four-equation system above can be decomposed into the zero-profit conditions as *two* equations in *two* unknowns—the wage and rental—and then the full-employment conditions, which involve both the factor prices (which affect a_{il} and a_{ik}) and the outputs. It would be especially convenient if we could uniquely solve for the factor prices from the zero-profit conditions, and then just substitute these into the full-employment conditions. This will be possible when the hypotheses of the following lemma, are satisfied.

LEMMA (FACTOR PRICE INSENSITIVITY)

So long as both goods are produced, and factor intensity reversals (FIRs) do not occur, then each price vector (p_1, p_2) corresponds to unique factor prices (w, r) .

This is a remarkable result, because it says that the factor endowments (L, K) do not matter for the determination of (w, r) . We can contrast this result with a one-sector economy, with production of $y = f(L, K)$, wages of $w = pf_L$, and diminishing marginal product $f_{LL} < 0$. In this case, any increase in the labor endowments would certainly reduce wages, so that countries with higher labor/capital endowments (L/K) would have lower wages. This is the result we normally expect. In contrast, the above lemma says that in a two-by-two economy, with a fixed product price p , it is possible for the labor force or capital stock to grow *without* affecting their factor prices! Thus, Leamer (1995) refers to this result as “factor price insensitivity.” Our goal in this section is to prove the result and also develop the intuition for why it holds.

Two conditions must hold to obtain this result: first, that both goods are produced; and second, that factor intensity reversals (FIRs) do not occur. To understand FIRs, consider figures 1.5 and 1.6. In the first case, presented in figure 1.5, we have graphed the two zero-profit conditions, and the unit-cost lines intersect only *once*, at point A. This illustrates the lemma: given (p_1, p_2) , there is a *unique* solution for (w, r) . But another case is illustrated in figure 1.6, where the unit-cost lines intersect *twice*, at points A and B. Then there are two possible solutions for (w, r) , and the result stated in the lemma no longer holds.

The case where the unit-cost lines intersect more than once corresponds to “factor intensity reversals.” To see where this name comes from, let us compute the labor and capital requirements in the two industries. We have already shown that a_{il} and a_{ik} are the derivatives of the unit-cost function with respect to factor prices, so it follows that the vectors (a_{il}, a_{ik}) are the *gradient vectors* to the iso-cost curves for the two industries

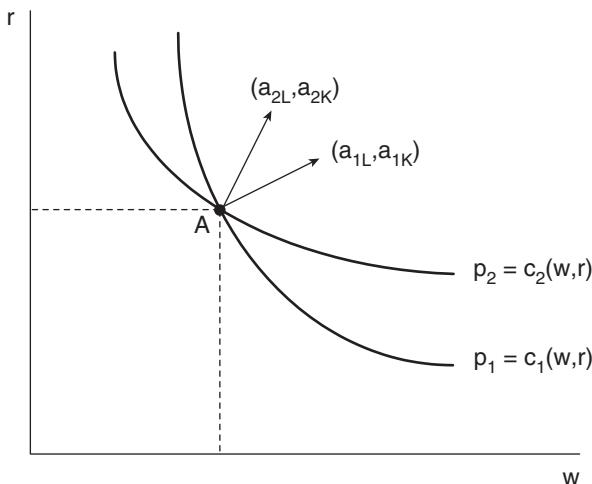


Figure 1.5

in figure 1.5. Recall from calculus that gradient vectors point in the direction of the maximum increase of the function in question. This means that they are *orthogonal* to their respective iso-cost curves, as shown by (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) at point A. Each of these vectors has slope (a_{ik}/a_{il}) , or the capital-labor ratio. It is clear from figure 1.5 that (a_{1L}, a_{1K}) has a smaller slope than (a_{2L}, a_{2K}) , which means that *industry 2 is capital intensive*, or equivalently, *industry 1 is labor intensive*.⁷

In figure 1.6, however, the situation is more complicated. Now there are two sets of gradient vectors, which we label by (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) at point A and by (b_{1L}, b_{1K}) and (b_{2L}, b_{2K}) at point B. A close inspection of the figure will reveal that industry 1 is *labor intensive* ($a_{1K}/a_{1L} < a_{2K}/a_{2L}$) at point A, but is *capital intensive* ($b_{1K}/b_{1L} > b_{2K}/b_{2L}$) at point B. This illustrates a *factor intensity reversal*, whereby the comparison of factor intensities changes at different factor prices.

While FIRs might seem like a theoretical curiosum, they are actually quite realistic. Consider the footwear industry, for example. While much of the footwear in the world is produced in developing nations, the United States retains a small number of plants. For sneakers, New Balance has a plant in Norridgewock, Maine, where employers earn about \$14 per hour.⁸ Some operate computerized equipment with up to twenty sewing machine heads running at once, while others operate automated stitchers guided by cameras, which allow one person to do the work of six. This is a far cry from the plants in Asia that produce shoes for Nike, Reebok, and other U.S. producers, using century-old technology and paying less than \$1 per hour. The technology used to make sneakers in Asia is like that of industry 1 at point A in figure 1.5, using labor-intensive

⁷Alternatively, we can totally differentiate the zero-profit conditions, holding prices fixed, to obtain $0 = a_{il}dw + a_{ik}dr$. It follows that the slope of the iso-cost curve equals $dr/dw = -a_{il}/a_{ik} = -L_i/K_i$. Thus, the slope of each iso-cost curve equals the relative demand for the factor on the horizontal axis, whereas the slope of the gradient vector (which is orthogonal to the iso-cost curve) equals the relative demand for the factor on the vertical axis.

⁸The material that follows is drawn from Aaron Bernstein, "Low-Skilled Jobs: Do They Have to Move?" *Business Week*, February 26, 2001, pp. 94–95.

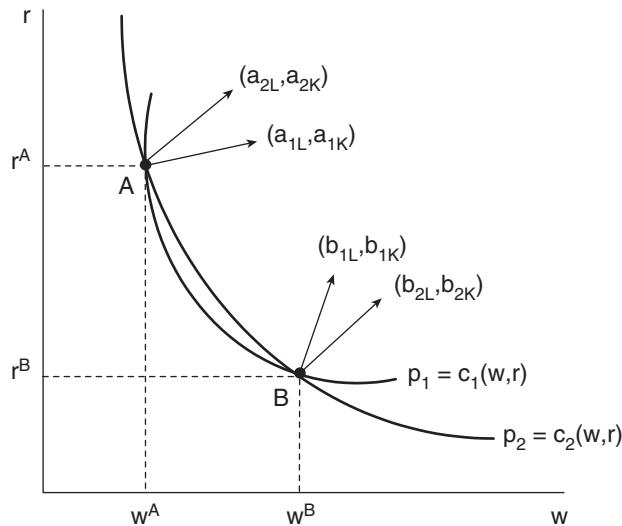


Figure 1.6

technology and paying low wages w^A , while industry 1 in the United States is at point B , paying higher wages w^B and using a capital-intensive technology.

As suggested by this discussion, when there are two possible solutions for the factor prices such as points A and B in figure 1.6, then some countries can be at one equilibrium and others countries at the other. How do we know which country is where? This is a question that we will answer at the end of the chapter, where we will argue that a *labor-abundant* country will likely be at equilibrium A of figure 1.6, with a low wage and high rental on capital, whereas a *capital-abundant* country will be at equilibrium B , with a high wage and low rental. Generally, to determine the factor prices in each country we will need to examine its full-employment conditions in addition to the zero-profit conditions.

Let us conclude this section by returning to the simple case of no FIR, in which the lemma stated above applies. What are the implications of this result for the determination of factor prices under free trade? To answer this question, let us sketch out some of the assumptions of the Heckscher-Ohlin model, which we will study in more detail in the next chapter. We assume that there are two countries, with identical technologies but different factor endowments. We continue to assume that labor and capital are the two factors of production, so that under free trade the equilibrium conditions (1.7) and (1.8) apply in *each* country with the *same* product prices (p_1, p_2). We can draw figure 1.5 for each country, and in the absence of FIR, this *uniquely* determines the factor prices in each countries. In other words, the wage and rental determined by figure 1.5 are *identical* across the two countries. We have therefore proved the factor price equalization (FPE) theorem, which is stated as follows.

FACTOR PRICE EQUALIZATION THEOREM (SAMUELSON 1949)

Suppose that two countries are engaged in free trade, having identical technologies but different factor endowments. If both countries produce both goods and FIRs do not occur, then the factor prices (w, r) are equalized across the countries.

The FPE theorem is a remarkable result because it says that *trade in goods* has the ability to equalize factor prices: in this sense, trade in goods is a “perfect substitute” for trade in factors. We can again contrast this result with that obtained from a one-sector economy in both countries. In that case, equalization of the product price through trade would certainly not equalize factor prices: the labor-abundant country would be paying a lower wage. Why does this outcome *not occur* when there are two sectors? The answer is that the labor-abundant country can *produce more of, and export*, the labor-intensive good. In that way it can fully employ its labor while still paying the same wages as a capital-abundant country. In the two-by-two model, the opportunity to disproportionately produce more of one good than the other, while exporting the amounts not consumed at home, is what allows factor price equalization to occur. This intuition will become even clearer as we continue to study the Heckscher-Ohlin model in the next chapter.

CHANGE IN PRODUCT PRICES

Let us move on now to the second of our key questions of the two-by-two model: if the product prices change, how will the factor prices change? To answer this, we perform comparative statics on the zero-profit conditions (1.7). Totally differentiating these conditions, we obtain

$$dp_i = a_{il}dw + a_{ik}dr \Rightarrow \frac{dp_i}{p_i} = \frac{wa_{il}}{c_i} \frac{dw}{w} + \frac{ra_{ik}}{c_i} \frac{dr}{r}, i = 1, 2. \quad (1.9)$$

The second equation is obtained by multiplying and dividing like terms, and noting that $p_i = c_i(w, r)$. The advantage of this approach is that it allows us to express the variables in terms of *percentage changes*, such as $d\ln w = dw/w$, as well as *cost-shares*. Specifically, let $\theta_{il} = wa_{il}/c_i$ denote the cost-share of labor in industry i , while $\theta_{ik} = ra_{ik}/c_i$ denotes the cost-share of capital. The fact that costs equal $c_i = wa_{il} + ra_{ik}$ ensures that the shares sum to unity, $\theta_{il} + \theta_{ik} = 1$. In addition, denote the percentage changes by $dw/w = \hat{w}$ and $dr/r = \hat{r}$. Then (1.9) can be re-written as

$$\hat{p}_i = \theta_{il}\hat{w} + \theta_{ik}\hat{r}, i = 1, 2. \quad (1.9')$$

Expressing the equations using these cost-shares and percentage changes follows Jones (1965) and is referred to as the “Jones algebra.” This system of equations can be written in matrix form and solved as

$$\begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix} = \begin{pmatrix} \theta_{1L} & \theta_{1K} \\ \theta_{2L} & \theta_{2K} \end{pmatrix} \begin{pmatrix} \hat{w} \\ \hat{r} \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{w} \\ \hat{r} \end{pmatrix} = \frac{1}{|\theta|} \begin{pmatrix} \theta_{2K} & -\theta_{1K} \\ -\theta_{2L} & \theta_{1L} \end{pmatrix} \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix}, \quad (1.10)$$

where $|\theta|$ denotes the determinant of the two-by-two matrix on the left. This determinant can be expressed as

$$\begin{aligned} |\theta| &= \theta_{1L}\theta_{2K} - \theta_{1K}\theta_{2L} \\ &= \theta_{1L}(1 - \theta_{2L}) - (1 - \theta_{1L})\theta_{2L} \\ &= \theta_{1L} - \theta_{2L} = \theta_{2K} - \theta_{1K} \end{aligned} \quad (1.11)$$

where we have repeatedly made use of the fact that $\theta_{il} + \theta_{ik} = 1$.

In order to fix ideas, let us assume henceforth that *industry 1 is labor intensive*. This implies that its labor cost-share in industry 1 exceeds that in industry 2, $\theta_{1L} - \theta_{2L} > 0$, so that $|\theta| > 0$ in (1.11).⁹ Furthermore, suppose that the relative price of good 1 *increases*, so that $\hat{p} = \hat{p}_1 - \hat{p}_2 > 0$. Then we can solve for the change in factor prices from (1.10) and (1.11) as

$$\hat{w} = \frac{\theta_{2K}\hat{p}_1 - \theta_{1K}\hat{p}_2}{|\theta|} = \frac{(\theta_{2K} - \theta_{1K})\hat{p}_1 + \theta_{1K}(\hat{p}_1 - \hat{p}_2)}{(\theta_{2K} - \theta_{1K})} > \hat{p}_1, \quad (1.12a)$$

since $\hat{p}_1 - \hat{p}_2 > 0$, and,

$$\hat{r} = \frac{\theta_{1L}\hat{p}_2 - \theta_{2L}\hat{p}_1}{|\theta|} = \frac{(\theta_{1L} - \theta_{2L})\hat{p}_2 - \theta_{2L}(\hat{p}_1 - \hat{p}_2)}{(\theta_{1L} - \theta_{2L})} < \hat{p}_2, \quad (1.12b)$$

since $\hat{p}_1 - \hat{p}_2 > 0$.

From the result in (1.12a), we see that the wage increases *by more* than the price of good 1, $\hat{w} > \hat{p}_1 > \hat{p}_2$. This means that workers can afford to buy more of good 1 (w/p_1 has gone up), as well as more of good 2 (w/p_2 has gone up). When labor can buy more of *both goods* in this fashion, we say that the *real wage* has increased. Looking at the rental on capital in (1.12b), we see that the rental r changes by *less than* the price of good 2. It follows that capital-owner can afford less of good 2 (r/p_2 has gone down), and also less of good 1 (r/p_1 has gone down). Thus the *real return to capital* has fallen. We can summarize these results with the following theorem.

STOLPER-SAMUELSON (1941) THEOREM

An increase in the relative price of a good will increase the real return to the factor used intensively in that good, and reduce the real return to the other factor.

To develop the intuition for this result, let us go back to the differentiated zero-profit conditions in (1.9'). Since the cost-shares add up to unity in each industry, we see from equation (1.9') that \hat{p}_i is a *weighted average* of the factor price changes \hat{w} and \hat{r} . This implies that \hat{p}_i necessarily lies in between \hat{w} and \hat{r} . Putting these together with our assumption that $\hat{p}_1 - \hat{p}_2 > 0$, it is therefore clear that

$$\hat{w} > \hat{p}_1 > \hat{p}_2 > \hat{r}. \quad (1.13)$$

Jones (1965) has called this set of inequalities the “magnification effect”: they show that any change in the product price has a *magnified effect* on the factor prices. This is an extremely important result. Whether we think of the product price change as due to export opportunities for a country (the export price goes up), or due to lowering import tariffs (so the import price goes down), the magnification effect says that there will be both gainers and losers due to this change. Even though we will argue in chapter 6 that there are gains from trade in some overall sense, it is still the case that trade opportunities have strong *distributional consequences*, making some people worse off and some better off!

⁹ As an exercise, show that $L_1/K_1 > L_2/K_2 \Leftrightarrow \theta_{1L} > \theta_{2L}$. This is done by multiplying the numerator and denominator on both sides of the first inequality by like terms, so as to convert it into cost-shares.

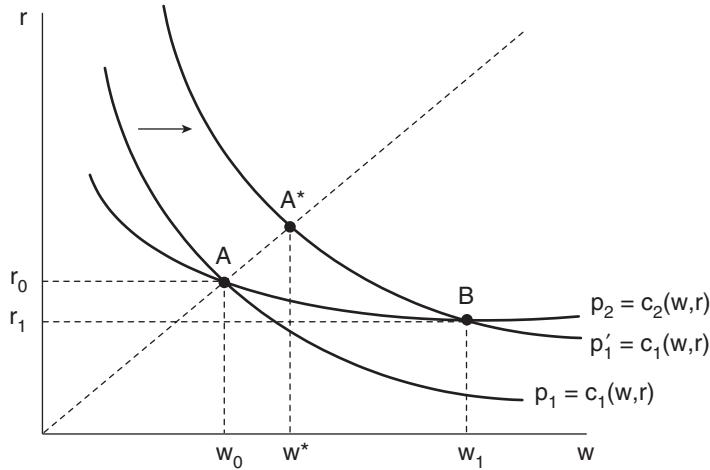


Figure 1.7

We conclude this section by illustrating the Stolper-Samuelson theorem in figure 1.7. We begin with an initial factor price equilibrium given by point A , where industry 1 is labor intensive. An increase in the price of that industry will shift out the iso-cost curve, and as illustrated, move the equilibrium to point B . It is clear that the wage has gone up, from w_0 to w_1 , and the rental has declined, from r_0 to r_1 . Can we be sure that the wage has increased in percentage terms *by more* than the relative price of good 1? The answer is yes, as can be seen by drawing a ray from the origin through the point A . Because the unit-cost functions are homogeneous of degree one in factor prices, moving along this ray increases p and (w, r) in the same proportion. Thus, at the point A^* , the increase in the wage exactly matched the percentage change in the price p_1 . But it is clear that the equilibrium wage increases *by more*, $w_1 > w^*$, so the percentage increase in the wage *exceeds* that of the product price, which is the Stolper-Samuelson result.

CHANGES IN ENDOWMENTS

We turn now to the third key question: if endowments change, how do the industry outputs change? To answer this, we hold the product prices *fixed* and totally differentiate the full-employment conditions (1.8) to obtain

$$\begin{aligned} a_{1L} dy_1 + a_{2L} dy_2 &= dL, \\ a_{1K} dy_1 + a_{2K} dy_2 &= dK. \end{aligned} \tag{1.14}$$

Notice that the a_{ij} coefficients *do not* change, despite the fact that they are functions of the factor prices (w, r) . These coefficients are fixed because p_1 and p_2 do not change, so from our earlier lemma, the factor prices are also fixed.

By rewriting the equations in (1.14) using the “Jones algebra,” we obtain

$$\begin{aligned} \frac{y_1 a_{1L}}{L} \frac{dy_1}{y_1} + \frac{y_2 a_{2L}}{L} \frac{dy_2}{y_2} &= \frac{dL}{L} \Rightarrow \lambda_{1L} \hat{y}_1 + \lambda_{2L} \hat{y}_2 = \hat{L} \\ \frac{y_1 a_{1K}}{K} \frac{dy_1}{y_1} + \frac{y_2 a_{2K}}{K} \frac{dy_2}{y_2} &= \frac{dK}{K} \Rightarrow \lambda_{1K} \hat{y}_1 + \lambda_{2K} \hat{y}_2 = \hat{K}. \end{aligned} \tag{1.14'}$$

To move from the first set of equations to the second, we denote the percentage changes $dy_1/y_1 = \hat{y}_1$, and likewise for all the other variables. In addition, we define $\lambda_{il} \equiv (y_i a_{il}/L) = (L_i/L)$, which measures the *fraction of the labor force employed in industry i*, where $\lambda_{1L} + \lambda_{2L} = 1$. We define λ_{ik} analogously as the fraction of the capital stock employed in industry i .

This system of equations is written in matrix form and solved as

$$\begin{bmatrix} \lambda_{1L} & \lambda_{2L} \\ \lambda_{1K} & \lambda_{2K} \end{bmatrix} \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \begin{pmatrix} \hat{L} \\ \hat{K} \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \frac{1}{|\lambda|} \begin{bmatrix} \lambda_{2K} & -\lambda_{2L} \\ -\lambda_{1K} & \lambda_{1L} \end{bmatrix} \begin{pmatrix} \hat{L} \\ \hat{K} \end{pmatrix}, \quad (1.15)$$

where $|\lambda|$ denotes the determinant of the two-by-two matrix on the left, which is simplified as

$$\begin{aligned} |\lambda| &= \lambda_{1L}\lambda_{2K} - \lambda_{2L}\lambda_{1K} \\ &= \lambda_{1L}(1 - \lambda_{1K}) - (1 - \lambda_{1L})\lambda_{1K} \\ &= \lambda_{1L} - \lambda_{1K} = \lambda_{2K} - \lambda_{2L}, \end{aligned} \quad (1.16)$$

where we have repeatedly made use of the fact that $\lambda_{1L} + \lambda_{2L} = 1$ and $\lambda_{1K} + \lambda_{2K} = 1$.

Recall that we assumed *industry 1 to be labor intensive*. This implies that the share of the labor force employed in industry 1 exceeds the share of the capital stock used there, $\lambda_{1L} - \lambda_{1K} > 0$, so that $|\lambda| > 0$ in (1.16).¹⁰ Suppose further that the endowments of labor is increasing, while the endowments of capital remains fixed such that $\hat{L} > 0$, and $\hat{K} = 0$. Then we can solve for the change in outputs from (1.15)–(1.16) as

$$\hat{y}_1 = \frac{\lambda_{2K}}{(\lambda_{2K} - \lambda_{2L})} \hat{L} > \hat{L} > 0 \quad \text{and} \quad \hat{y}_2 = \frac{-\lambda_{1K}}{|\lambda|} \hat{L} < 0. \quad (1.17)$$

From (1.17), we see that the output of the labor-intensive industry 1 expands, whereas the output of industry 2 contracts. We have therefore established the Rybczynski theorem.

RYBCZYNSKI (1955) THEOREM

An increase in a factor endowment will increase the output of the industry using it intensively, and decrease the output of the other industry.

To develop the intuition for this result, let us write the full-employment conditions in vector notation as:

$$\begin{pmatrix} a_{1L} \\ a_{1K} \end{pmatrix} y_1 + \begin{pmatrix} a_{2L} \\ a_{2K} \end{pmatrix} y_2 = \begin{pmatrix} L \\ K \end{pmatrix}. \quad (1.8')$$

We have already illustrated the gradient vectors (a_{il} , a_{ik}) to the iso-cost curves in figure 1.5 (with not FIR). Now let us take these vectors and regraph them, in figure

¹⁰ As an exercise, show that $L_1/K_1 > L/K > L_2/K_2 \Leftrightarrow \lambda_{1L} > \lambda_{1K}$ and $\lambda_{2K} > \lambda_{2L}$.

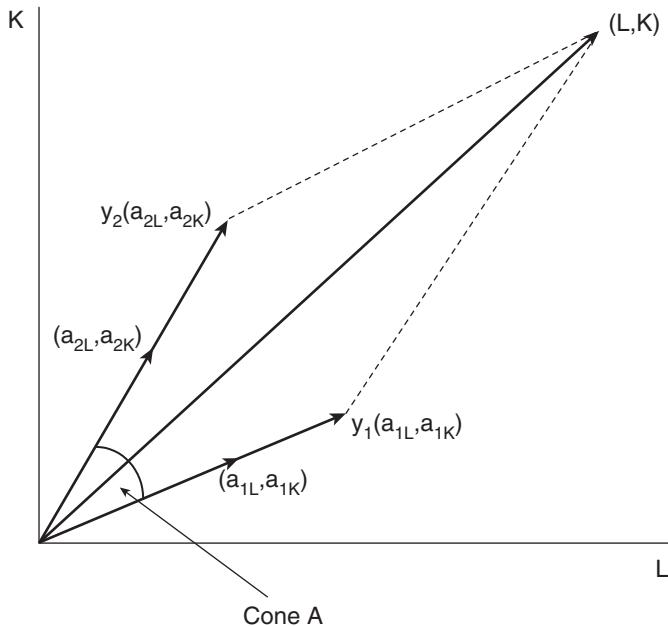


Figure 1.8

1.8. Multiplying each of these by the output of their respective industries, we obtain the total labor and capital demands $y_1(a_{1L}, a_{1K})$ and $y_2(a_{2L}, a_{2K})$. Summing these as in (1.8') we obtain the labor and capital endowments (L, K) . But this exercise can also be performed in reverse: for any endowment vector (L, K) , there will be a *unique* value for the outputs (y_1, y_2) such that when (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) are multiplied by these amounts, they will sum to the endowments.

How can we be sure that the outputs obtained from (1.8') are positive? It is clear from figure 1.8 that the outputs in both industries will be positive if and only if the endowment vector (L, K) lies *in between* the factor requirement vectors (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) . For this reason, the space spanned by these two vectors is called a “cone of diversification,” which we label by cone A in figure 1.8. In contrast, if the endowment vector (L, K) lies *outside* of this cone, then it is *impossible* to add together any positive multiples of the vectors (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) and arrive at the endowment vector. So if (L, K) lies outside of the cone of diversification, then it must be that only *one good* is produced. At the end of the chapter, we will show how to determine which good it is.¹¹ For now, we should just recognize that when only one good is produced, the factor prices are determined by the marginal products of labor and capital as in the one-sector model, and will certainly depend on the factor endowments.

Now suppose that the labor endowment increases to $L' > L$, with no change in the capital endowment, as shown in figure 1.9. Starting from the endowments (L', K) , the *only* way to add up multiples of (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) and obtain the endowments is to *reduce* the output of industry 2 to y'_2 , and *increase* the output of industry 1 to y'_1 . This means that not only does industry 1 absorb the entire amount of the extra labor

¹¹ See problem 1.5.

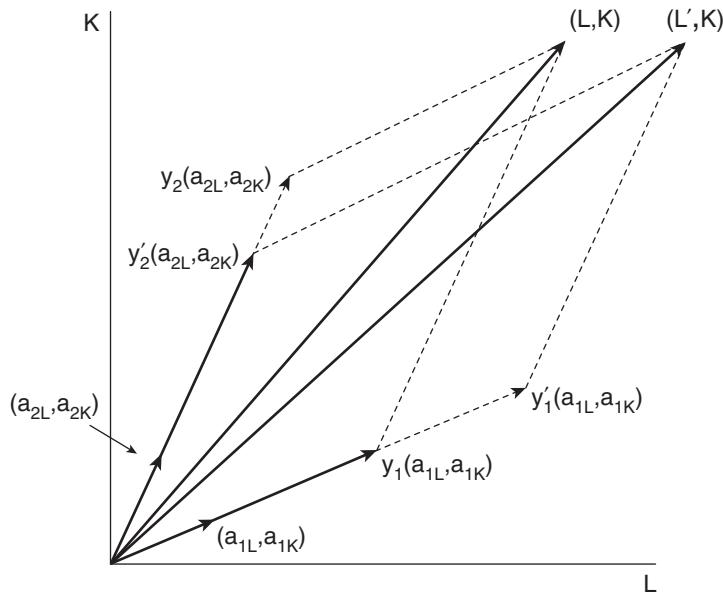


Figure 1.9

endowment, but it also absorbs further labor and capital from industry 2 so that its ultimate labor/capital ratio is unchanged from before. The labor/capital ratio in industry 2 is also unchanged, and this is what permits both industries to pay exactly the same factor prices as they did before the change in endowments.

There are many examples of the Rybczynski theorem in practice, but perhaps the most commonly cited is what is called the “Dutch Disease.”¹² This refers to the discovery of oil off the coast of the Netherlands, which led to an increase in industries making use of this resource. (Shell Oil, one of the world’s largest producers of petroleum products, is a Dutch company.) At the same time, however, other “traditional” export industries of the Netherlands contracted. This occurred because resources were attracted away from these industries and into those that were intensive in oil, as the Rybczynski theorem would predict.

We have now answered the three questions raised earlier in the chapter: how are factor prices determined; how do changes in product prices affect factor prices; and how do changes in endowments affect outputs? But in answering all of these, we have relied on the assumptions that *both goods are produced*, and also that factor intensity reversals do not occur, as was stated explicitly in the FPE theorem. In the remainder of this chapter we need to investigate both of these assumptions, to understand either when they will hold or the consequences of their not holding.

We begin by tracing through the changes in the outputs induced by changes in endowments, along the equilibrium of the production possibility frontier. As the labor endowment grows in figure 1.9, the PPF will shift out. This is shown in figure 1.10, where the outputs will shift from point A to point A' with an increase of good 1 and reduction of good 2, at the unchanged price p . As the endowment of labor rises, we

¹²See, for example, Corden and Neary (1982) and Jones, Neary, and Ruane (1987).

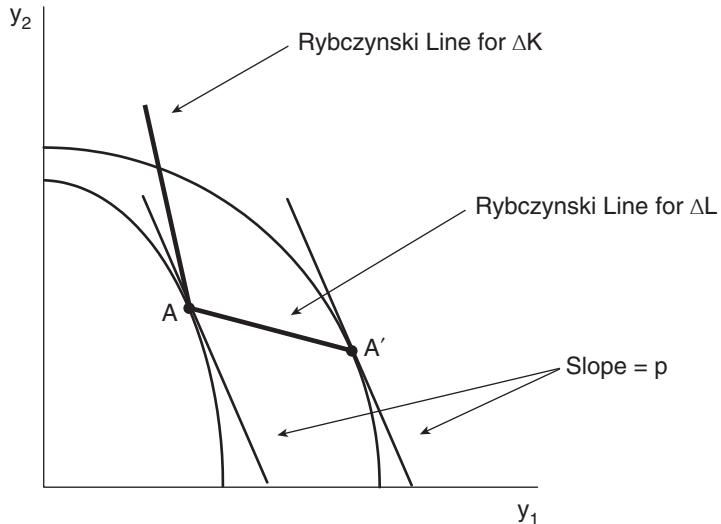


Figure 1.10

can join up all points such as A and A' where the slopes of the PPFs are equal. These form a downward-sloping line, which we will call the Rybczynski line for changes in labor (ΔL). The Rybczynski line for ΔL indicates how outputs change as labor endowment expands.

Of course, there is also a Rybczynski line for ΔK , which indicates how the outputs change as the capital endowment grows: this would lead to an increase in the output of good 2, and reduction in the output of good 1. As drawn, both of the Rybczynski lines are illustrated as *straight* lines: can we be sure that this is the case? The answer is yes: the fact that the product prices are fixed along a Rybczynski line, implying that factor prices are also fixed, ensures that these are straight lines. To see this, we can easily calculate their slopes by differentiating the full-employment conditions (1.8). To compute the slope of the Rybczynski line for ΔL , it is convenient to work with the full-employment condition for *capital*, since that endowment does not change. Total differentiating (1.8) for capital gives

$$a_{1K}y_1 + a_{2K}y_2 = K \Rightarrow a_{1K}dy_1 + a_{2K}dy_2 = 0 \Rightarrow \frac{dy_2}{dy_1} = -\frac{a_{1K}}{a_{2K}}. \quad (1.18)$$

Thus, the slope of the Rybczynski line for ΔL is the negative of the ratio of capital/ output in the two industries, which is constant for fixed prices. This proves that the Rybczynski lines are indeed straight.

If we continue to increase the labor endowment, outputs will move downward on the Rybczynski line for ΔL in figure 1.10, until this line hits the y_1 axis. At this point the economy is fully specialized in good 1. In terms of figure 1.9, the vector of endowments (L, K) is coincident with the vector of factor requirements (a_{1L}, a_{1K}) in industry 1. For further increases in the labor endowment, the Rybczynski line for ΔL then moves right along the y_1 axis in figure 1.10, indicating that the economy remains specialized in good 1. This corresponds to the vector of endowments (L, K)

lying outside and below the cone of diversification in figure 1.9. With the economy fully specialized in good 1, factor prices are determined by the marginal products of labor and capital in that good, and the earlier “factor price insensitivity” lemma no longer applies.

FACTOR PRICE EQUALIZATION REVISITED

Our finding that the economy produces both goods whenever the factor endowments remain *inside* the cone of diversification allows us to investigate the FPE theorem more carefully. Let us continue to assume that there are no FIRs, but now rather than *assuming* that both goods are produced in both countries, we will instead derive this as an *outcome* from the factor endowments in each country. To do so, we engage in a thought experiment posed by Samuelson (1949) and further developed by Dixit and Norman (1980).

Initially, suppose that labor and capital are *free to move* between the two countries until their factor prices are equalized. Then all that matters for factor prices are the *world* endowments of labor and capital, and these are shown as the length of the horizontal and vertical axis in figure 1.11. The amounts of labor and capital choosing to reside at home are measured relative to the origin 0, while the amounts choosing to reside in the foreign country are measured relative to the origin 0^* ; suppose that this allocation is at point B . Given the world endowments, we establish equilibrium prices for goods and factors in this “integrated world equilibrium.” The factor prices determine the demand for labor and capital in each industry (assuming no FIR), and using

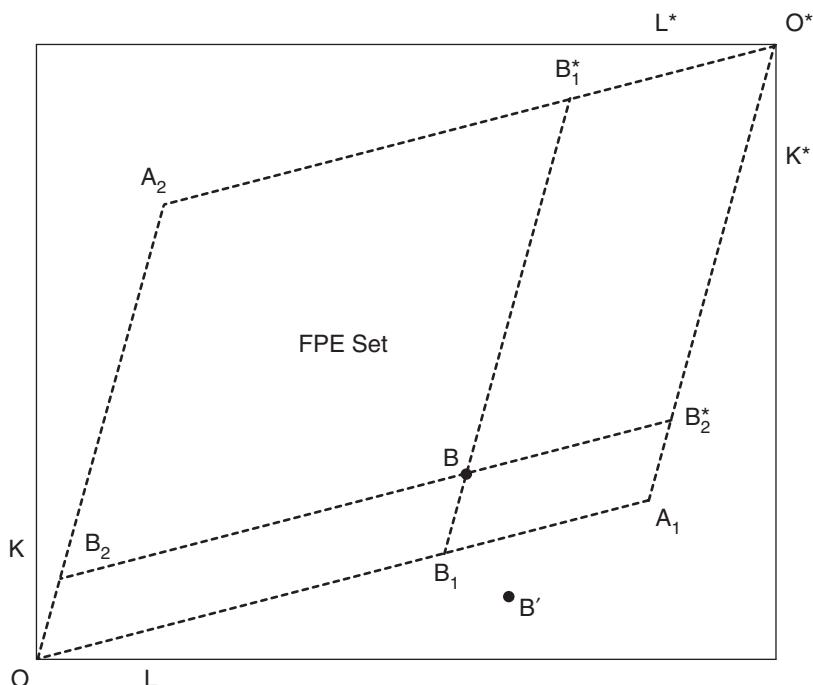


Figure 1.11

these, we can construct the diversification cone (since factor prices are the same across countries, then the diversification cone is also the same). Let us plot the diversification cone relative to the home origin 0 , and again relative to the foreign origin 0^* . These cones form the parallelogram $0A_10^*A_2$.

For later purposes, it is useful to identify precisely the points A_1 and A_2 on the vertices of this parallelogram. The vectors $0A_i$ and 0^*A_i are proportional to (a_{iL}, a_{iK}) , the amount of labor and capital used to produce one unit of good i in each country. Multiplying (a_{iL}, a_{iK}) by world demand for good i , D_i^w , we then obtain the *total* labor and capital used to produce that good, so that $A_i = (a_{iL}, a_{iK})D_i^w$. Summing these gives the total labor and capital used in world demand, which equals the labor and capital used in world production, or world endowments.

Now we ask whether we can achieve exactly the same world production and equilibrium prices as in this “integrated world equilibrium,” but *without* labor and capital mobility. Suppose there is some allocation of labor and capital endowments across the countries, such as point B . Then can we produce the same amount of each good as in the “integrated world equilibrium”? The answer is clearly yes: with labor and capital in each country at point B , we could devote $0B_1$ of resources to good 1 and $0B_2$ to good 2 at home, while devoting $0^*B_1^*$ to good 1 and $0^*B_2^*$ toward good 2 abroad. This will ensure that the same amount of labor and capital worldwide is devoted to each good as in the “integrated world equilibrium,” so that production and equilibrium prices must be the same as before. Thus, we have achieved the same equilibrium but without factor mobility. It will become clear in the next chapter that there is still *trade in goods* going on to satisfy the demands in each country.

More generally, for *any allocation* of labor and capital within the parallelogram $0A_10^*A_2$ both countries remain diversified (producing both goods), and we can achieve the same equilibrium prices as in the “integrated world economy.” It follows that factor prices *remain equalized across countries* for allocations of labor and capital within the parallelogram $0A_10^*A_2$, which is referred to as the *factor price equalization (FPE) set*. The FPE set illustrates the range of labor and capital endowments between countries over which both goods are produced in both countries, so that factor price equalization is obtained. In contrast, for endowments *outside* of the FPE set such as point B' , then at least one country would have to be fully specialized in one good and FPE no longer holds.

FACTOR INTENSITY REVERSALS

We conclude this chapter by returning to a question raised earlier: when there are “factor intensity reversals” giving multiple solutions to the zero-profit conditions, how do we know which solution will prevail in each country? To answer this, it is necessary to combine the zero-profit with the full-employment conditions, as follows.

Consider the case in figure 1.6, where the zero-profit conditions allows for two solutions to the factor prices. Each of these determine the labor and capital demands shown orthogonal to the iso-cost curves, labeled as (a_{1L}, a_{1K}) and (a_{2L}, a_{2K}) , and (b_{1L}, b_{1K}) and (b_{2L}, b_{2K}) . We have redrawn these in figure 1.12, after multiplying each of them by the outputs of their respective industries. These vectors create *two* cones of diversification, labeled as cones A and B . Initially, suppose that the factor endowments for each country lie within one cone or the other (then we will consider the case where the endowments are outside both cones).

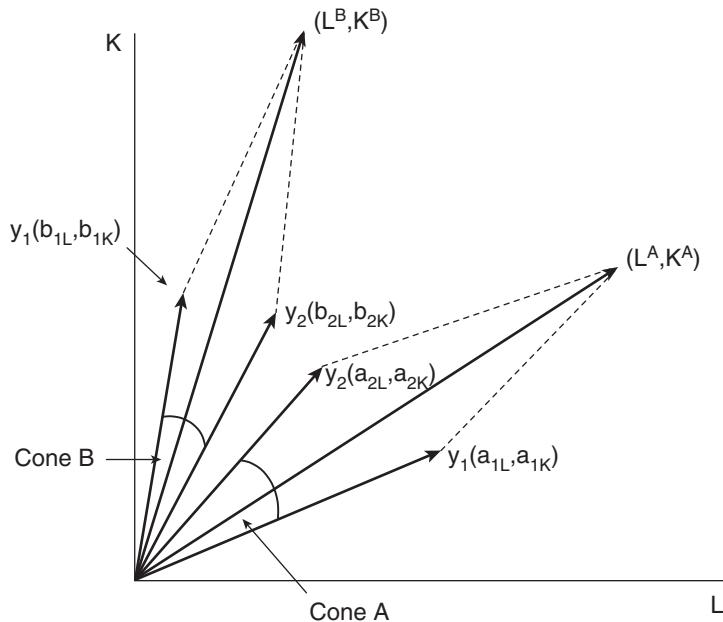


Figure 1.12

Now we can answer the question of which factor prices will apply in each country: a *labor-abundant* economy, with a high ratio of labor/capital endowments, such as (L^A, K^A) in cone A of figure 1.12, will have factor prices given by (w^A, r^A) in figure 1.6, with low wages; whereas a *capital-abundant* economy with a high ratio of capital/labor endowments such as shown by (L^B, K^B) in cone B of figure 1.12, will have factor prices given by (w^B, r^B) in figure 1.6, with high wages. Thus, factor prices depend on the endowments of the economy. A labor-abundant country such as China will pay low wages and a high rental (as in cone A), while a capital-abundant country such as the United States will have high wages and a low rental (as in cone B). Notice that we have now reintroduced a link between factor endowments and factor prices, as we argued earlier in the one-sector model: when there are FIR in the two-by-two model, factor prices vary systematically with endowments *across* the cones of diversification, even though factor prices are independent of endowments *within* each cone.

What if the endowment vector of a country does not lie in either cone? Then the country will be fully specialized in one good or the other. Generally, we can determine which good it is by tracing through how the outputs change as we move through the cones of diversification, and it turns out that outputs depend *non-monotonically* on the factor endowments.¹³ For example, textiles in South Korea or Taiwan expanded during the 1960s and 1970s, but contracted later as capital continued to grow. Despite the complexity involved, many trade economists feel that countries do in fact produce in different cones of diversification, and taking this possibility into account is a topic of research.¹⁴

¹³See problem 1.5.

¹⁴Empirical evidence on whether developed countries fit into the same cone is presented by Debaere and Demiroğlu (2003), and the presence of multiple cones is explored by Leamer (1987), Harrigan and Zakravsek (2000), and Schott (2003).

CONCLUSIONS

In this chapter we have reviewed several two-sector models: the Ricardian model, with just one factor, and the two-by-two model, with two factors, both of which are fully mobile between industries. There are other two-sector models, of course: if we add a third factor, treating capital as specific to each sector but labor as mobile, then we obtain the Ricardo-Viner or “specific-factors” model, as will be discussed in chapter 3. We will have an opportunity to make use of the two-by-two model throughout this book, and a thorough understanding of its properties—both the equations and the diagrams *labor-abundant economy*—is essential for all the material that follows.

One special feature of this chapter is the dual determination of factor prices, using the unit-cost function in the two industries. This follows the dual approach of Woodland (1977, 1982), Mussa (1979), and Dixit and Norman (1980). Samuelson (1949) uses a quite different diagrammatic approach to prove the FPE theorem. Another method that is quite commonly used is the so-called Lerner (1952) diagram, which relies on the production rather than cost functions.¹⁵ We will not use the Lerner diagram in this book, but it will be useful to understand some articles, for example, Findlay and Grubert (1959) and Deardorff (1979), so we include a discussion of it in the appendix to this chapter.

This is the only chapter where we do not present any accompanying empirical evidence. The reader should not infer from this that the two-by-two model is unrealistic: while it is usually necessary to add more goods or factors to this model before confronting it with data, the relationships between prices, outputs, and endowments that we have identified in this chapter will carry over in some form to more general settings. Evidence on the pattern of trade is presented in the next chapter, where we extend the two-by-two model by adding another country, and then many countries, trading with each other. We also allow for many goods and factors, but for the most part restrict attention to situations where factor price equalization holds. In chapter 3, we examine the case of many goods and factors in greater detail, to determine whether the Stolper-Samuelson and Rybczynski theorems generalize and also how to estimate these effects. In chapter 4, evidence on the relationship between product prices and wages is examined in detail, using a model that allows for trade in intermediate inputs. The reader is already well prepared for the chapters that follow, based on the tools and intuition we have developed from the two-by-two model. Before moving on, you are encouraged to complete the problems at the end of this chapter.

APPENDIX: THE LERNER DIAGRAM AND FACTOR PRICES

The Lerner (1952) diagram for the two-by-two model can be explained as follows: With perfect competition and constant returns to scale, we have that revenue = costs in both industries. So let us choose a special isoquant in each industry such that revenue = 1. In each industry, we therefore choose the isoquant $p_i y_i = 1$, or

$$Y = f_i(L_i, K_i) = 1/p_i \Rightarrow wL_i + rK_i = 1.$$

¹⁵This diagram was used in a seminar presented by Abba Lerner at the London School of Economics in 1933, but not published until 1952. The history of this diagram is described at the “Origins of Terms in International Economics,” maintained by Alan Deardorff at <http://www-personal.umich.edu/~alandear/glossary/orig.html>. See also Samuelson (1949, 181 n.1).

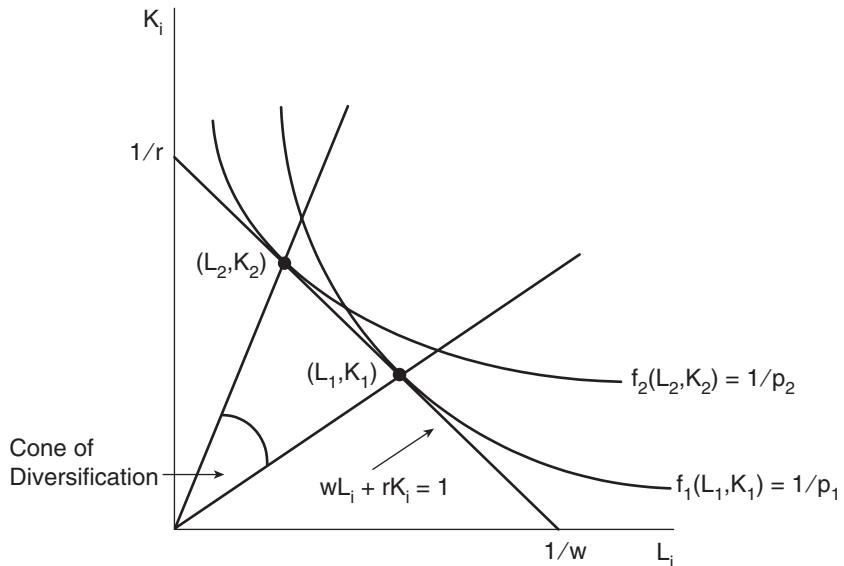


Figure 1.13

Therefore, from cost minimization, the $1/p_i$ isoquant in each industry will be *tangent* to the line $wL_i + rK_i = 1$. This is the same line for both industries, as shown in figure 1.13.

Drawing the rays from the origin through the points of tangency, we obtain the cone of diversification, as labeled in figure 1.13. Furthermore, we can determine the factor prices by computing where $wL_i + rK_i = 1$ intersects the two axes: $L_i = 0 \Rightarrow K_i = 1/r$, and $K_i = 0 \Rightarrow L_i = 1/w$. Therefore, given the prices p , we determine the two isoquants in figure 1.13, and drawing the (unique) line tangent to both of these, we determine the factor prices as the intercepts of this line. Notice that these equilibrium factor prices do not depend on the factor endowments, provided that the endowment vector lies within the cone of diversification (so that both goods are produced). We have thus obtained an alternative proof of the “factor price insensitivity” lemma, using a primal rather than dual approach. Furthermore, with two countries having the same prices (through free trade) and technologies, then figure 1.13 holds in both of them. Therefore, their factor prices will be equalized.

Lerner (1952) also showed how figure 1.13 can be extended to the case of factor intensity reversals, in which case the isoquants intersect twice. In that case there will be *two* lines $wL_i + rK_i = 1$ that are tangent to both isoquants, and there are two cones of diversification. This is shown in figure 1.14. To determine which factor prices apply in a particular country, we plot its endowments vector and note which cone of diversification it lies in: the factor prices in this country are those applying to that cone. For example, the endowments (L^A, K^A) will have the factor prices (w^A, r^A) , and the endowments (L^B, K^B) will have the factor prices (w^B, r^B) . Notice that the labor-abundant country with endowments (L^A, K^A) has the low wage and high rental, whereas the capital-abundant country with endowments (L^B, K^B) has the high wage and low rental.

How likely is it that the isoquants of industries 1 and 2 intersect twice, as in figure 1.14? Lerner (1952, 11) correctly suggested that it depends on the elasticity of substitution between labor and capital in each industry. For simplicity, suppose that each

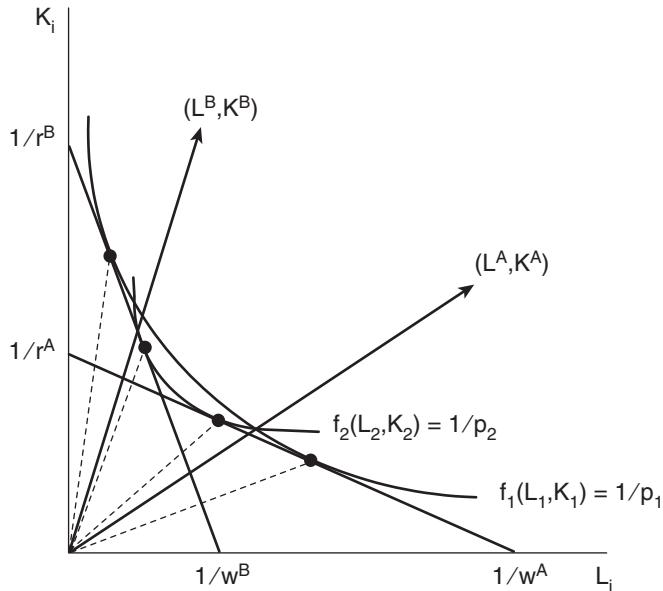


Figure 1.14

industry has a constant elasticity of substitution production function. If the elasticities are the same across industries, then it is impossible for the isoquants to intersect twice. If the elasticities of substitution differ across industries, however, and we choose prices p_i , $i = 1, 2$, such that the $1/p_i$ isoquants intersect at least once, then it is *guaranteed* that they intersect twice. Under exactly the same conditions, the iso-cost lines in figure 1.6 intersect twice. Thus, the occurrence of FIR is very likely once we allow elasticities of substitution to differ across industries. Minhas (1962) confirmed that this was the case empirically, and discussed the implications of FIR for factor prices and trade patterns. This line of empirical research was dropped thereafter, perhaps because FIR seemed too complex to deal with, and has been picked up again more recently (see note 14 of this chapter).

PROBLEMS

- 1.1 Rewrite the production function $y_1 = f_1(L_1, K_1)$ as $y_1 = f_1(v_1)$, and similarly, $y_2 = f_2(v_2)$. Concavity means that given two points $y_1^a = f_1(v_1^a)$ and $y_1^b = f_1(v_1^b)$, and $0 \leq \lambda \leq 1$, then $f_1(\lambda v_1^a + (1 - \lambda)v_1^b) \geq \lambda y_1^a + (1 - \lambda)y_1^b$. Similarly for the production function $y_2 = f_2(v_2)$. Consider two points $y^a = (y_1^a, y_2^a)$ and $y^b = (y_1^b, y_2^b)$, both of which can be produced while satisfying the full-employment conditions $v_1^a + v_2^a \leq V$ and $v_1^b + v_2^b \leq V$, where V represents the endowments. Consider a production point midway between these, $\lambda y^a + (1 - \lambda)y^b$. Then use the concavity of the production functions to show that this point can *also* be produced while satisfying the full-employment conditions. This proves that the production possibilities set is *convex*. (Hint: Rather than showing that $\lambda y^a + (1 - \lambda)y^b$ can be produced while satisfying the full-employment

- conditions, consider instead allocating $\lambda v_1^a + (1-\lambda)v_1^b$ of the resources to industry 1, and $\lambda v_2^a + (1-\lambda)v_2^b$ of the resources to industry 2.)
- 1.2 Any function $y = f(v)$ is homogeneous of degree α if for all $\lambda > 0$, $f(\lambda v) = \lambda^\alpha f(v)$. Consider the production function $y = f(L, K)$, which we assume is homogeneous of degree one, so that $f(\lambda L, \lambda K) = \lambda f(L, K)$. Now differentiate this expression with respect to L , and answer the following: Is the marginal product $f_L(L, K)$ homogeneous, and of what degree? Use the expression you have obtained to show that $f_L(L/K, 1) = f_L(L, K)$.
- 1.3 Consider the problem of maximizing $y_1 = f_1(L_1, K_1)$, subject to the full-employment conditions $L_1 + L_2 \leq L$ and $K_1 + K_2 \leq K$, and the constraint $y_2 = f_2(L_2, K_2)$. Set this up as a Lagrangian, and obtain the first-order conditions. Then use the Lagrangian to solve for dy_1/dy_2 , which is the slope of the production possibilities frontier. How is this slope related to the marginal product of labor and capital?
- 1.4 Consider the problem of maximizing $p_1 f_1(L_1, K_1) + p_2 f_2(L_2, K_2)$, subject to the full-employment constraints $L_1 + L_2 \leq L$ and $K_1 + K_2 \leq K$. Call the result the GDP function $G(p, L, K)$, where $p = (p_1, p_2)$ is the price vector. Then answer the following:
- What is $\partial G / \partial p_i$? (Hint: we solved for this in the chapter.)
 - Give an economic interpretation to $\partial G / \partial L$ and $\partial G / \partial K$.
 - Give an economic interpretation to $\partial^2 G / \partial p_i \partial L = \partial^2 G / \partial L \partial p_i$, and $\partial^2 G / \partial p_i \partial K = \partial^2 G / \partial K \partial p_i$.
- 1.5 Trace through changes in outputs when there are factor intensity reversals. That is, construct a graph with the capital endowment on the horizontal axis, and the output of goods 1 and 2 on the vertical axis. Starting at a point of diversification (where both goods are produced) in cone A of figure 1.12, draw the changes in output of goods 1 and 2 as the capital endowment grows outside of cone A, into cone B, and beyond this.

The Heckscher-Ohlin Model

We begin this chapter by describing the Heckscher-Ohlin (HO) model with two countries, two goods, and two factors (or the two-by-two-by-two model). This formulation is often called the Heckscher-Ohlin-Samuelson (HOS) model, which is based on the work of Paul Samuelson, who developed a mathematical model from the original insights of Eli Heckscher and Bertil Ohlin.¹ The goal of that model is to predict the pattern of trade in goods between the two countries, based on their differences in factor endowments. Following this, we present the multigood, multifactor extension that is associated with the work of Vanek (1968), and is often called the Heckscher-Ohlin-Vanek (HOV) model. As we shall see, in this latter formulation we do not attempt to keep track of the trade pattern in individual goods but instead compute the “factor content” of trade, that is, the amounts of labor, capital, land, and so on embodied in the exports and imports of a country.

The factor-content formulation of the HOV model has led to a great deal of empirical research, beginning with Leontief (1953) and continuing with Leamer (1980), Bowen, Leamer, and Sveikauskas (1987), Trefler (1993a, 1995), Davis and Weinstein (2001), with many other writers in between. We will explain the twists and turns in this chain of empirical research. The bottom line is that the HOV model performs quite poorly empirically unless we are willing to dispense with the assumption of identical technologies across countries. This brings us back to the earlier tradition of the Ricardian model of allowing for technological differences, which also implies differences in factor prices across countries. We will show several ways that technological differences can be incorporated into an “extended” HO model, with their empirical results, and this remains a question of ongoing research.

HECKSCHER-OHLIN-SAMUELSON (HOS) MODEL

The basic assumptions of the HOS model were already introduced in the previous chapter: identical technologies across countries; identical and homothetic tastes across countries; differing factor endowments; and free trade in goods (but not factors). For the most part, we will also assume away the possibility of factor intensity

¹The original 1919 article of Heckscher and the 1924 dissertation of Ohlin have been translated from Swedish and edited by Harry Flam and June Flanders and published as Heckscher and Ohlin (1991).

reversals. Provided that all countries have their endowments *within* their “cone of diversification,” this means that factor prices are equalized across countries

We begin by supposing that there are just two countries, two sectors and two factors, exactly like the two-by-two model we introduced in chapter 1. We shall assume that the home country is labor abundant, so that $L/K > L^*/K^*$. We will also assume that good 1 is labor intensive. The countries engage in free trade, and we also suppose that trade is balanced (value of exports = value of imports). Then the question is, What is the pattern of trade in goods between the countries? This is answered by the following theorem.

HECKSCHER-OHLIN THEOREM

Each country will export the good that uses its abundant factor intensively.

Thus, under our assumptions the home country will export good 1 and the foreign country will export good 2. To prove this, let us take a particular case of the factor endowment differences $L/K > L^*/K^*$, and assume that the labor endowments are identical in the two countries, $L^* = L$, while the foreign capital endowments exceed that at home, $K^* > K$.² In order to derive the pattern of trade between the countries, we proceed by first establishing what the relative product price is in each country *without* any trade, or *in autarky*. As we shall see, the pattern of autarky prices can then be used to predict the pattern of trade: a country will export the good whose free trade price is higher than its autarky price, and import the other.

Let us begin by illustrating the *home autarky* equilibrium, at point A in figure 2.1. We have assumed a representative consumer with homothetic tastes, so we can use indifference curves to reflect demand. The autarky equilibrium is established where an indifference curve is tangent to the home production possibility frontier (PPF), at point A. The price line drawn tangent to the PPF and indifference curve has a slope of (negative) the autarky relative price of good 1, $p^a \equiv p_1^a/p_2^a$. Let us now consider the foreign PPF, which is drawn outside the home PPF in the figure 2.1. In order to determine where the foreign autarky equilibrium lies, let us initially suppose that p^a is *also* the autarky equilibrium abroad, and see whether this assumption leads to a contradiction.

If p^a is also the autarky price in the foreign country, then production must occur at the tangency between the price line with slope p^a and the foreign PPF, or at point B'. Notice that from the Rybczynski theorem, point B' must lie *above and to the left* of point A: the higher capital endowment abroad leads to more of good 2 and less of good 1. The price line through point B' acts like a budget constraint for the representative consumer in the foreign country, so that the consumer chooses the highest indifference curve on this price line. Since tastes are homothetic, the foreign representative consumer will demand the two goods in exactly the same proportion as does the home representative consumer. In other words, the foreign consumption point must lie on the budget constraint through point B', and also on a ray from the origin through point A. Thus, foreign consumption must occur at point C', which is *above and to the right* of point A. Since points B' and C' do not coincide, we have arrived at a contradiction: the relative price p^a at home *cannot* equal the autarky price abroad, and on the

²Because of the assumptions of identical homothetic tastes and constant returns to scale, the result we are establishing remains valid if the labor endowments also differ across countries.

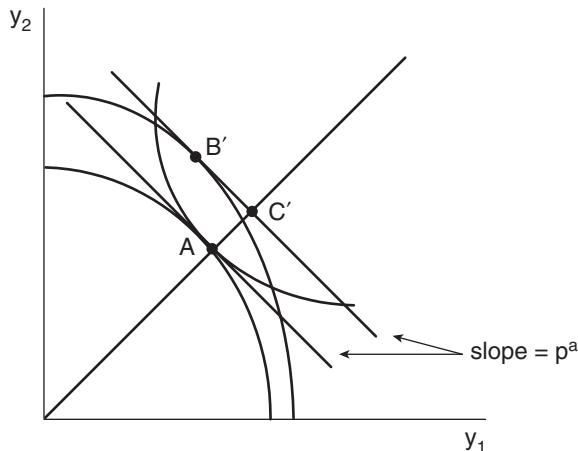


Figure 2.1

contrary, at this price there is an excess demand for good 1 in the foreign country. This excess demand will bid up the relative price of good 1, so that the foreign autarky price must be *higher* than at home, $p^{a*} > p^a$.

To establish the free trade equilibrium price, let $z(p)$ denote the excess demand for good 1 at any prevailing price p at home, while $z^*(p^*)$ denotes the excess demand for good 1 abroad. World excess demand at a common price is therefore $z(p) + z^*(p)$, and a free-trade equilibrium occurs when world excess demand is equal to zero. The home autarky equilibrium satisfies $z(p^a) = 0$, and we have shown above that $z^*(p^a) > 0$. It follows that $z(p^a) + z^*(p^a) > 0$. If instead we reversed the argument in figure 2.1, and started with the foreign autarky price satisfying $z^*(p^{a*}) = 0$, then we could readily prove that $z(p^{a*}) < 0$, so at the foreign autarky price there is excess supply of good 1 at home. It follows that world excess demand would satisfy $z(p^{a*}) + z^*(p^{a*}) < 0$. Then by continuity of the excess demand functions, there must be a price p , with $p^{a*} > p > p^a$, such that $z(p) + z^*(p) = 0$. This is the equilibrium price with free trade.

Let us illustrate the free trade equilibrium, in figure 2.2. In panel (a) we show the equilibrium at home, and in panel (b) we show the equilibrium in the foreign country. Beginning at the home autarky point A , the relative price of good 1 *rises* at home, $p > p^a$. It follows that production will occur at a point like B , where the price line through point B has the slope p . Once again, this price line acts as a budget constraint for the representative consumer, and utility is maximized at point C . The difference between production at point B and consumption at point C is made up through exporting good 1 and importing good 2, as illustrated by the “trade triangle” drawn. This trade pattern at home establishes the HO theorem stated above. In the foreign country, the reverse pattern occurs: the relative price of good 1 *falls*, $p^{a*} > p$, and production moves from the autarky equilibrium point A^* to production at point B^* and consumption at point C^* , where good 1 is imported and good 2 is exported. Notice that the trade triangles drawn at home and abroad are *identical* in size: the exports of one country must be imports of the other.³

³Also, notice that the slope of the hypotenuse of the trade triangle for the home country is $(\text{import}_2 / \text{export}_1) = p$. It follows that $(p \cdot \text{export}_1) = \text{import}_2$, so that trade is balanced, and this also holds for the foreign country.

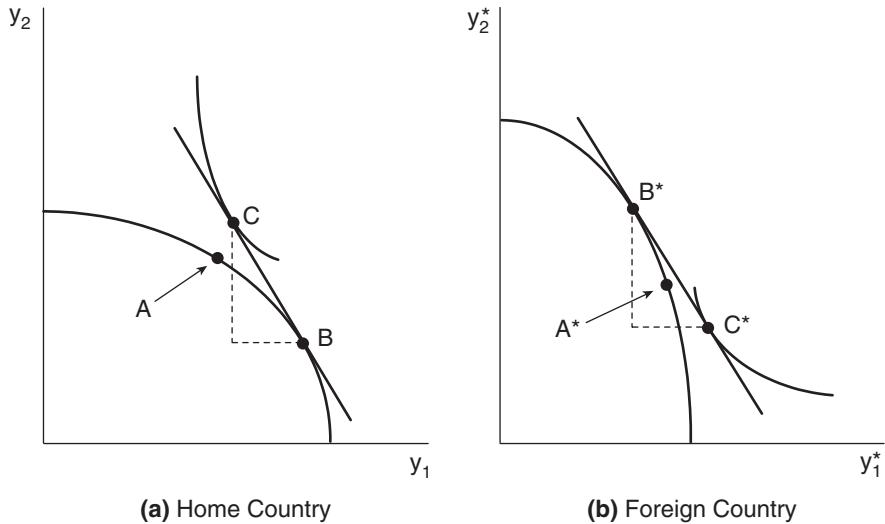


Figure 2.2

In addition to establishing the trade pattern, the HO model has precise implications about who gains and who loses from trade: the *abundant* factor in each country gains from trade, and the *scarce* factor loses. This result follows from the pattern of price changes ($p^* > p > p^d$) and the Stolper-Samuelson theorem. With the relative price of good 1 rising at home, the factor used intensively in that good (labor) will gain in real terms, and the other factor (capital) will lose. Notice that labor is the abundant factor at home. The fact that $L/K > L^*/K^*$ means that labor would have been earning less in the home autarky equilibrium than in the foreign autarky equilibrium: its marginal product at home would have been lower (in both goods) than abroad. However, with free trade the home country can shift production toward the labor-intensive good, and export it, thereby absorbing the abundant factor without lowering its wage. Indeed, factor prices are *equalized* in the two countries after trade, as we argued in the previous chapter. Thus, the abundant factor, whose factor price was bid down in autarky, will gain from the opening of trade, while the scarce factor in each country loses.

Our presentation of the HO model above is about as far as most discussion of this model goes at the undergraduate level. After showing something like figure 2.2, it would be common to provide some rough data or anecdotes to illustrate the HO theorem (e.g., the United States is abundant in scientists, so it exports high-tech goods; Canada is abundant in land, so it exports natural resources, etc.). As plausible as these illustrations are, it turns out that the HO model is a *rather poor* predictor of actual trade patterns, indicating that its assumptions are not realistic. It has taken many years, however, to understand why this is the case, and we begin this exploration by considering the earliest results of Leontief (1953).

LEONTIEF'S PARADOX

Leontief (1953) was the first to confront the HO model with data. He had developed the set of input-output accounts for the U.S. economy, which allowed him to compute

TABLE 2.1: Leontief's (1953) Test

	Exports	Imports
Capital (\$ million)	\$2.5	\$3.1
Labor (person-years)	182	170
Capital/Labor (\$/person)	\$13,700	\$18,200

Note: Each column shows the amount of capital or labor needed per \$1 million worth of exports or imports into the United States, for 1947.

the amounts of labor and capital used in each industry for 1947. In addition, he utilized U.S. trade data for the same year to compute the amounts of labor and capital used in the production of \$1 million of U.S. exports and imports. His results are shown in table 2.1.

Leontief first measured the amount of capital and labor required for \$1 million worth of U.S. exports. This calculation requires that we measure the labor and capital used *directly*, that is, in each exporting industry, and also those factors used *indirectly*, that is, in the industries that produce intermediate inputs that are used in producing exports. From the first row of table 2.1, we see that \$2.5 million worth of capital was used in \$1 million of exports. This amount of capital seems much too high, until we recognize that what is being measured is the *capital stock*, so that only the annual depreciation on this stock is actually used. For labor, 182 person-years were used to produce the exports. Taking the ratio of these, we find that each person employed in producing exports (directly or indirectly) is working with \$13,700 worth of capital.

Turning to the import side of the calculation, we immediately run into a problem: it is not possible to measure the amount of labor and capital used in producing imports unless we have knowledge of the *foreign* technologies, which Leontief certainly did not know in 1953! Indeed, it is only recently that researchers have begun to use data on foreign technologies to test the HO model, as we will describe later in the chapter. So Leontief did what many researchers have done since: he simply used the *U.S. technology* to calculate the amount of labor and capital used in imports. Does this invalidate the test of the HO model? Not really, because recall that an assumption of the HO model is that technologies are the same across countries. Thus, under the null hypothesis that the HO model is true, it would be valid to use the U.S. technology to measure the labor and capital used in imports. If we find that this null hypothesis is rejected, then one explanation would be that the assumption of identical technologies is false.

Using the U.S. technology to measure the labor and capital used in imports, both directly and indirectly, we arrive at the estimates in the last column of table 2.1: \$3.1 million of capital, 170 person-years, and so a capital/labor ratio in imports of \$18,200. Remarkably, this is *higher* than the capital/labor ratio found for U.S. exports! Under the presumption that the United States was capital-abundant in 1947, this appears to contradict the HO theorem. Thus, this finding came to be called “Leontief’s paradox.”

A wide range of explanations have been offered for this paradox:

- U.S. and foreign technologies are not the same;
- By focusing only on labor and capital, Leontief ignored land;
- Labor should have been disaggregated by skill (since it would not be surprising to find that U.S. exports are intensive in skilled labor);

- The data for 1947 may be unusual, since World War II had just ended;
- The U.S. was not engaged in free trade, as the HO model assumes.

These reasons are all quite valid criticisms of the test that Leontief performed, and research in the years following his test aimed to redo the analysis while taking into account land, skilled versus unskilled labor, other years, and so on. This research is well summarized by Deardorff (1984a), and the general conclusion is that the paradox continued to occur in some cases. It was not until two decades later, however, that Leamer (1980) provided the definitive critique of the Leontief paradox: it turned out that Leontief had performed the wrong test! That is, even if the HO model is true, it turns out the capital/labor ratios in export and imports, as reported in table 2.1, should not be compared. Instead, an alternative test should be performed. The test that Leamer proposed relies on the “factor content” version of the Heckscher-Ohlin model, developed by Vanek (1968), which we turn to next.

HECKSCHER-OHLIN-VANEK (HOV) MODEL

Let us now consider many countries, indexed by $i=1,\dots,C$; many industries, indexed by $j=1,\dots,N$; and many factors, indexed by k or $\ell=1,\dots,M$. We will continue to assume that technologies are identical across countries, and that factor price equalization prevails under free trade. In addition, we assume that tastes are identical and homothetic across countries.

Let the $(M \times N)$ matrix $A = [a_{jk}]'$ denote the amounts of labor, capital, land, and other primary factors needed for one unit of production in each industry.⁴ Notice that this matrix applies in any country. The *rows* measure the different factors k , $\ell=1,\dots,M$, while the *columns* of this matrix measure the different industries $j=1,\dots,N$. For example, with just two industries using only labor and capital, this matrix would be $A = \begin{bmatrix} a_{1k} & a_{2k} \\ a_{1k} & a_{2k} \end{bmatrix}$.

Next, let Y^i denote the $(N \times 1)$ vector of outputs in each industry for country i , and let D^i denote the $(N \times 1)$ vector of demands of each good, so that $T^i = Y^i - D^i$ equals the vector of *net exports* for country i . The *factor content of trade* is then defined as $F^i \equiv AT^i$, which is an $(M \times 1)$ vector. We will denote individual components of this vector as F_k^i , where a positive value indicates that the factor is exported, while a negative value indicates that the factor is imported. For example, with just labor and capital, the factor content of trade is

$$\begin{pmatrix} F_\ell^i \\ F_k^i \end{pmatrix} \equiv AT^i.$$

The goal of the HOV model is to relate the factor content of trade AT^i to the underlying endowments of country i . To do so, we can proceed by computing AY^i and AD^i . The term AY^i equals the demand for factors in country i . Analogous to the full-employment conditions studied in chapter 1, AY^i equals the endowments of country i , which we write as $AY^i = V^i$. Turning to AD^i , this term is simplified by using our

⁴This matrix should include both the *direct* primary factors used in the production of each good, and the *indirect* primary factors used through the intermediate inputs. In practice, the indirect factors are measured using the input-output matrix for the economy. That is, denoting the $(M \times N)$ *direct* factor requirements by \tilde{A} and the $(N \times N)$ input-output matrix by B , we compute the *total* factor requirements as $A = \tilde{A}(I - B)^{-1}$.

assumption of identical and homothetic tastes. Since product prices are equalized across countries by free trade, it follows that the consumption vectors of all countries must be *proportional* to each other. We shall write this as $D^i = s^i D^w$, where D^w denotes the *world* consumption vector and s^i is the share of country i in world consumption.⁵ It follows that $AD^i = s^i AD^w$. Note that if trade is balanced, then s^i also equals country i 's share of world GDP.⁶ Since world consumption must equal world production, we therefore obtain $AD^i = s^i AD^w = s^i AY^w = s^i V^w$, where the last equality is the full-employment condition at the world level.

Making use of these expressions for AY^i and AD^i , we have therefore proved

$$F^i \equiv AT^i = V^i - s^i V^w, \quad (2.1)$$

which is a statement of the Heckscher-Ohlin-Vanek (HOV) theorem. In terms of individual factors, this is written as $F_k^i = V_k^i - s^i V_k^w$. If country i 's endowment of factor k relative to the world endowment exceeds country i 's share of world GDP ($V_k^i/V_k^w > s^i$), then we say that country i is *abundant* in that factor. In that case, (2.1) says that the factor-content of trade in factor k should also be positive ($F_k^i > 0$), and conversely if country i is scarce in factor k ($V_k^i/V_k^w < s^i$).

What does the HOV theorem tell us about the Leontief test? To answer this, let us focus on just two elements of the factor-content vector, for labor and capital. These are written as

$$F_k^i = K^i - s^i K^w, \quad (2.2a)$$

$$F_\ell^i = L^i - s^i L^w, \quad (2.2b)$$

where F_k^i and F_ℓ^i are the computed factor contents of trade, and K^i and L^i are the capital and labor endowments for country i . Following Leamer (1980), we define capital to be abundant relative to labor in country i if $K^i/K^w > L^i/L^w$. Then using (2.2), the implications of capital abundance are:

THEOREM (LEAMER 1980)

If capital is abundant relative to labor in country i , then the HOV theorem (2.1) implies that the capital/labor ratio embodied in *production* for country i exceeds the capital/labor ratio embodied in *consumption*:

$$K^i/L^i > (K^i - F_k^i)/(L^i - F_\ell^i). \quad (2.3)$$

Proof:

From equation (2.2), we have $K^w = (K^i - F_k^i)/s^i$ and $L^w = (L^i - F_\ell^i)/s^i$. It follows that $K^i/K^w = s^i K^i / (K^i - F_k^i)$ and $L^i/L^w = s^i L^i / (L^i - F_\ell^i)$. Then $K^i/K^w > L^i/L^w$ implies that $K^i/(K^i - F_k^i) > L^i/(L^i - F_\ell^i)$, which is rewritten as (2.3). QED

⁵Letting p denote the vector of prices, then $p'D^i = s^i p'D^w$ so that $s^i = p'D^i/p'D^w$.

⁶Continuing from note 5, if trade is balanced so that expenditure = income in each country, then $p'D^i = p'Y^i$ and so $s^i = p'Y^i/p'Y^w = GDP^i/GDP^w$.

To interpret this result, note that K^i and L^i are simply the endowments of capital and labor, or alternatively, the capital and labor embodied in *production*. If we subtract the content of these factors embodied in trade, then what we end up with can be defined as the factor-content of *consumption*, or $K^i - F_k^i$ and $L^i - F_\ell^i$. Then equation (2.3) states that the capital/labor ratio embodied in *production* (on the left) must exceed the capital/labor ratio embodied in *consumption* (on the right).

The results from making this comparison for the United States in 1947 are shown in table 2.2. In the first column we list the capital and labor endowments for the United States, and in the second column we show the capital and labor embodied in consumption. Taking the ratio of these, it is indeed the case that the capital/labor ratio embodied in production exceeds that in consumption. This is the precise application of the HOV theorem, and it turns out to be satisfied for the United States in 1947, contrary to what Leontief concluded. Thus, there was no paradox after all!

It is useful to see the HOV theorem and Leamer's result in a diagram. In figure 2.3, the length of the horizontal axis is the world labor endowment $L^w = L^1 + L^2$, and the length of the vertical axis is the world capital endowment $K^w = K^1 + K^2$. The origin for country 1 is in the lower-left corner, and for country 2 is in the upper-right corner. Thus, any point in the world endowment box measures the endowments (L^i, K^i) of the two countries. Suppose that the endowments are at the point V^i where country 1 is capital-abundant, $K^1/L^1 > K^w/L^w > K^2/L^2$.

Under the assumptions of the HOV model, the consumption of each country D^i is proportional to world consumption D^w , which means that the factor content of consumption AD^i is proportional to $AD^w = V^w$. In other words, the factor content of consumption must lie along the *diagonal* in the world endowment box, as illustrated by point AD^i . Therefore, a line from point V^i to point AD^i measures the factor content of trade. In figure 2.3, country 1 exports F_k^1 of capital services and imports F_ℓ^1 of labor services. With balanced trade, the slope of the line between V^i and AD^i measures the ratio of factor prices.

The theorem of Leamer (1980) states that if country 1 is capital abundant, as illustrated, then the capital/labor ratio embodied in *production* must exceed the capital/labor ratio embodied in *consumption*. That is, since the consumption point AD^i must lie on the diagonal, it is necessarily to the right of and below the endowment point V^i . While this is graphically obvious, note that it does not depend in any way on whether trade is balanced or not. For example, if country 1 is running a trade surplus (with the value of production exceeding consumption), then we should move the consumption point AD^i to the left down the diagonal. This would have no effect whatsoever on the capital/labor *ratio* embodied in consumption as compared to the capital/labor *ratio* embodied in production. So Leamer's test of the HOV theorem in (2.3) is completely robust to having nonbalanced trade. Indeed, Leamer (1980) argues that this was the key problem with the way that Leontief did the original test: Leontief's method of testing

TABLE 2.2: Leamer's (1980) Reformulation of the Leontief Test

	Production	Consumption
Capital (\$ billion)	\$327	\$305
Labor (person-years)	47 million	45 million
Capital/Labor (\$/person)	\$6,949	\$6,737

Note: Each column shows the amount of capital or labor embodied in production or consumption in the United States, for 1947.

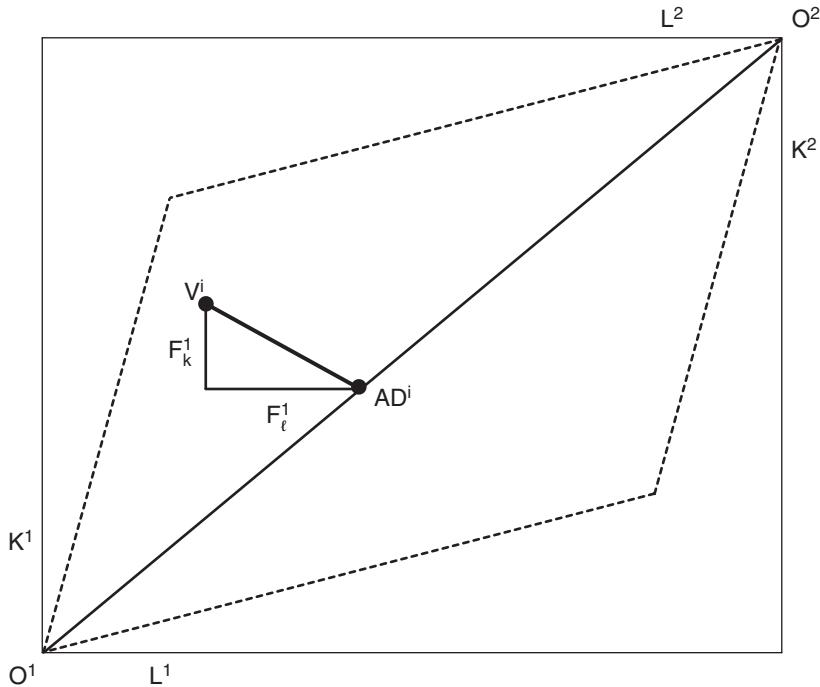


Figure 2.3

the HOV theorem was not valid with nonbalanced trade, and in 1947, the United States had a trade surplus and was exporting both labor and capital as embodied in trade.⁷

PARTIAL TESTS OF THE HOV THEOREM

The statement of the HOV theorem in (2.1) tells us immediately how a complete test of the theory should be performed: simply compute the left-hand side (using data on trade T^i and technology A), compute the right-hand side (using data on endowments V^i and V^w), and compare them. Depending on how well they match up, we can judge whether the theorem is an empirical success or not. This complete test requires both trade and endowments data for many countries, and technology data for at least one country. While such data are readily available today,⁸ this was not the case two or three decades ago. Accordingly, many researchers, including Leontief himself, performed what we can call “partial tests” of the HOV model, using only two rather than all three types of data. Before considering the complete test, we will review several other partial tests that were performed.

⁷This situation is shown in figure 2.3 by moving the consumption AD^i so far to the left and down the diagonal that both capital services F_k^i and labor services F_l^i are exported. Brecher and Choudhri (1982a) point out that the United States being a net exporter of labor is itself a paradox: from (2.2b), $F_l^i > 0$ if and only if $L^i > s^i L^w$, which implies $p'D^i L^i < p'D^w L^w$ using the expression for s^i from note 5. In other words, the United States will be exporting labor if and only if its per capita income is less than the world average, which is plainly false!

⁸Input-output tables as well as bilateral trade data for many countries are available from the Organization for Economic Cooperation and Development (OECD) STAN and I-O datasets, as well as from the World Input-Output Database, or WIOD (see Timmer et al. 2014).

Let us assume that the number of goods equals the number of factors, so that A is a square matrix, which we assume is invertible. Then we can rewrite (2.1) as

$$T^i = A^{-1}(V^i - s^i V^w). \quad (2.4)$$

This equation could be tested in several ways. First, if we think about the matrix A^{-1} as data, then we could run a regression of T^i on A^{-1} , and coefficients obtained would serve as an estimate of the relative abundance ($V^i - s^i V^w$) of each factor. Baldwin (1971) performed a test similar to this, but instead of regressing T^i on A^{-1} , he actually regressed T^i on A' . With two factors and three goods, for example, A' is the matrix

$$\begin{bmatrix} a_{1L} & a_{1K} \\ a_{2L} & a_{2K} \\ a_{3L} & a_{3K} \end{bmatrix}.$$

Thus, Baldwin regressed the *adjusted net exports*⁹ of each industry on its *labor and capital requirements* for one unit of production.

Using data for sixty U.S. industries for years around 1960, and disaggregating workers by their types, Baldwin (1971) obtained the following result:

$$\begin{aligned} \text{Adjusted net exports} = & -1.37^* \text{ (physical capital/worker)} + 7011^* \text{ (scientists/worker)} \\ & - 1473 \text{ (managers/worker)} + 71 \text{ (clerical staff/worker)} \\ & + 1578^* \text{ (craftsmen \& foremen/worker)} + 248 \text{ (operatives/worker)} \\ & + 761 \text{ (unskilled employees/worker)} + 845^* \text{ (farmers/worker)} \\ & + \text{(other variables included for industry scale and unionization)}, \\ & N = 60, R^2 = 0.44, ^* = \text{significant at 95\% level} \end{aligned}$$

Thus, looking across the U.S. industries, Baldwin finds that those industries using more scientists, craftsmen and foremen, or farmers relative to total workers will tend to have *higher* exports. The importance of scientists and farmers in predicting U.S. exports is not surprising at all, since the United States is abundant in skilled labor and land; and the importance of craftsmen and foremen is perhaps reasonable, too. What is surprising, however, is the *negative* coefficient found on the very first variable, physical capital/worker. Taken literally, this coefficient says that U.S. industries using *more* capital per worker will tend to export *less*. This is exactly the opposite of what we would expect if the United States were capital abundant. Thus, this result appears to be similar to the “paradox” found by Leontief.

Various writers after Baldwin have redone the type of regression shown above, with mixed results: sometimes the capital coefficient is positive, but other times it is again negative (see the survey by Deardorff 1984a). What are we to make of these results? Well, as we argued for Leontief’s original paradox, it can be questioned whether Baldwin’s approach is a *valid test* of the HOV model. From (2.4), a valid test would be to regress T^i on A^{-1} , but instead Baldwin regressed T^i on A' . To see the consequences of this, use the ordinary least squares (OLS) formula for the coefficients β that would be obtained from this regression:

$$T^i = A'\beta \Rightarrow \hat{\beta} = (AA')^{-1}AT^i = (AA')^{-1}(V^i - s^i V^w), \quad (2.5)$$

where the final equality follows from using the HOV theorem (2.1).

⁹ Adjusted net exports are defined as industry exports per million dollars of total exports minus industry imports per million dollars of total imports.

Thus, we see that $\hat{\beta}$ is a *contaminated estimate* of the vector of relative factor endowments ($V^i - s^i V^w$). Namely, rather than equaling the relative factor endowments vector, $\hat{\beta}$ equals the positive definite matrix $(AA')^{-1}$ times $(V^i - s^i V^w)$. Because $(AA')^{-1}$ is certainly not the identity matrix, it is entirely possible that the *sign pattern* of the elements in $\hat{\beta}$ will differ from the sign pattern of $(V^i - s^i V^w)$.¹⁰ In other words, even if the United States was capital abundant and the HOV theorem held, it would still be possible for the Baldwin regression to find a negative coefficient on capital: this does not contradict the HOV theorem, because like the original Leontief approach, it is the wrong test.

The careful reader might point out, however, that possible differences in the sign pattern of $\hat{\beta}$ and $(V^i - s^i V^w)$ can presumably be *checked for* using the actual data on $(AA')^{-1}$. This is exactly the approach taken by Bowen and Sveikauskas (1992), who argue that such sign reversals are very unlikely to occur in practice. Thus, the Baldwin regression may be acceptable in practice, and in fact, it has been used in research by Romalis (2004).¹¹ As a descriptive tool to show how trade is related to industry factor requirements this regression makes good sense, but as a definitive *test* of the HOV theorem it is inadequate for the reasons we have described.

A second “partial test” of the HOV theorem in equation (2.4) has been undertaken by Leamer (1984). In contrast to Baldwin, Leamer tested (2.4) by treating factor endowments ($V^i - s^i V^w$) as data, while estimating the elements of A^{-1} . To follow this test, notice that (2.4) applies across all countries i . Focusing on a single industry j , and letting the elements of A^{-1} be written as β_{jk} , we can write (2.4) in scalar form as

$$T_j^i = \sum_{k=1}^M \beta_{jk} (V_k^i - s^i V_k^w), \quad i = 1, \dots, C. \quad (2.6)$$

Notice that the summation in (2.6) is across factors, while the observations are across countries. Thus, to estimate this regression we would combine the observation for a *single* industry j across multiple countries $i = 1, \dots, C$, where the coefficients β_{jk} are estimated. These coefficients should be interpreted as Rybczynski effects and, as shown in chapter 1, can be positive or negative.

To run this regression, we first need to choose some aggregation scheme for net exports and factors of production. Leamer (1984) works with the trade data for sixty countries in two years, 1958 and 1975. The trade data are organized according to the Standard Industrial Trade Classification (SITC). Because these are *traded* goods, they include only merchandise, that is, agriculture, mining, and manufacturing. Leamer organizes these goods into ten aggregates, as shown in the first column of table 2.3: two primary products (petroleum and raw materials); four crops (forest products, tropical agricultural products, animal products, and cereals); and four manufactured products (labor-intensive manufactures, capital-intensive manufactures, machinery, and chemicals).¹² Implicitly there is an eleventh product in the economy, making up all the nontraded goods, so Leamer includes total GNP as an aggregate to reflect this. There are also eleven factors of production, listed along the top row of table 2.3: capital,

¹⁰See Leamer and Bowen (1981) and Aw (1983).

¹¹Romalis's model relies on the monopolistic competition framework, which we do not cover until chapter 5. It is recommended for further reading after that.

¹²The method used to form these aggregates involves looking at the cross-country correlations of the disaggregate products in each group with the aggregate itself. That is, “cereals” is an aggregate because countries that export this product in total tend to have high exports of all the products within this group. It turns out that the ratios of capital to worker and professional to all workers are also reasonably similar within the aggregates. Feenstra and Hanson (2000) consider how the aggregation scheme might lead to bias in calculations of the factor content of trade.

TABLE 2.3: Estimates of Rybczynski Effects from Leamer (1984)

	Capital	Professional & Technical Workers	Non-professional Workers	Illiterate Workers	Land1	Land2	Land3	Land4	Coal	Minerals	Oil	R ²
GNP	453**	15,683**	-445*	-480**	30**	-49**	81*	-27*	9.0**	2.8*	2.1**	0.997
Primary Products												
Petroleum	-18.4**	-248	-.8	31*	0.5	7.9**	-5.4	0.3	-1.0**	-0.4	0.6**	0.92
Raw materials	-8.9**	127	-31.2*	-2.3	0.89	0.47	1.1	3.3**	0.45**	0.86**	0.04*	0.86
Crops												
Forest products	-1.7	34	-18.8	8.4	-0.45	-2.7*	0.3	3.3*	-0.17	0.53*	0.08*	0.53
Tropical agricultural products	-2.9**	-30.1	8.9	5.6	2.0**	-1.1	4.5*	-3.9**	-0.30**	0.44**	0.05*	0.72
Animal products	-0.5	-42	-29.3	14.0	0.09	-0.34	5.6*	-1.3	0.17	0.28	0.05	0.22
Cereals	-4.5**	70	-51.0*	-6.8	1.3*	-4.8**	16.1**	-3.6*	0.39**	0.97**	0.24**	0.86
Manufacturing												
Labor intensive	1.9	-397	42*	7.4	-0.97	.00	-0.6	-1.6	-0.10	-0.09	-0.07*	0.13
Capital intensive	18**	-1,900**	116**	49**	-0.89	3.2*	-12**	-0.63	-0.12	-0.46*	-0.17**	0.86
Machinery	29**	-1,471*	38	33	-1.6	-1.5	-24**	-7.9*	1.0**	-1.1*	-0.27**	0.76
Chemicals	4.1**	-154	-16.0	3.0	-0.57	-0.30	-6.0*	-1.0	0.3**	-0.15	-0.04*	0.51

Source: Leamer (1984), 163.

Notes: GNP is scaled in thousands of dollars; capital is in millions.

* |t| > 1.

** |t| > 2.

three types of workers (distinguished by skill), four types of land (distinguished by climate), and three natural resources. So the number of goods equals the number of factors by construction.

The estimates of β_{jk} from regressing the 1975 net exports of each of these aggregates on the country factor endowments are shown in table 2.3. Leamer stresses that the point estimates should not be taken too seriously, but the sign pattern of the coefficients is of interest, especially when the estimates are significantly different from zero. In interpreting the coefficients as Rybczynski effects, it is noteworthy that an increase in the capital endowment is associated with a *rise* in the net exports of all the manufactured goods, and the same is true for an increase in the endowments of nonprofessional or illiterate workers. Conversely, increases in most types of land, or of professional and technical workers, are associated with *declines* in the net exports of the manufactured products. The interpretation of these results is that increases in land favor agriculture over manufacturing, and increases in professional and technical workers favor nontraded services over manufacturing.

Because the estimated coefficients can be positive or negative they do not provide any basis to *test* the HOV theorem. Instead, we have to rely on the overall fit of the regression: a *linear* relationship between endowments and trade, as in (2.6), is implied by the theory and can be assessed by the R^2 of the regressions. In the final column of table 2.3 we show the R^2 from each cross-country regression, and these vary from 0.13 to nearly unity. Admittedly, these fits are a *very weak* test of the theory, because we generally do not want to rely on only the R^2 of an equation to judge the validity of the theory underlying it. So while this early approach of Leamer (1984) provides valid results on the Rybczynski effects (which we shall refer to again in the next chapter), it is of limited usefulness in testing the HOV theorem.

We summarize our results so far in table 2.4, which indicates the data used and the basic methods of the various authors. These various “partial tests” have used only two

TABLE 2.4: Tests of the Heckscher-Ohlin Model

Authors	Data used				Method
	Trade	Technology	Factor Endowments		
Leontief (1953)	Yes	U.S.	No		Compared (K/L) ratio of exports and imports
Leamer (1980)	Yes	U.S.	No		Compared (K/L) ratio in production and consumption
Baldwin (1971)	Yes	U.S.	No		$T^i = \underbrace{A'}_{\text{data}} (\underbrace{V^i - s^i V^w}_{\text{data}})$
Leamer (1984)	Yes	No	Yes		$T^i = \underbrace{A^{-1}}_{\text{data}} (\underbrace{V^i - s^i V^w}_{\text{data}})$
Bowen, Leamer, and Sveikaukas (1987)	Yes	U.S.	Yes		Sign test and rank test
Trefler (1993a)	Yes	U.S.	Yes		Allow for productivity parameters π_k^i
Trefler (1995)	Yes	U.S.	Yes		Allow for productivity parameters δ^i (and more)
Davis and Weinstein (2001)	Yes	Many countries	Yes		Estimate A^i from data

sources of data, and we now turn to various complete tests of the HOV theorem, using all three sources of data (trade, technology, and factor endowments).

COMPLETE TEST OF THE HOV THEOREM

The first complete test of the HOV theorem was by Bowen, Leamer, and Sveikaukas (1987). They proposed two tests of equation (2.1): *a sign test*

$$\text{sign}(F_k^i) = \text{sign}(V_k^i - s^i V_k^w), i=1,\dots,C; k=1,\dots,M;$$

and a *rank test*

$$F_k^i > F_\ell^i \Leftrightarrow (V_k^i - s^i V_k^w) > (V_\ell^i - s^i V_\ell^w), i=1,\dots,C; k, \ell=1,\dots,M.$$

The first of these tests simply compares the sign pattern on the left and right sides of (2.1). With M factors and C countries, there are MC observations in total, and we are interested in what *percentage* of these have the same sign on the two sides of the equation. Notice that a completely random pattern of signs such as obtained by flipping a coin would still generate correct signs 50 percent of the time in a large sample. Therefore, the sign test must do considerably better than this in order to conclude that the HOV theorem is successful.

The rank test involves a pairwise comparison of all factors for each country, so there are $M(M-1)/2$ pairs for each of C countries. If the computed factor contents of one factor exceeds that of a second factor, then we check whether the relative abundance of that first factor also exceeds the relative abundance of the second factor. Again, a completely random assignment of factor abundance and relative endowments would imply that in 50 percent of the comparisons in a large sample, the rank test would be satisfied, so we would hope that the actual data perform considerably better than this.

In fact, the sign and rank test both fail miserably when confronted with actual data. Using data for twenty-seven countries and twelve factors, Bowen, Leamer, and Sveikaukas found that the sign test was satisfied in about 61 percent of the cases, and the rank test was satisfied in about 49 percent of them: about what we would expect from flipping a coin. These negative results were confirmed in later calculations by Trefler (1995), with a sample of thirty-three countries and nine factors, and his results for each country are shown in table 2.5. Overall, the sign test was successful in only 50 percent of the cases when using U.S. technology to measure the matrix A (column (2) in table 2.5).¹³ The rank test was successful in only 60 percent of cases (column 4). This lack of success indicates that some assumption of the HOV model is drastically wrong!¹⁴

¹³Table 2.5 corrects the expenditure shares s^i for trade imbalance. From note 6, $s^i = p'D^i/p'D^w$. Noting that $GDP^i = p'Y^i = p'D^i + p'T^i$, it follows that when $p'T^i \neq 0$ due to trade imbalances, we should measure the shares by $s^i = (GDP^i - p'T^i)/GDP^w$.

¹⁴Rather than conduct this test over the entire sample of countries, we might instead restrict attention to the factor content of trade for northern versus southern countries, whose factor endowments are quite different. Taking this approach, Debaere (2003) obtains much greater success with the sign test, which is reformulated to compare the difference in endowments of two factors with the difference in their factor content of trade. Then comparing northern and southern countries using Trefler's (1995) data, Debaere finds that the reformulated sign test is successful in 84 percent of cases for capital relative to labor, whereas for northern countries it is successful in 60 percent of cases and for southern countries in only 49 percent of cases.

TABLE 2.5: Results from the HOV model

GDP per capita	SIGN HOV		RANK HOV		ESTIMATES		
	F	F^{δ}	F	F^{δ}	δ^i	t-statistic	
Country	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bangladesh	0.04	0.33	0.78	0.75	0.78	0.03	47.71
Pakistan	0.08	0.33	0.67	0.72	0.78	0.09	32.10
Indonesia	0.11	0.22	0.67	0.67	0.67	0.10	39.51
Sri Lanka	0.12	0.22	0.56	0.42	0.67	0.09	14.85
Thailand	0.16	0.22	0.67	0.69	0.72	0.17	23.80
Colombia	0.21	0.33	0.89	0.81	0.86	0.16	18.41
Panama	0.23	0.33	0.78	0.56	0.78	0.28	3.24
Yugoslavia	0.30	0.56	0.67	0.44	0.61	0.29	11.35
Portugal	0.30	0.22	0.78	0.53	0.58	0.14	9.63
Uruguay	0.31	1.00	0.11	0.72	0.53	0.11	19.46
Greece	0.35	0.11	0.56	0.47	0.75	0.45	4.63
Ireland	0.39	0.67	0.44	0.53	0.39	0.55	2.91
Spain	0.41	0.22	0.78	0.39	0.69	0.42	9.40
Israel	0.60	0.67	0.89	0.39	0.69	0.49	2.91
Hong Kong	0.61	0.67	0.89	0.83	0.72	0.40	4.12
New Zealand	0.62	0.44	0.22	0.53	0.61	0.38	7.89
Austria	0.65	0.56	0.67	0.53	0.47	0.60	3.03
Singapore	0.66	0.56	1.00	0.61	0.61	0.48	2.11
Italy	0.66	0.67	0.33	0.78	0.67	0.60	7.16
U.K.	0.66	0.67	0.78	0.58	0.64	0.58	8.04
Japan	0.66	0.78	0.67	0.78	0.78	0.70	7.15
Belgium	0.67	0.67	0.78	0.61	0.53	0.65	2.73
Trinidad	0.69	0.67	1.00	0.50	0.53	0.47	1.25
Netherlands	0.69	0.44	0.44	0.53	0.47	0.72	2.66
Finland	0.70	0.33	0.44	0.47	0.50	0.65	2.17
Denmark	0.72	0.44	0.44	0.53	0.42	0.73	1.92
West Germany	0.73	0.56	0.67	0.81	0.78	0.78	3.80
France	0.73	0.33	0.33	0.08	0.22	0.74	4.84
Sweden	0.75	0.44	0.44	0.67	0.36	0.57	4.09
Norway	0.82	0.44	0.44	0.61	0.78	0.69	1.80
Switzerland	0.91	0.89	0.89	0.56	0.47	0.79	1.41
Canada	0.95	0.56	0.22	0.89	0.56	0.55	9.82
U.S.A.	1.00	0.89	0.56	0.92	0.72	1.00	
All Countries	0.50	0.62	0.60	0.62			

Source: Trefler (1995) and empirical exercise 2.1 and 2.2.

Notes: Column (1) is per capita GDP relative to U.S. per capita GDP. Columns (2) and (4) report the results of the sign and rank tests, assuming that all countries have the U.S. technology. Columns (3) and (5) report the results of the sign and rank tests, allowing for uniform technological differences δ^i across countries. Column (6) reports the estimates of δ^i and column (7) their asymptotic t-statistic for the null hypothesis $\delta^i = 1$.

Bowen, Leamer, and Sveikauskas along with Trefler (1995) perform several diagnostic tests on the data to determine which assumptions of the HOV model is most likely to be responsible for its failure, and conclude that the assumption of *equal technologies across countries* is especially bad. Accordingly, they develop an extended version of the HOV model that allows for different technologies across countries—a proposal originally made by Leontief (1953) as one explanation for the “paradox.”

MODELING DIFFERENT TECHNOLOGIES ACROSS COUNTRIES

There are two ways that technological differences can be introduced into the HOV model. One approach is to model *the productivity of factors* in different countries, whereas another approach is to model *differences in the factor requirements matrix A*. These approaches are closely related, of course: saying that a factor is 10 percent less productive in one country is the same as saying that 10 percent more of that factor is needed per unit of production. But thinking about them as distinct will be convenient for our discussion.

Trefler (1993a) takes the first approach and allows *all factors in every country to differ in their productivities*. The only exception to this is for the United States, which he uses as the benchmark country, with factor productivities normalized at unity. So let π_k^i denote the productivity of factor k in country i relative to its productivity in the United States. In terms of efficiency units, the *effective endowment* of factor k in country i becomes $\pi_k^i V_k^i$. Let A now denote the amount of effective factors needed per unit of output in each industry. We continue to assume that factor price equalization (FPE) holds in terms of effective factor prices, so with identical technologies, the matrix A is the same across countries. Thus, we continue to measure the factor content of trade as $F^i \equiv AT^i$. Then the HOV equation (2.1) is rewritten in terms of effective factor endowments as:

$$F_k^i = \pi_k^i V_k^i - s^i \sum_{j=1}^C \pi_k^j V_k^j, \quad i=1, \dots, C; \quad k=1, \dots, M. \quad (2.7)$$

There are $M \times C$ equations in (2.7), and $M \times (C - 1)$ productivity parameters. These equations are not independent, however. For any factor k , if we sum (2.7) across countries $i=1, \dots, C$ then we must obtain zero on both sides: exports equal imports for the world, even when measured in terms of factor contents. So we can drop the equations for one country, and we do so for the United States. This leaves $M \times (C - 1)$ equations in $M \times (C - 1)$ parameters. These equations will be independent “almost everywhere,” that is, for almost all sets of data on endowments and factor contents of trade. Furthermore, the productivity parameters enter (2.7) *linearly*, so the $M \times (C - 1)$ equations in (2.7) can be inverted to solve for these uniquely. Thus, we have established the following theorem.

THEOREM (TREFLER 1993A)

Allowing for all factors in all but one country to differ in their productivities π_k^i , for almost all datasets there will be a solution for productivities π_k^i such that the HOV equation (2.7) holds *with equality* for $i=1, \dots, C; \quad k=1, \dots, M$.

The unusual feature of this result is that the HOV equation no longer becomes testable: it holds as an identity by the choice of productivity parameters! How, then, are we to judge whether the HOV model is a reasonable description of reality? Trefler recommends two methods: first, we need to check whether the productivity parameters are *positive* (there is nothing in the inversion of (2.7) that guarantees this); second, we might compare these parameters to other economic data to evaluate how “reasonable” the productivity parameters are. For example, it makes sense to compare the labor

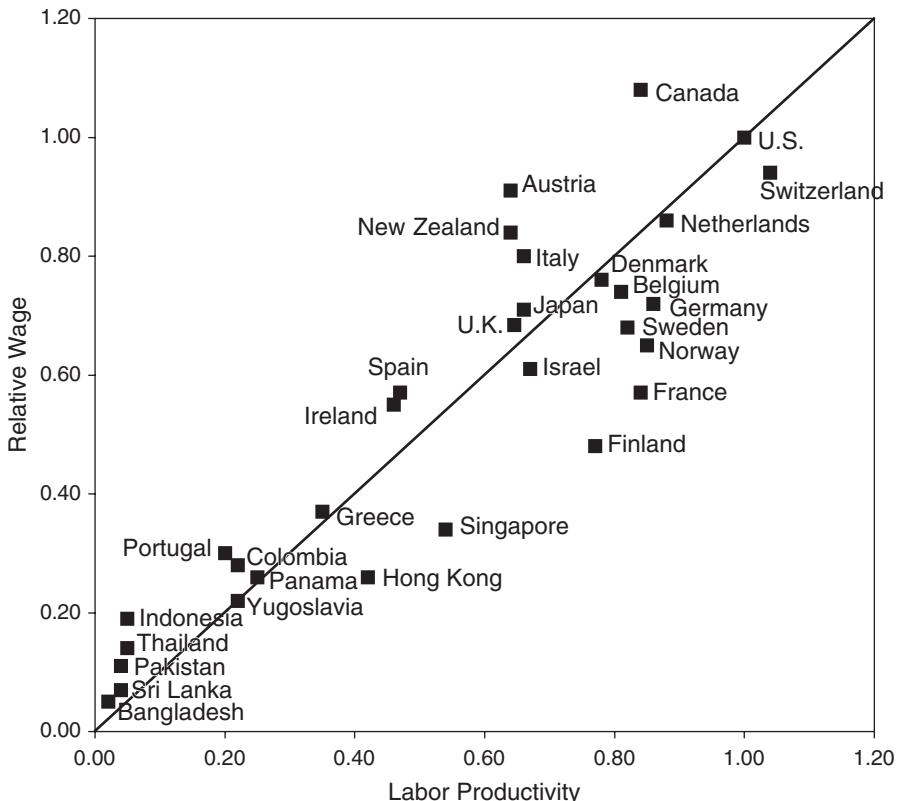


Figure 2.4

productivity parameters to wages across countries. If factor price equalization holds in terms of *effective* endowments, then it should be the case that the productivity parameters π_k^i match the factor prices of country i quite closely.

In figure 2.4, we show the results that Trefler (1993a) obtains for *labor*.¹⁵ On the horizontal axis we measure the estimated productivities π_k^i , and on the vertical axis are data on wages across countries (both of these are relative to the United States). It is clear that the observations on these two variables across countries follow each other quite closely: the correlation between estimated π_k^i and actual wages is 0.9. Since we think that cross-country differences in wages should indeed reflect workers' productivities, this close correspondence between the two in figure 2.4 lends support to Trefler's extension of the HOV model.¹⁶

Turning to the second method of introducing productivity differences, Trefler (1995) allows the factor requirements matrix A^i to differ across countries, while assuming that the factor endowments do not differ in efficiency units. In this analysis,

¹⁵In empirical exercise 2.3 you are provided with Trefler's data and are asked to compute the productivities π_k^i .

¹⁶Gabaix (1999) disagrees, and argues that even when the labor content of trade is zero, the solution for the productivities of labor π_k^i from equation (2.7) is simply GDP per capita, as you are asked to show in problem 2.1. Because the labor content of trade is small, it is therefore not surprising that the solution for π_k^i is highly correlated with wages, which are highly correlated with GDP per capita.

he chooses to *restrict* the range of differences in matrices A^i so that the HOV equations give *less than a perfect fit*. This means that we are back in the traditional econometric approach: choosing the parameters of A^i so as to minimize the sum of squared residuals of the HOV equations. To implement this, we need to model A^i in terms of some productivity differences across countries. The initial case that Trefler considers is where the matrices A^i differ by a *uniform amount* across countries:

$$\delta^i A^i = A^{\text{US}}, \quad (2.8)$$

which implies $A^i = A^{\text{US}}/\delta^i$. With $\delta^i < 1$, this means that $A^i > A^{\text{US}}$, so that country i is *less productive* and requires more labor, capital, and other resources for a unit of production relative to that of the United States.

With the assumption (2.8), we need to rederive the HOV equation. From the assumption of full employment, we still have $A^i Y^i = V^i$. We continue to make the assumption of identical and homothetic tastes across countries, so the factor content of trade is computed as

$$A^i T^i = A^i Y^i - A^i D^i = V^i - A^i (s^i D^w) = V^i - A^i \left(s^i \sum_{j=1}^C Y^j \right). \quad (2.9)$$

To make further progress with (2.9), let us multiply both sides by δ^i . On the left, we would obtain $\delta^i A^i T^i = A^{\text{US}} T^i$. That is, we define $F^{i\text{US}}$ as the factor content of trade for country i , computed by using the U.S. factor requirements matrix. (Recall that this is what Leontief used, as well as all other researchers discussed above). By also multiplying the right-hand side of (2.9) by δ^i , the matrix A^i appearing in the last term becomes instead $\delta^i A^i = A^{\text{US}}$. Since this term now does not depend on the country i , it can be brought *inside* the summation on the far right of (2.9), to obtain $\sum_{j=1}^C A^{\text{US}} Y^j = \sum_{j=1}^C \delta^i A^i Y^j = \sum_{j=1}^C \delta^i V^j$, where the final equality makes use of the full-employment conditions in country $j=1, \dots, C$. Substituting these results back into (2.9), we have therefore derived

$$F^{i\text{US}} \equiv A^{\text{US}} T^i = \delta^i V^i - \left(s^i \sum_{j=1}^C \delta^j V^j \right). \quad (2.10)$$

This becomes the statement of the HOV theorem when we allow for uniform technological differences across countries: the factor content of trade, measured with the U.S. technology on the left, should equal the relative factor endowments on the right, where these are each adjusted by the technological parameter of each country. Unlike equation (2.7), however, (2.10) will certainly not hold exactly. We therefore introduce an additive error and choose the parameters δ^i to minimize the sum of squared residuals of this equation.

The estimates for δ^i obtained by Trefler (1995) from equation (2.10) are shown in column (6) of table 2.5, and column (7) reports their asymptotic t -statistic to test the null hypothesis $\delta^i = 1$. Clearly, most countries have a technological development that is significantly less advanced than that in the United States. Furthermore, there is a close relationship between δ^i in column (6) and each country's GDP per capita relative to that of the United States reported in column (1): the correlation between them is 0.89. If we presume that the differences in GDP per capita across countries are principally explained by technology, then this strong correlation lends support to the model.

How should we compare the models estimated by Trefler 1993a and Trefler 1995? The complete set of cross-country factor productivity differences π_k^i allows for a perfect fit in the HOV equations, but this is because there are so many “free” parameters. Allowing only scalar productivity differences δ^i is much more parsimonious, with $(C - 1)$ free parameters rather than $M(C - 1)$. One approach is to ask how much the fit of the HOV equations improves by this more limited choice of parameters. That is, we can compare the variance of the left side and estimated right side of the HOV equation in (2.1), (2.7), or (2.10) and use this as a measure of the R^2 of the model. In conjunction with the “reasonableness” of the estimates themselves, this can be used to evaluate the model.

Consider the original HOV equations (2.1), without any productivity parameters at all. For Trefler’s data, the variance of the factor contents F_k^i relative to the variance of the country endowments $V_k^i - s^i V_k^w$ turns out to be only 0.032! Trefler (1995) refers to this as the mystery of “missing trade”: measured factor contents of trade are very small in comparison with relative endowments $V_k^i - s^i V_k^w$. At the other extreme, when we allow for the complete set of cross-country factor productivity differences π_k^i , we obtain a perfect fit in the HOV equations (2.7), so that the ratio of the left side and right side is unity. In between these, we can consider the uniform productivity differences δ^i , chosen to minimize the sum of squared residuals in (2.10). Then the R^2 from this equation turns out to be 0.486. In other words, nearly *one-half* of the “missing trade” is explained by allowing for uniform productivity differences across countries. This is an encouraging result! Recomputing the sign test—which compares the signs of the left- and right-hand side of (2.10)—it turns out that the correct sign is obtained in 62 percent of cases (column (3) in table 2.6), rather than the 50 percent obtained without the scalar productivity differences δ^i .¹⁷

Trefler (1995) goes on to consider other modifications to the HOV framework, including a *limited* set of factor productivity differences (allowed to differ between rich and poor countries), distinguishing consumption versus investment goods, and a “home country” bias in consumption. He concludes that the contribution of these to further explaining “missing trade” is more limited, and he prefers the initial model that allows for uniform productivity differences.¹⁸

USING DATA ON DIFFERENT TECHNOLOGIES ACROSS COUNTRIES

The reader will notice in table 2.4 that most complete tests of the HOV theorem, using all three types of data, have still relied on the U.S. technology matrix. This is again due to data limitations: factor requirements for other countries, concorded to the same set of factors and industries, have only recently become available.¹⁹ With the availability of this data, the next logical step is to make use of it to construct A^i across countries, rather than only using the U.S. technology matrix. This is the approach taken by Davis and Weinstein (2001). We will defer a discussion of their approach until the

¹⁷If instead we compute the sign test while *weighting* each observation in (2.10) by the factor content of trade, then the same signs are obtained in 71 percent of cases without allowing for any technological differences between countries, and 93 percent of cases when allowing for the scalar differences δ^i .

¹⁸Another extension not considered by Trefler is nonhomothetic tastes, which is discussed in the conclusions to this chapter.

¹⁹See note 8 of this chapter.

next chapter, since it relies on material that we have not yet covered. We note at this point, however, that Davis and Weinstein are careful to *estimate* the differences in the technology matrices A^i across countries, rather than using the *actual data* for A^i , since the latter approach would run the risk of having the constructed factor content of trade equal to the countries' relative endowments by construction. This is the same issue that Trefler (1993a) had, where he allowed for enough productivity parameters for the HOV equations to fit perfectly. If the actual technology matrices are used to construct the factor content of trade, the HOV equation can again hold as an identity, as we now show.

In addition to using the actual data on A^i across countries, let us assume that the output of every good is exported to each country in proportion to the purchasing country's GDP.²⁰ We now keep track of the *bilateral* trades between countries, so let X^{ij} denote the *gross exports of goods from country i to country j* . (Note that this is related to the net exports of country i by $T^i = \sum_{j \neq i} X^{ij} - \sum_{j \neq i} X^{ji}$). Using our assumption regarding bilateral exports, these may be written as

$$X^{ij} = s^j Y^i, \quad (2.11)$$

where $s^j = \text{GDP}^j / \text{GDP}^w$ is the ratio of country j 's GDP to world GDP.²¹ In order to measure the factor content of these bilateral exports, we use the primary factor requirements of the exporting country i , and define $F^{ij} = A^i X^{ij}$ as the factor contents of exports from country i to country j . It follows from (2.11) that

$$F^{ij} \equiv A^i X^{ij} = s^j A^i Y^i = s^j V^i, \quad (2.11')$$

where the final equality follows from full employment.

Using simple algebra we can rewrite the relative endowments of country i as,

$$V^i - s^i \left(\sum_j V^j \right) = (1 - s^i) V^i - s^i \left(\sum_{j \neq i} V^j \right) = \left(\sum_{j \neq i} s^j \right) V^i - s^i \left(\sum_{j \neq i} V^j \right). \quad (2.12)$$

Now make use of the factor content of exports in (2.11'), $F^{ij} = s^j V^i$. This term is summed over all destination countries $j \neq i$, and appears as the first term on the right side of (2.12), which is interpreted as the *factors used in country i to produce exports for all countries $j \neq i$* . Similarly, the second term on the right side of (2.12) is the *factors used in every country $j \neq i$ to produce exports to country i* . Thus, using (2.11') to replace the terms on the right of (2.12), we have shown the following theorem:

THEOREM (TREFLER AND ZHU 2010)

If the output of every good is exported to each country in proportion to the purchasing country's GDP, and the factor content of bilateral trade is measured with the exporting countries' technologies, then

²⁰This assumption is somewhat stronger than identical and homothetic tastes, and is modeled by the "gravity equation." We will discuss this equation in detail in chapters 5 and 6.

²¹We are ignoring trade imbalances, as discussed in note 13 of this chapter.

$$V^i - s^i \left(\sum_j V^j \right) = \left(\sum_{j \neq i} F^{ij} \right) - \left(\sum_{j \neq i} F^{ji} \right). \quad (2.13)$$

The first term on the right in (2.13) is the factor content of exports from country i to all countries $j \neq i$, and the second term is the factor content of imports from all countries $j \neq i$ to country i . The factor content of country i 's multilateral net exports is measured by the difference between these, and (2.13) therefore shows that the factor content of country i 's multilateral net exports (on the right) equals its relative factor endowments (on the left). We stress that this HOV equation is *automatically satisfied* if the hypotheses of the theorem holds true.

This result, due to Trefler and Zhu (2010), serves as a warning about using actual technology data to construct the factor content of trade. As in Trefler (1993a), which allowed for unlimited differences in factor productivities across countries, using the actual technologies can quickly lead us to an *identity* between the factor content of trade and country's relative endowments, as in (2.13). This would not then be a test of the theory. Trefler and Zhu (2010) further show how this result can be extended to accommodate trade in intermediate inputs, and rightfully suggest that future work needs to take this identity into account.

OTHER TESTS OF TRADE

We conclude this chapter by briefly considering some other approaches to testing models of trade. We begin with the most general test of the “law of comparative advantage,” as proposed by Deardorff (1984b), which does not depend on the Heckscher-Ohlin model. While this test is completely general, it unfortunately relies on information on *autarky* prices, which is seldom available. Accordingly, we move toward alternative tests of the HO model that do not rely on FPE, and also do not require information on autarky prices. Such a test has been proposed by Helpman (1984a), building on the earlier work of Brecher and Choudhri (1982b). We shall review this test and the empirical application by Choi and Krishna (2004).

Deardorff's test of comparative advantage can be motivated with the simple two-by-two-by-two model with which we began this chapter. In figure 2.2(a), the home country has its production shifted from point A to point B , and its consumption shifted from point A to point C , when moving from autarky to free trade. But both the shift in production and the shift in consumption can be linked by some inequalities that follow from profit maximization and revealed preference. To develop these, let us denote the autarky production of country i by the vector Y^{ia} , the autarky price vector by p^{ia} , and the free-trade production by Y^i . Then profit maximization ensures that $p^{ia'} Y^{ia} \geq p^{ia'} Y^i$. In other words, the value of production in autarky must exceed the value obtained if producers choose the feasible (but not optimal) point Y^i instead. In autarky, production equals consumption, so that $p^{ia'} Y^{ia} = p^{ia'} D^{ia}$. Now consider the consumption vector chosen under free trade, D^i . As is evident from figure 2.2, and will be proven carefully in chapter 6, the representative consumer is *better off* with this free-trade consumption than in autarky: there are gains from trade. It follows from revealed preference that the representative consumer in country i could not afford the free-trade consumption vector at the autarky prices, that is, $p^{ia'} D^i \geq p^{ia'} D^{ia}$. Pulling together these various inequalities, we therefore have

$$p^{ia'} D^i \geq p^{ia'} D^{ia} = p^{ia'} Y^{ia} \geq p^{ia'} Y^i. \quad (2.14)$$

Comparing the first and last of these expressions, we see that $p^{ia'}(D^i - Y^i) \geq 0$. But since the net exports of country i are denoted by $T^i = Y^i - D^i$, it follows that $p^{ia'} T^i \leq 0$. In addition, we have implicitly assumed above that trade is balanced for all countries, meaning that $p' T^i = 0$, where p is the vector of free-trade prices. Therefore, we have derived the key result of Deardorff (1984b):

$$(p^{ia} - p)' T^i \leq 0. \quad (2.15)$$

In other words, countries will tend to export those goods ($T_j^i > 0$) whose autarky prices are below their trade prices, and import those goods ($T_j^i < 0$) whose autarky prices are above their trade prices. This certainly held in our two-by-two-by-two model at the beginning of the chapter, where the home country exported good 1 (with $p > p^a$) and imported good 2. The inequality in (2.15) shows that this type of result holds in general, with any number of goods and factors.

While we have derived this result for a comparison of autarky and free trade, Deardorff (1984b) shows that it also holds for a comparison of autarky with any form of trade that may be *restricted* by the use of tariffs. So we can measure the trade prices and quantities using actual data, but as we have already mentioned, it is difficult to measure the autarky equilibrium. One unique case where this equilibrium was observed in historical data was Japan in the mid-1850s, which then rapidly moved to more open trade with the Meiji restoration in the late 1860s. Bernhofen and Brown (2004a) use data on Japan's pattern of trade from 1868 to 1872, together with various sources of autarky price information from around the 1850s, to directly test the prediction in (2.15). They find that this prediction is strongly confirmed, thereby supporting this general statement of comparative advantage.

To develop a test that does not rely on autarky information, let us return to the assumptions of the HO model: identical technologies across countries, with constant returns to scale. In contrast to our initial formulation of the HOV model, however, we no longer assume that FPE holds. We are then allowing the countries to be in different cones of diversification. When countries trade freely in goods at the price vector p , we denote the GDP of country i by $G(p, V^i)$, where V^i are the endowments of country i . Note that under our assumption of identical technologies, we do not need to index the GDP function itself by the country i . We again keep track of the *bilateral* trades between countries, and let X^{ij} denote the *gross exports* of goods *from country i to country j* . In order to measure the factor content of these exports, we use the primary factor requirements of the *exporting* country i , and define $F^{ij} \equiv A^i X^{ij}$ as the factor contents of exports from country i to country j , measured with the exporting country's technology.

Suppose, however, that rather than having country j import X^{ij} of goods, we directly allowed it to import the amount F^{ij} *in factors*. Under our maintained assumption that technologies are the same in both countries, it would be *feasible* for country j to directly produce the amount X^{ij} in goods, using the same methods as employed in country i . In general, however, country j could do better than that: because factor prices differ between the countries, the importing country might not want to use the same methods as the exporter, and will end up producing more by choosing methods attuned to its own factor prices. Thus, if we evaluated GDP in country j using the hypothetical factor imports F^{ij} , it must *exceed* the value obtained with the import of goods:

$$\begin{aligned}
p'(Y^j + X^{ij}) &\leq G(p, V^j + F^{ij}) \\
&\leq G(p, V^j) + [\partial G(p, V^j)/\partial V^j]' F^{ij} \\
&= p' Y^j + w^j' F^{ij}.
\end{aligned} \tag{2.16}$$

The second line of (2.16) follows because the GDP function is concave in the factor endowments, and the last line follows because the derivative of the GDP function with respect to endowments equals the vector of factor prices w^j .

Taking the difference between the first and last lines of (2.16), we therefore have $p' X^{ij} \leq w^j' F^{ij}$, so that the value of bilateral exports in goods is less than the value of embodied factors, using the *importing* country's factor prices. If instead we had used the *exporting* country's factor prices, however, then with constant returns to scale the value of output would just equal the value of factors used, so that $p' X^{ij} = w^i' F^{ij}$. Combining this with the previous inequality we therefore have

$$(w^j - w^i)' F^{ij} \geq 0. \tag{2.17}$$

Furthermore, by repeating this exercise but focusing on net exports from country j to country i , measured using *country j*'s technology, we can derive that

$$(w^j - w^i)' F^{ji} \leq 0. \tag{2.18}$$

Putting together these two inequalities, we therefore obtain the key result of Helpman (1984a):

$$(w^j - w^i)' (F^{ij} - F^{ji}) \geq 0. \tag{2.19}$$

These inequalities can be interpreted as saying that factors embodied in trade should flow toward the country with higher factor price, that is, if factor k has a higher price in country j , $(w_k^j - w_k^i) > 0$, then we should have $F_k^{ij} > 0$ in (2.17) or $F_k^{ji} < 0$ in (2.18) for at least some factors k , so that (2.17)–(2.19) hold when we add up across factors.

To summarize, (2.17)–(2.19) give us testable hypotheses on bilateral factor content of trade between countries that *do not* depend on FPE, though they do presume identical technologies. Choi and Krishna (2004) implement this test for bilateral trade flows between eight countries (Canada, Denmark, France, Germany, Korea, Netherlands, the United Kingdom, and the United States) using data for 1980. They considered two different ways of measuring the rental price of capital, and two different levels of disaggregating labor. Considering first the one-way trade flows in (2.17) and (2.18), they find that these equations have the correct sign in about 52 percent to 55 percent of cases, depending on the methods used for measuring factors. Even when a one-way flow has the incorrect sign, but is small in magnitude, it is still possible that the two-way factor flow in equation (2.19) will have the correct sign. This is confirmed by Choi and Krishna, who find that the *combined* equation (2.19) is satisfied for 72 percent to 75 percent of cases, which is a quite respectable success rate.

Choi and Krishna (2004) also show how (2.17) and (2.18) can be generalized to allow for scalar technological differences across countries, much like in (2.8). In that case, they find that the one-way trade flows in (2.17) and (2.18) have the correct sign in 55 percent to 59 percent of cases, while the two-way trade flows in (2.19) have the correct sign in 79 percent of cases. It is somewhat surprising that correcting for scalar technology differences between countries did not lead to a greater improvement in the

test results, but then again, this correction improved the “sign test” for Trefler (1995) from 50 percent to only 62 percent. Trefler’s acceptance of the scalar productivity differences as an *improvement* over the conventional HOV model was based on other criterion, and in particular, their ability to help explain the “missing trade.” For the bilateral flows analyzed by Choi and Krishna, the fact that the combined equation (2.19)—emphasized by Helpman (1984a)—is satisfied in about three-quarters of cases indicates support for the theory and suggests that dropping FPE is an important direction for further research.

CONCLUSIONS

We started this chapter with the Leontief paradox and, after explaining this away using the results of Leamer (1980), went on to argue that the complete tests of the HOV model fail sadly under the conventional assumptions of this model: identical homothetic tastes and identical technologies with FPE across countries. As we begin to loosen these assumptions, the model performs better, and when we allow for unlimited differences in productivities of factors across countries, as in Trefler (1993a), then the resulting HOV equations will hold as an identity. Between these two extremes, Trefler (1995) shows that a parsimonious specification of technological differences between countries—allowing for a uniform difference from the technology of the United States—is still able to greatly improve the fit of the HOV equation. Research such as Davis and Weinstein (2001), which we review in the next chapter, generalizes these technological differences and further explains how we account for the differences between the factor content of trade and relative endowments.

So what is left for further research? While it is dangerous to predict the course of future research, we can suggest two areas that deserve further attention. First, it is worth making a distinction between *accounting* for global trade volumes and *testing hypotheses* related to trade. When we attempt to match the right- and left-hand sides of the HOV equation, such as by introducing productivity parameters, we are engaged in an accounting exercise. With enough free parameters this is bound to lead to equality of the HOV equation, as we found in (2.13), which uses the actual technologies of each country but relies on an assumption about exports. On the other hand, when we test hypotheses such as (2.17)–(2.19), we are making use of economic behavior: the GDP function is concave in the factor endowments, leading to (2.17)–(2.19), if and only if producers economize on factor inputs when their prices rise. There seems to be a difference between these two approaches, and ordinarily in economics we are more interested in testing economic behavior.

Second, even if we accept that the HOV equation can fit perfectly by allowing for sufficient differences between technologies across countries, this begs the question, *Where do these differences in technology come from?* In the original work of Heckscher and Ohlin, they rejected the technology differences assumed by Ricardo in favor of a world where knowledge flowed across borders. We have since learned that this assumption of technological similarity across countries was empirically false at the time they wrote (see Estevadeordal and Taylor 2002a, 2002b), as well as in recent years (Trefler 1993a, 1995; Davis and Weinstein 2001). So we are back in the world of Ricardo, where technological differences are a major determinant of trade patterns. Such differences can hardly be accepted as exogenous, however, and surely must be explicable based on underlying causes. Increasing returns to scale might be one explanation,

as will be discussed in chapters 5 and 6, and this has been incorporated into the HOV framework by Antweiler and Trefler (2002). Economy-wide increasing returns are also suggested by the literature on “endogenous growth,” which we will discuss in chapter 11. Other authors have argued that geography and climate (Sachs 2001), or colonial institutions (Acemoğlu, Johnson, and Robinson 2001), or social capital (Jones and Hall 1999), or the efficiency with which labor is utilized (Clark and Feenstra 2003) must play an important role in the technological capabilities of countries.

Beyond differences in technologies across countries, we might also investigate differences in *tastes*. Hunter and Markusen (1988) argue that differences in per capita income are indeed an important determinant of trade, and show how the HOV equations can be extended to allow for a nonhomothetic linear expenditure system.²² Caron, Fally, and Markusen (2014) further show that nonhomothetic preferences in the context of an HOV production structure can resolve a number of puzzles in international trade. For example, the extent of “missing trade” in the HOV model is considerably reduced by nonhomothetic preferences, once we recognize that the skilled-labor intensity of good’s production is correlated with the income elasticity in consumption. Since rich countries are skilled-labor abundant, those countries consume relatively more of the skill-intensive goods and services that they are specialized in producing, and so they trade more with one another than with poorer countries.²³ Other recent work shows that nonhomothetic preferences contribute in an interesting way to the prices charged across countries (Simonovska 2015) and to the gains from trade enjoyed by consumers of differing incomes within and between countries (Fajgelbaum and Khandelwal 2014).

PROBLEMS

- 2.1 Suppose that the factor content of trade is zero on the left side of equation (2.7). Then solve for the implied values of the productivities π_k^i in one country relative to another.
- 2.2 An assumption of the Heckscher-Ohlin-Vanek model, as outlined in this chapter, is that tastes are identical across countries and also homothetic. The latter is an unrealistic assumption because of Engel’s Law: expenditures on food are a declining share of total income. Following Hunter and Markusen (1988), we can introduce nonhomothetic tastes using a linear expenditure system. Denoting per capita consumption of good i in any country j by d_i^j , we suppose that the utility function is given by

$$U^j = \prod_{i=1}^N (d_i^j - \bar{d}_i)^{\phi_i}, \text{ with } \sum_{i=1}^N \phi_i = 1.$$

The parameters $\bar{d}_i \geq 0$ in the utility function are interpreted as the “minimum consumption” of each commodity $i=1, \dots, N$. They are assumed to be the same across countries, as are the parameters $\phi_i > 0$.

²²You are asked to work through this case in problem 2.2.

²³Fieler (2011) also discusses how nonhomothetic preferences can improve the fit of trade models. She introduces these preferences into the model of Eaton and Kortum (2002), which is discussed in chapter 6.

- (a) Assuming that *per capita* income I^j is large enough to afford the “minimum consumption,” show that per capita demand for each commodity in country j is given by

$$d_i^j = \bar{d}_i + \phi_i (I^j - \sum_{k=1}^N p_k \bar{d}_k) / p_i,$$

where $p_i, i = 1, \dots, N$, are the commodity prices.

- (b) Assuming that prices are the same across countries due to free trade, we can normalize them at unity, and rewrite demand in the more compact form

$$d_i^j = \delta_i + \phi_i I^j,$$

where $\delta_i \equiv (\bar{d}_i - \phi_i \sum_{k=1}^N \bar{d}_k)$ are the values of “minimum consumption” measured relative to their mean. Interpret the coefficients δ_i .

- (c) Multiply the per capita demand by country population L^j , to obtain total demand, $D_i^j = L^j d_i^j = \delta_i L^j + \phi_i E^j$, where $E^j \equiv L^j I^j$ denotes *total expenditure* in country j . We can write this in vector notation as, $D^j = \delta L^j + \phi E^j$. Using this demand function, rederive the HOV equation (2.1), and interpret the new equation.

EMPIRICAL EXERCISES

In these exercises, you will reproduce some of the empirical results from Trefler (1993a, 1995). To complete the exercise, the Excel file “hov_pub.csv” should be stored in the directory c:\Empirical_Exercise\Chapter_2\hov_pub.csv. After this, run the STATA program “hov_pub.do”, which will create a new STATA data file “trefler.dta”. Then do the following exercises:

- 2.1 Given identical technologies across countries, run the program “sign_rank_1.do” to conduct the sign test, rank test, and test for missing trade. Use the results in “sign_rank_1.log” to replicate columns (2) and (4) in table 2.5.
- 2.2 Given uniform technological differences across countries, run the program “sign_rank_2.do” to redo the sign test, rank test, and missing trade. Use the results in “sign_rank_2.log” to replicate column (3) and (5), given column (6) in table 2.5.
- 2.3 Allowing all factors in each country to have different productivities, now run the program “compute_pi.do” to compute factor productivities π_k^i as Trefler 1993a. Note that there are nine factors in the original dataset, but these are now aggregated to just four factors, which are labor (endowment 1), capital (endowment 2), cropland (endowment 3), and pasture (endowment 4). Using the results in “pi.log” or alternatively in the data files “pi_1.dta, pi_2.dta, pi_3.dta, pi_4.data”, answer the following:
 - (a) Which factor has the most *negative* productivities estimated?
 - (b) What is the correlation between the estimated labor productivity and the productivities of other factors? What is the correlation between each factor productivity and GDP per capita (which you can find in the file “trefler.dta”)?

Many Goods and Factors

The Heckscher-Ohlin-Vanek (HOV) model we studied in the previous chapter allowed for many goods and factors, but for much of that discussion we maintained the assumption of factor price equalization. This theorem was proved in chapter 1 only for the case of two goods and two factors (and no factor intensity reversals). A natural question to ask, with many goods and factors, is whether factor price equalization will continue to hold? More generally, what happens to the other theorems we discussed in chapter 1—the Stolper-Samuelson and Rybczynski theorems? Are there some generalizations of these that apply to the many-goods and many-factors case? These are the topics we shall address in this chapter.

We begin with the case where the number of goods and factors is the same, the so-called “even” case. We will see that the factor price equalization (FPE) theorem naturally generalizes from the case of two goods and two factors to N goods and N factors. Versions of the Stolper-Samuelson and Rybczynski theorems also continue to hold, though not quite as strongly as we obtained in the two-by-two case.

Then we turn to “uneven” cases, beginning with two goods and three factors. We can think of the factors as labor, which is mobile between both industries, and a specific factor in each industry, such as capital or land, which does not move between them. This is the so-called “specific factors” or Ricardo-Viner model. It will become immediately clear that factor price equalization *does not* hold, but rather, that factor prices prevailing in a country depend on its factor endowments. The results of the Stolper-Samuelson and Rybczynski theorems are also modified, depending on whether we are considering a change in the mobile factor (labor) or the specific factors. The results we obtain for this case illustrate those that apply whenever there are more factors than goods.

Empirically, the case of an equal number of goods and factors or more factors than goods can be neatly analyzed using a country’s GDP function. We discuss how these functions can be estimated using data on a set of industries over time. These techniques were pioneered by Diewert (1974) and applied in an international context by Kohli (1978, 1991), Woodland (1982), and others.¹ Harrigan (1997) provides an application to the OECD (Organization for Economic Cooperation and Development) countries that we shall review, whereby the output shares of various industries are related to country prices, productivity, and factor endowments. By estimating these share equations, we can then compute the effect of endowment

¹Burgess (1974a, 1974b, 1976) provides estimates of an aggregate cost rather than GDP function for the United States, as we shall discuss.

changes on output shares, namely, the Rybczynski effects. This methodology allows us to explain the specialization of countries across various industries. In addition, the GDP function allows us to estimate the dual Stolper-Samuelson effects, as we shall discuss in this chapter and the next.

Finally, we consider the case of more goods than factors, starting with three goods and two factors. Using the results of Dixit and Norman (1980), we argue that there is a wide range of possible factor endowments across countries such that FPE continues to hold, provided that technologies are the same across countries. However, the amount of production occurring in each country is indeterminate when factor prices are equalized. A test for production indeterminacy has been implemented by Bernstein and Weinstein (2002), as we review.² An elegant generalization of the case with more goods than factor is when there is a *continuum* of goods, as in Dornbusch, Fischer, and Samuelson (1977) for the Ricardian model, as we shall discuss. Dornbusch, Fischer, and Samuelson (1980) extend this framework to the Heckscher-Ohlin case, and this model allows us to complete our presentation of empirical results from Davis and Weinstein (2001) that we began in the last chapter.

EQUILIBRIUM CONDITIONS

We will suppose that there are $i=1, \dots, N$ goods, and $j=1, \dots, M$ factors. The production functions are written $y_i = f_i(v_i)$, where $v_i = (v_{i1}, \dots, v_{iM})$ is the vector of factor inputs. As usual we assume that the production functions are positive, increasing, concave, and homogeneous of degree one for all $v_i \geq 0$. Denoting the vector of factor prices by w , the unit-cost functions are $c_i(w) \equiv \min\{w'v_i | f_i(v_i) \geq 1\}$, which are also positive, non-decreasing, concave and homogeneous of degree one for all $w > 0$. The zero-profit conditions are then stated as

$$p_i = c_i(w), \quad i=1, \dots, N. \quad (3.1)$$

The second set of equilibrium conditions is full employment of resources. As in chapter 1, we write $\partial c_i / \partial w = a_i(w)$ as the amount of factors used for *one unit* of production, and it follows that the total inputs used in industry i are $v_i = y_i a_i(w)$. We denote the elements of the vector $a_i(w)$ by $a_{ij}(w)$, $j=1, \dots, M$. Then the full-employment conditions are stated as

$$\sum_{i=1}^N a_{ij}(w) y_i = V_j, \quad j=1, \dots, M, \quad (3.2)$$

where V_j is the endowment of resource j . In matrix notation, let $A = [a_1(w)', \dots, a_N(w)']'$ denote the $M \times N$ matrix of primary factors needed for one unit of production in each industry, where the *columns* of this matrix measure the different industries $i=1, \dots, N$. The full-employment conditions (3.2) are then written compactly as

$$AY = V, \quad (3.2')$$

where Y is the $(N \times 1)$ vector of output and V is the $(M \times 1)$ vector of factor endowments.

²The empirical methods discussed in this chapter draw upon the survey by Harrigan (2003).

The equilibrium conditions (3.1) and (3.2) are $N + M$ equations in $N + M$ unknowns, namely, the factor prices $w_j, j=1, \dots, M$ and outputs $y_p, p=1, \dots, N$. To analyze these, it will be helpful to use the GDP function for the economy, which is defined as

$$G(p, V) \equiv \max_{v_i \geq 0} \sum_{i=1}^N p_i f_i(v_i) \quad \text{subject to } \sum_{i=1}^N v_i \leq V, \quad (3.3)$$

where $p = (p_1, \dots, p_N)$ and $V = (V_1, \dots, V_M)$ are the price and endowment vectors. To solve this maximization problem, we substitute the constraint (written with equality) into the objective function, and write the output of good 1 as $f_1(V - \sum_{i=2}^N v_i)$, so that the maximized value of GDP becomes, $G(p, V) = p_1 f_1(V - \sum_{i=2}^N v_i) + \sum_{i=2}^N p_i f_i(v_i)$. From the envelope theorem, we can differentiate this with respect to p and V while holding the optimal inputs choices v_i fixed, $i = 2, \dots, N$. Then we obtain

$$(a) \frac{\partial G}{\partial p_i} = f_i(v_i) = y_p$$

which is the output of industry i ; and

$$(b) \frac{\partial G}{\partial V_j} = p_1 \frac{\partial f_1}{\partial v_{1j}}$$

which is the factor price w_j .

Furthermore, by Young's theorem we know that, $\frac{\partial^2 G}{\partial p_i \partial V_j} = \frac{\partial^2 G}{\partial V_j \partial p_i}$, so it follows that

$$(c) \frac{\partial^2 G}{\partial p_i \partial V_j} = \frac{dw_j}{dp_i} = \frac{\partial^2 G}{\partial V_j \partial p_i} = \frac{dy_i}{dV_j}. \quad (3.4)$$

Samuelson (1953–54) called conditions (3.4) the “reciprocity relations,” and they show that the Stolper-Samuelson derivatives are identical to the Rybczynski derivatives! Results (a)–(c) hold regardless of the number of goods and factors, but require differentiability of the GDP function. We shall comment on the validity of this assumption later.

We now consider some special cases.

SAME NUMBER OF GOODS AND FACTORS

With $N = M$, the zero-profit conditions (3.1) become N equations in N unknowns. The question, then, is whether we can obtain a solution for the wages w and if this solution is unique. In chapter 1, we discussed the case of “factor intensity reversals,” whereby the iso-cost lines of the unit-costs curves for two industries intersect twice (see figure 1.6). This means that there are some prices for the two industries such that the iso-cost lines are tangent, as shown in figure 3.1 at point B . At these factor prices, the factor requirements vectors $a_i(w)$ in the two industries are proportional.

A factor intensity reversal (FIR) means that there are factor prices w at which two columns of the matrix A are proportional, so that A is singular. Actually, a factor intensity reversal occurs even if the components of $a_1(w)$ and $a_2(w)$ for just two factors are proportional, that is, point B in figure 3.1 holds for factors 1 and 2, even though the demand for the other factors in these industries is not proportional. To rule out this case, we therefore require a stricter assumption than the nonsingularity of A .

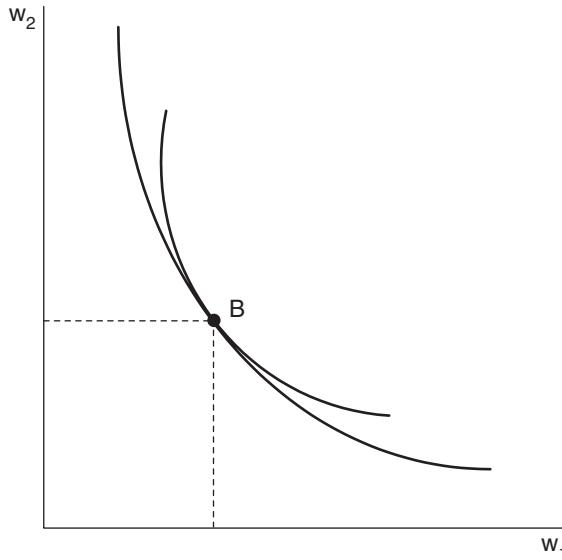


Figure 3.1

Samuelson (1953–54) first proposed the condition that the leading principal minors of A are all nonnegative, that is,

$$a_{11} \neq 0, \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} \neq 0, \dots, \begin{vmatrix} a_{11} & \cdots & a_{N1} \\ \vdots & \ddots & \vdots \\ a_{1N} & \cdots & a_{NN} \end{vmatrix} \neq 0. \quad (3.5)$$

By a natural reordering of the goods and factors, the signs of all these determinants can be taken as positive. However, Nikaido (1972) showed that this condition was not enough to ensure that the system of equations (3.1) has a unique solution for w given any $p > 0$. Instead, it had to be strengthened in the following way: there exist bounds $b > 0$ and $B > 0$ such that

$$0 < b \leq a_{11}, \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix}, \dots, \begin{vmatrix} a_{11} & \cdots & a_{N1} \\ \vdots & \ddots & \vdots \\ a_{1N} & \cdots & a_{NN} \end{vmatrix} \leq B. \quad (3.6)$$

Then we have the following “factor price insensitivity” result, as named by Leamer (1995).

LEMMA (SAMUELSON 1953–54; NIKAIDO 1972)

Assume $N = M$ and that condition (3.6) holds for all $w > 0$. Then for all $p > 0$, the set of equations (3.1) has a unique solution for factor prices $w > 0$.

This lemma implies that the factor prices can be written as a function of product prices, or $w(p)$, independent of endowments. Note that this implies a special form for

the GDP function (3.3). With constant returns to scale in all industries, GDP can be written in general as $G(p, V) = \sum_{j=1}^M w_j(p, V)V_j$, which is just the sum of payment to primary factors. Under the conditions of the above lemma, this formula is simplified to

$$G(p, V) = \sum_{j=1}^M w_j(p)V_j. \quad (3.7)$$

In particular, this implies that $\partial^2 G / \partial V_j \partial V_i = 0$, so there are no diminishing returns to the accumulation of a factor in the economy overall.

Now suppose that two countries share the same technologies, and therefore the same unit-cost functions, and through free trade they face the same prices p . Provided that both countries are producing all $i = 1, \dots, N$ goods, then this lemma establishes that they will have the same factor prices, so FPE is obtained. The set of endowments that allow both countries to be producing $y_i > 0$ for $i = 1, \dots, N$, can be computed from the full-employment conditions (3.2). This will be an FPE set, analogous to figure 1.9. In the case of equal goods and factors, the FPE set is a higher-dimension analogue to figure 1.9: rather than a parallelogram in two dimensions, we would have a hyperparallelogram in N dimensions. So for factor price equalization, the extension from two goods and two factors to N goods and N factors is obtained quite easily.

What about the Stolper-Samuelson and Rybczyski theorems? To see what sort of generalization can be obtained for these, let us totally differentiate the zero-profit conditions (3.1), and apply the Jones algebra, to obtain

$$\hat{p}_i = \sum_{j=1}^M \theta_{ij} \hat{w}_j, \quad i = 1, \dots, M, \quad (3.8)$$

where $\theta_{ij} = w_j a_{ij} / c_i$ denotes the cost-share of factor j in industry i . As we saw in chapter 1, the changes in commodity prices are *weighted averages* of the changes in factor prices. For a change in the price of each commodity i , holding other goods' prices fixed, there must then exist factors j and k such that $\hat{w}_j \geq \hat{p}_i$ and $\hat{w}_k \leq 0$. Under additional conditions we can establish that these inequalities are strict,³ and therefore, factor j has gained in real terms and factor k has lost in real terms. The Stolper-Samuelson theorem generalizes in this sense: *for a change in the price of each good, there will exist some factor that gains in real terms and another that loses*.

A similar generalization holds for the Rybczynski theorem. To see this, differentiate the full-employment conditions (3.2) with respect to endowment V_k , holding prices (and therefore factor prices) fixed, obtaining

$$\sum_{i=1}^N \frac{dy_i}{dV_k} a_{ij}(w) = 0, \quad j, k = 1, \dots, M, \quad j \neq k, \quad (3.9a)$$

$$\sum_{i=1}^N \frac{dy_i}{dV_k} a_{ik}(w) = 1. \quad (3.9b)$$

³Ethier (1974, 1984) assumed that $a_{ij} > 0$ for all factors and industries. Jones and Scheinkman (1977, 929) use the weaker assumption that every industry employs at least two inputs, and each input is employed in at least two industries. However, they both assume that with a small change in prices every good is still produced. This means that the factor intensities a_{ij} cannot be proportional in two industries.

Provided that the matrix A is nonsingular at the prevailing factor prices, we can use the conditions (3.9) to uniquely determine the Rybczynski derivatives dy_i/dV_k . From condition (3.9b), it must be the case that $dy_i/dV_k > 0$ for some good i . Using this in (3.9a), there must exist some other good j for which $dy_j/dV_k < 0$. Thus, we have shown that *for an increase in the endowment of each factor, there must be a good whose output rises and another good whose output falls.*

This generalization of the Rybczynski theorem is fine, but our results for the Stolper-Samuelson theorem are not fully satisfactory. Recall that in the two-by-two model, labor and capital *each* suffer from an increase in the price of *some* good. Does this apply in the $N \times N$ case? Jones and Scheinkman (1977) answer this in the affirmative.

THEOREM (JONES AND SCHEINKMAN 1977)

Assume that $N = M$, and that the matrix A is nonsingular at the prevailing factor prices. Then for each factor, there must be a good such that an *increase* in the price of that good will *lower* the real return to the factor.

The proof of this theorem follows from (3.9) and the reciprocity relations (3.4): starting with any factor k , we know that $dy_j/dV_k < 0$ for some good j , and so it follows that $dw_k/dp_j < 0$ for some good j . Since all other prices are fixed, it follows that w_k/p_i has fallen, along with w_k/p_j , so that the real return to factor k has been reduced. Jones and Scheinkman interpret this theorem as saying that “each factor has a good that is a natural enemy,” in the sense that raising the price of that good will *lower* the real return to the factor. However, in the general case of $N = M > 2$ goods and factors, we *cannot* prove in general that “each factor has a good that is a natural friend” that is, for each factor, there need not exist a good such that increasing the price of that good will *raise the real return* to the factor.⁴ We have already argued that raising a price will increase the real return to *some* factor, but we cannot establish that *each* factor will benefit in real terms due to an increase in the price of some good. So factors have “natural enemies” but not necessarily “natural friends.”

Empirical estimates of the Rybczynski effects in the “even” case have already been shown in the previous chapter, from Leamer (1984). Harrigan (1995) reexamines a similar set of regressions but using industry outputs as the dependent variable rather than net trade. The explanatory variables are country factor endowments, and the linear relationship between outputs and factor endowments follows directly from inverting (3.2'), obtaining $Y = A^{-1}V$, or

$$y_j^i = \sum_{k=1}^M \beta_{jk} V_k^i \quad i=1, \dots, C; j, k=1, \dots, N, \quad (3.10)$$

where i now indexes countries, j indexes goods, k indexes factors, and β_{jk} denotes the elements of A^{-1} . Harrigan uses data for the OECD countries over 1970–85 for ten manufacturing sectors and four factor supplies: capital, skilled and unskilled labor, and land. Notice that the number of goods exceeds the number of factors in this formulation, even though inverting (3.2') assumes that $N = M$. We can think of Harrigan’s

⁴Ethier (1984, proposition 20) argues that there will be a “qualified” friend, in the sense that raising some commodity price will raise that factor price enough to *lower* the aggregate income of all other factors.

approach as assuming $N = M$ in terms of the “true” number of goods and factors, but then aggregating factors in the data so that there are only four. More generally, this discussion points out that whether or not there is an equal number of goods and factors cannot be judged by simply “counting” them: instead, we would need to test some hypotheses to make a judgment, as we shall discuss in the sections that follow.

The results obtained by regressing industry outputs on factor endowments, over the panel of OECD countries and years, are similar to those obtained by Leamer (1984): for each manufacturing industry there is at least one factor with a negative Rybczynski effect, indicating that an increase in that endowment would reduce the manufacturing output. These factors were usually skilled or unskilled labor, and sometimes land. Conversely, capital has a positive coefficient β_{jk} in all ten regressions, indicating that an increase in this endowment will raise manufacturing output. Beyond these results, Harrigan (1995) is able to explore properties of the panel dataset that Leamer (1984) could not since he only had data for two years. Harrigan finds that *fixed effects* for countries are very important, indicating that there are systematic differences across countries not captured by the theory underlying (3.10). The most significant assumption in applying (3.10) to panel data is that *technologies are the same* across countries and time. We will be eliminating this assumption later in this chapter.

MORE FACTORS THAN GOODS

We turn now to the case of more factors than goods, $M > N$. We can again differentiate the zero-profit conditions to obtain (3.8), so our generalization of the Stolper-Samuelson theorem continues to hold: an increase in the price of each good will lead to a rise in the real return to some factor, and a fall in the real return to another. However, our generalization of the Rybczynski theorem *does not* hold: when we differentiate the full-employment conditions we do not obtain (3.9), because now a change in factor endowments leads to a change in their prices, so the factor intensities a_{ij} will change. Furthermore, the fact that (3.9) does not hold means that the theorem due to Jones and Scheinkman (1977) will not hold either: it is no longer the case that each factor has a good that is a “natural enemy.”

To illustrate these results, we focus on the special case of two goods and three factors, in what is called the “specific factors,” or the Ricardo-Viner model. This model has been analyzed by Jones (1971), Mayer (1974), Mussa (1974), and Neary (1978), among others. Let us return to the notation that the two sectors use labor and capital, with production functions $y_i = f_i(L_i, K_i)$. We now assume that the capital stocks K_i are *fixed* in each industry, so that labor is the only factor that moves between them. This framework is sometimes viewed as a “short run” version of the Heckscher-Ohlin model, for a time span short enough such that capital is fixed.

The full employment condition for labor is $L_1 + L_2 \leq L$, and so the GDP function for the sector-specific economy is

$$G(p, L, K_1, K_2) \equiv \max_{L_i \geq 0} \sum_{i=1}^2 p_i f_i(L_i, K_i) \quad \text{subject to } L_1 + L_2 \leq L. \quad (3.11)$$

To maximize GDP, labor will flow between the two industries until its marginal value product is the same in each. This first-order condition is stated as

$$w = p_1 f_{1L}(L_1, K_1) = p_2 f_{2L}(L_2, K_2). \quad (3.12)$$

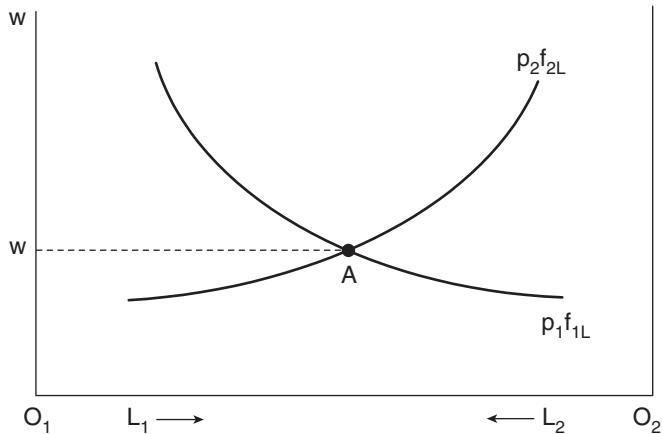


Figure 3.2

With the capital stocks fixed, (3.12) is easily solved for the labor allocation, as is illustrated in figure 3.2. The horizontal axis measures the total labor endowment L , which is allocated between industry 1 (measured from its origin O_1) and industry 2 (measured from its origin O_2). Concavity of the production functions implies a diminishing marginal product of labor, so both curves $p_i f_{iL}$ are drawn as downward sloping relative to their origins. The equilibrium wage is determined at the intersection of these two curves, point A .

Note that the position of the marginal product curves depends on the amount of capital in each industry, and therefore, so does the equilibrium wage at point A . Thus, with two countries having the same technology and facing the same prices due to free trade, but with differing endowments, there will *not be* factor price equalization: rather, the price of each factor will be inversely related to the countries' endowments.⁵

Now suppose that there is an exogenous increase in the price of good 1. What is the effect on factor prices? In figure 3.3, the marginal value product curve of industry 1 will shift up by $\Delta p_1 f_{1L}$, as illustrated. It is immediately clear that the wage increases along with the allocation of labor to industry 1. From (3.12), we can compute the increase in the wage as

$$\frac{dw}{dp_1} = f_{1L} + p_1 f_{1LL} \frac{dL_1}{dp_1} < f_{1L} = \frac{w}{p_1}, \quad (3.13)$$

where the inequality follows from $f_{1LL} < 0$ and $dL_1/dp_1 > 0$. With $dp_1 > 0$, it follows immediately from (3.13) that $dw/w = \hat{w} < dp_1/p_1 = \hat{p}_1$. The fact that $\hat{w} > 0$ means that workers can buy more of good 2, whose price is fixed, but the fact that $\hat{w} < \hat{p}_1$ means that (w/p_1) has fallen, so that workers can afford less of good 1.⁶ Thus, the change in the real wage is *ambiguous*: whether a worker is better off or worse off depends on his

⁵See problem 3.1.

⁶This result can also be seen from figure 3.3, since the vertical shift of the marginal value product curve exceeds the increase in the relative wage, $\Delta w < \Delta p_1 f_{1L}$, with f_{1L} evaluated at the initial labor allocation at point A . Dividing both sides of this equation by $w = p_1 f_{1L}$, we obtain $\Delta w/w < \Delta p_1/p_1$.

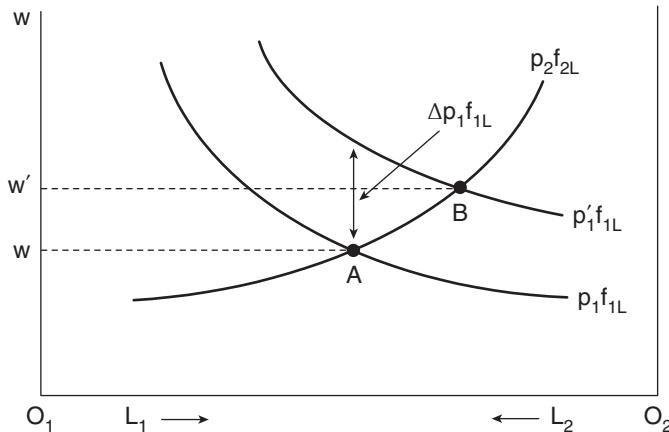


Figure 3.3

or her relative consumption of the two goods. This illustrates that the theorem due to Jones and Scheinkman (1977) no longer holds with more factors than goods.

What about the rental to capital in each industry? From the zero-profit conditions (3.8) the change in the prices of the goods is a weighted average of the change in wage and rentals. Denoting the latter by \hat{r}_i , the fact that $0 < \hat{w} < \hat{p}_1$ implies

$$\hat{r}_2 < 0 < \hat{w} < \hat{p}_1 < \hat{r}_1. \quad (3.14)$$

If we allow for a change in the price of good 2, with $\hat{p}_2 < \hat{p}_1$, then these inequalities become

$$\hat{r}_2 < \hat{p}_2 < \hat{w} < \hat{p}_1 < \hat{r}_1. \quad (3.14')$$

This shows how the “magnification effect” present in the two-by-two model is modified when capital is specific to each sector. In that case, the wage of labor is “caught in the middle,” whereas the returns to the *specific factors* are still a *magnified* version of the price changes.

The effects of changes in the endowments of factors are also straightforward to derive. Suppose first that the capital stock in industry 1 rises. This increases the marginal product of labor in that industry, which leads to a higher wage, as shown in figure 3.4.⁷ Thus, labor benefits from an increase in the capital stock in either industry. Furthermore, we can apply (3.8) while setting $\hat{p}_1 = 0$ since we are assuming that product prices are fixed. The increase in the wage must be offset by a *fall* in the rental on capital in both industries.⁸ So all specific factors suffer from an increase in their stock in any industry. Figure 3.4 also shows that the labor allocated to industry 1 increases, while the labor allocated to industry 2 decreases. It follows that y_1 rises but y_2 falls. So an increase in either specific factor will raise the output of that industry but lower the output of the other industry.

⁷With only two factors, and a production function $f(L, K)$ that is concave and homogeneous of degree one, it must be the case that $f_{LK} > 0$.

⁸Exercise: use (3.8) with $\hat{p}_1 = 0$ to show which specific factor has the greatest fall in its rental.

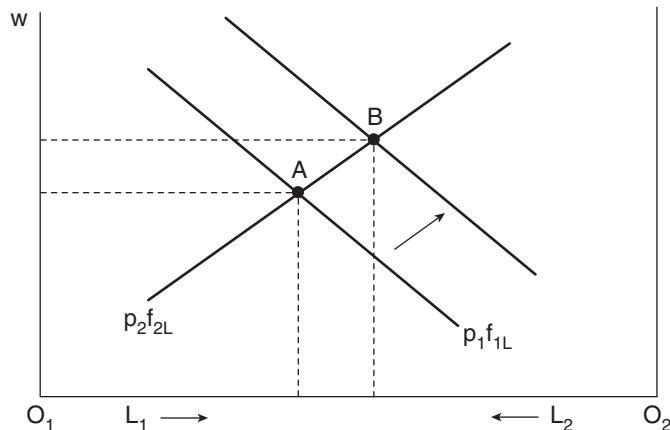


Figure 3.4

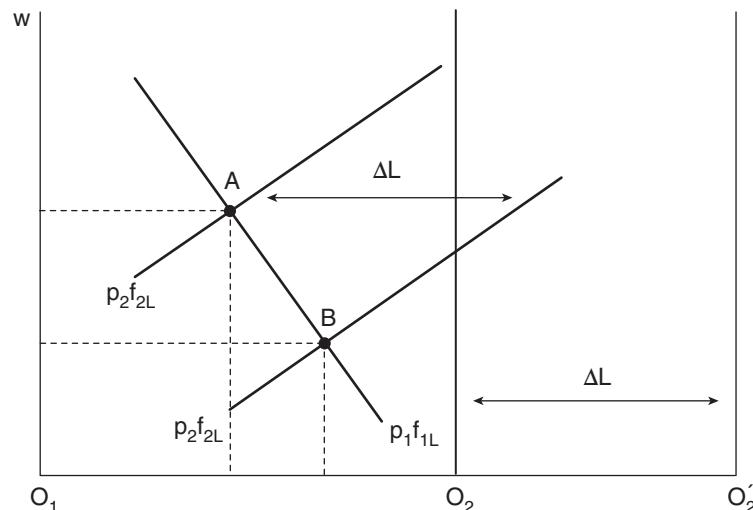


Figure 3.5

Next, consider an increase in the labor endowment. This expands the labor axis in figure 3.5, and as the origin O_2 shifts to the right, it carries along with it the $p_2 f_{2L}$ curve. The equilibrium shifts from point A to point B , and it is clear that the wage falls while the increase in labor in industry 1 is less than the total increase in the labor endowment. These two results imply that the specific factors in *both* industries experience a rise in their rentals, and the output of both industries increases. So for an increase in labor, which is the mobile factor, there is no longer a Rybczynski-like effect.

ESTIMATING THE GDP FUNCTION

The number of goods and factors, as well as whether factors are mobile or not, influences the form of the GDP function. To capture these effects empirically we can hypothesize a very general functional form for GDP and then use data on industry

output and prices, as well as factor endowments and prices, to estimate the GDP function. By taking the derivatives of the estimated GDP function with respect to prices and factor endowments, we can then obtain the Rybczynski and Stolper-Samuelson derivatives that apply to an actual economy.

The first step is to choose a functional form for GDP. A convenient choice is the translog function, which was introduced by Diewert (1974, 139) and initially used in the international trade literature by Kohli (1978). This function is written as

$$\begin{aligned} \ln G = & \alpha_0 + \sum_{i=1}^N \alpha_i \ln p_i + \sum_{k=1}^M \beta_k \ln V_k + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \ln p_i \ln p_j \\ & + \frac{1}{2} \sum_{k=1}^M \sum_{\ell=1}^M \delta_{k\ell} \ln V_k \ln V_\ell + \sum_{i=1}^N \sum_{k=1}^M \varphi_{ik} \ln p_i \ln V_k \end{aligned} \quad (3.15)$$

where p_i denotes the prices of products $i=1, \dots, N$, and V_k denotes the endowments of factors $k=1, \dots, M$. In order to ensure that the translog GDP function is homogeneous of degree one in prices, we impose the requirements⁹

$$\sum_{i=1}^N \alpha_i = 1 \quad \text{and} \quad \sum_{i=1}^N \gamma_{ij} = \sum_{i=1}^N \varphi_{ik} = 0. \quad (3.16)$$

In addition, to ensure that the GDP function is homogeneous of degree one in endowments, we impose the requirements

$$\sum_{k=1}^M \beta_k = 1 \quad \text{and} \quad \sum_{k=1}^M \delta_{k\ell} = \sum_{k=1}^M \varphi_{ik} = 0. \quad (3.17)$$

The usefulness of the translog function comes from computing its first derivatives, $\partial \ln G / \partial \ln p_i = (\partial G / \partial p_i)(p_i/G)$. Because $\partial G / \partial p_i$ equals the output of sector i , it follows that $(\partial G / \partial p_i)(p_i/G)$ equals the *share of sector i in GDP*, which we denote by the output-shares $s_i = p_i y_i / G$. Thus, differentiating (3.15) with respect to $\ln p_i$ we obtain

$$s_i = \alpha_i + \sum_{j=1}^N \gamma_{ij} \ln p_j + \sum_{k=1}^M \varphi_{ik} \ln V_k, \quad i=1, \dots, N. \quad (3.18)$$

In addition, we can compute the first derivatives with respect to factor endowments, obtaining $\partial \ln G / \partial \ln V_k = (\partial G / \partial V_k)(V_k/G)$. Because $\partial G / \partial V_k$ equals the factor price of endowment k , it follows that $(\partial G / \partial V_k)(V_k/G)$ equals the *share of GDP devoted to factor k*, which we denote by the factor shares $s_k = w_k V_k / G$. Thus, differentiating (3.15) with respect to $\ln V_k$ we obtain

$$s_k = \beta_k + \sum_{\ell=1}^M \delta_{k\ell} \ln V_\ell + \sum_{i=1}^N \varphi_{ik} \ln p_i, \quad k=1, \dots, M. \quad (3.19)$$

Equations (3.18) and (3.19) are $N + M$ linear equations that can be estimated using annual data on output and factor shares, product prices, and endowments. The dependent variables in each of (3.18) and (3.19) sum to unity, which means that one of the equations can be derived from the others. Under these conditions, one of the equations is dropped from each of (3.18) and (3.19) before the system is estimated, leaving

⁹Without loss of generality, we also impose the symmetry restrictions $\gamma_{ij} = \gamma_{ji}$ and $\delta_{k\ell} = \delta_{\ell k}$. See problem 3.2 to show that (3.16) and symmetry ensure that the translog cost function is homogeneous of degree one in prices.

$(N + M - 2)$ equations. With annual observations over T years, this means that there are $(N + M - 2)T$ observations used to estimate the coefficients γ_{ij} , φ_{ik} and $\delta_{k\ell}$.¹⁰ The great advantage of estimating this system of shares equations, rather than just the GDP function in (3.15), is that there are more observations and degrees of freedom. Even with just a modest number of annual observations, such as twenty, the researcher can expect to obtain reasonably precise estimates of the coefficients.

The coefficients φ_{ik} in (3.18) measure the response of each output share to changes in the endowments, and are referred to as *output elasticities*. These are similar to the Rybczynski effects, which are normally defined as the impact of a change in endowments on the *level* of output rather than its *share*. To make this conversion, write the quantity of each output as $\ln y_i - \ln(s_i G/p_i)$. Differentiating this with respect to an endowment $\ln V_k$ using (3.15) and (3.18), we obtain the Rybczynski elasticity:

$$\frac{\partial \ln y_i}{\partial \ln V_k} = \frac{\varphi_{ik}}{s_i} + s_k. \quad (3.20)$$

Thus, the coefficients φ_{ik} together with the input and output shares can be used to calculate the Rybczynski elasticity.

Similarly, write the price of each factor as $\ln w_k = \ln(s_k G/V_k)$. Differentiating this with respect to $\ln p_i$ using (3.15) and (3.19), it follows that the Stolper-Samuelson elasticities are

$$\frac{\partial \ln w_k}{\partial \ln p_i} = \frac{\varphi_{ik}}{s_k} + s_i. \quad (3.21)$$

In addition, we can differentiate $\ln w_k = \ln(s_k G/V_k)$ with respect to the factor endowments to obtain

$$\frac{\partial \ln w_k}{\partial \ln V_\ell} = \begin{cases} \frac{\delta_{kk}}{s_k} + s_k - 1, & \text{if } k = \ell, \\ \frac{\delta_{k\ell}}{s_k} + s_\ell, & \text{if } k \neq \ell. \end{cases} \quad (3.22)$$

These elasticities allow us to test the hypothesis of “factor price insensitivity,” meaning that the GDP function takes the special form in (3.7). While this functional form cannot be imposed globally for the translog GDP function, it can be evaluated locally by testing whether (3.22) is zero. This means that we test $\delta_{kk} = s_k(1 - s_k)$ and $\delta_{k\ell} = -s_k s_\ell$, which can be performed at one year (e.g., the midpoint) of the sample. In summary, given the estimated coefficients from the translog share equations, we can readily compute the Rybczynski and Stolper-Samuelson elasticities and also test for “factor price insensitivity.”

Turning to empirical applications, Kohli (1978) estimates a translog GDP function for Canada, while Kohli (1990a) estimates it for U.S. data. Focusing on the U.S. application, Kohli uses the five major GDP components as goods: consumption (C), investment (I), government services (G), exports (X), and imports (M). Notice that he treats exports and imports as an output and input, respectively, in the production

¹⁰The hypotheses of symmetry and homogeneity, in (3.16) and (3.17), can be tested by first estimating the translog system without these restrictions and comparing the system estimated with the restrictions.

process. This is quite different from our treatment in the previous chapter, where net exports were the difference between consumption and production: now we are defining exports and imports independently from consumption. This makes sense if exports are differentiated from domestic goods and if imports are mainly intermediate inputs. Both of these are plausible hypotheses, and we will explore the role of imports as intermediate inputs more in the next chapter.

For primary inputs Kohli (1990a) identifies labor and capital, so he obtains five output-share equations (3.18) (where imports have a negative share) and two factor-share equations (3.19). Dropping one equation from each system, we then have five equations in total, and with data over forty years (1948–87) this gives two hundred observations to estimate a total of twenty-five parameters of the translog system.¹¹ In his results, Kohli finds that the Rybczynski elasticities (3.20) are *positive* for the impact of labor on the supply of exports, investment goods and the demand for imports, and also for the impact of the capital stock on the production of consumption goods. In other words, exports and investment goods appear to be *labor* intensive, while consumption goods appear to be *capital* intensive, as measured by these Rybczynski elasticities. The signs of the Stolper-Samuelson elasticities in (3.21) are the same, of course, so that an increase in the import price hurts labor, while an increase in the export price benefits labor.

This sign pattern of elasticities found by Kohli is somewhat puzzling, since we normally think of investment goods as capital intensive in their production, and the same is true for U.S. exports, whereas consumption goods (including many nontraded services) would seem to be labor intensive. Kohli (1990a) finds the opposite of the expected results, confirmed in Kohli (1993a) using a different functional form. Perhaps the difficulty is that the macroeconomic variables C , I , and G are not the best way to form output aggregates, despite the fact that these categories are readily available from the national accounts. Indeed, the finding that U.S. exports appear to be labor intensive already appeared in the early results of Leontief (1953), discussed in chapter 2, and one explanation for this “paradox” (besides the critique of Leamer 1980) was that goods or factors should be disaggregated further.

An alternative aggregation scheme is used by Burgess (1974a,b, 1976), who estimates an aggregate cost rather than GDP function for the United States. He also includes imports as an input into the cost function, along with labor and capital, and for outputs uses traded goods (durables and nondurables) and nontraded goods (non-government services and structures). Burgess notes that the relative price movements within each of these aggregates are similar, which is one justification for the formation of an aggregate. Using data for the United States over thirty-two years (1948–69), Burgess finds that traded goods are *capital* intensive whereas nontraded goods are *labor* intensive. This matches our prior beliefs about U.S. exports versus nontraded services. In addition, he finds that nontraded goods have a lower cost-share of imports than do traded goods. This means that an increase in the price of imports has a greater impact on traded goods and, because this sector is capital intensive, leads, to a fall in the

¹¹In addition to five constant terms α_i and β_k in the share equations (3.18) and (3.19) (after dropping one from each system), Kohli (1990a) includes five time trends reflecting technological change. In addition, the five-by-five matrix $[\gamma_{ij}]$ has ten free parameters after taking into account symmetry $\gamma_{ij} = \gamma_{ji}$ and the homogeneity restrictions (3.16); the two-by-two matrix $[\hat{\delta}_{kk}]$ has one free parameter after taking into account symmetry and the homogeneity restrictions (3.17); and the five-by-two matrix $[\varphi_{ik}]$ has four free parameters after taking into account (3.16) and (3.17).

return to capital and a rise in the relative wage.¹² In this sense, tariffs benefit labor in the United States (which is the opposite of the result found by Kohli).

Burgess (1976) also tests the hypothesis that the aggregate cost function $C(Y, w)$ can be written in the linear form

$$C(Y, w) = \sum_{i=1}^N y_i c_i(w), \quad (3.23)$$

where the functions $c_i(w)$ are interpreted as *unit-cost* functions within the industries $i=1, \dots, N$. When (3.23) *does not* hold, it means that we cannot identify distinct industries producing each of the outputs; rather, these outputs are produced *jointly* from the economy's factor endowments. So (3.23) is referred to as a test of "nonjoint production." We have not yet emphasized that nearly all of our results on the Stolper-Samuelson and Rybczynski theorems so far in this book rely on *nonjoint* production. In contrast, if output is produced jointly from one or more inputs, there is no presumption that "magnification effects" of the type we have discussed hold. Notice the similarity between the linear form for the aggregate cost function in (3.23) and the linear form for the GDP function in (3.7). We argued that (3.7) is a test of "factor price insensitivity," whereas (3.23) is a test of nonjoint production, but sometimes these concepts are used nearly interchangeably.¹³

Burgess (1976) finds that (3.23) is rejected at exactly the 1 percent of significance. Thus, there is some evidence for the United States that the production of traded and nontraded goods occurs under *joint* production. Likewise, Kohli (1981) also rejects the hypothesis of a nonjoint aggregate cost function using U.S. data. Considering the hypothesis of "factor price insensitivity," which means that the GDP function can be written in the linear form (3.7), this can be evaluated from the elasticities (3.22) reported by Kohli (1990a, 1993a). He finds that factor returns do depend on endowments—with higher endowments leading to lower returns on that factor—but only weakly so (standard errors are not reported).

Notice that these tests of (3.7) and (3.23) are actually combining two distinct hypotheses: perfect mobility of capital between sectors and nonjointness in production. So when we reject (3.7) and (3.23), we are not sure which of these hypotheses is really responsible. Kohli (1993b) addresses this by developing a model where capital is *sector specific*, though the researcher does not observe the amount of capital in each sector. Only observing the *aggregate* capital stock, one way to proceed is to assume that the capital devoted to each sector is always in *fixed proportion* to the aggregate stock. In that case, the GDP function is written as a function of aggregate labor (mobile between the two sectors) and aggregate capital (in fixed proportion in each sector).

Using this sector-specific structure, Kohli (1993b) develops a test for nonjointness in production that he refers to as "almost nonjointness." This test can be applied to either the GDP or cost function, and is weaker than the tests in (3.7) and (3.23) because the return to capital in each sector need not be the same. He implements this test on a two-by-two aggregation structure of the U.S. economy, where labor and capital produce consumption and investment goods. In the results, he finds that "almost nonjointness" is rejected for the aggregate cost function but is *not rejected* for the GDP

¹²This result is developed in problem 3.3.

¹³For example, Kohli (1983; 1991, 44) refers to (3.23) as "nonjointness in input quantities" and (3.7) as "nonjointness in output quantities."

function. Even on the cost side, the parameter estimates of the cost function do not differ that much when “almost nonjointness” is imposed. So Kohli concludes that a *sector-specific structure* of the U.S. economy is broadly consistent with the annual data, and is much preferable to perfect mobility of capital combined with nonjointness as in (3.7) and (3.23). The implied Stolper-Samuelson and Rybczynski elasticities for the sector-specific structure are as we discussed them in the previous section.

A final application of the translog GDP function that we shall discuss is due to Harrigan (1997). Building on the prior work of Leamer (1984) and Harrigan (1995), he is interested in estimating Rybczynski elasticities on cross-country and time-series data. A limitation of early work was the assumption that technologies were the same across countries. Harrigan (1997) argues that this can be generalized by instead assuming that the GDP function is the same across countries while *allowing* for Hicks-neutral differences in the industry production functions. Thus, if industry production functions are denoted by $y_i = A_i f_i(v_i)$, where the coefficients A_i can differ across countries, then the GDP function becomes

$$G(A_1 p_1, \dots, A_N p_N, V) \equiv \max_{v_i \geq 0} \sum_{i=1}^N p_i A_i f_i(v_i) \quad \text{subject to } \sum_{i=1}^N v_i \leq V. \quad (3.24)$$

Notice that we have multiplied the prices p_i appearing in the GDP function by the Hicks-neutral productivity parameters A_i , because that is how they appear on the right side of (3.24). If we adopt a translog form for the GDP function, then the output share equations in (3.18) would be written as

$$s_i = \alpha_i + \sum_{j=1}^N \gamma_{ij} \ln(A_j p_j) + \sum_{k=1}^M \varphi_{ik} \ln V_k, \quad i=1, \dots, N. \quad (3.25)$$

That is, the output shares of each industry depend on their prices *and* productivities (with the same coefficient applying to each), as well as the factor endowments.

We will be discussing the measurement of industry productivity further in chapters 4 and 10, as well as properties of the GDP function in (3.24). For the present purposes, we note that Harrigan (1997) assumes that countries are engaged in free trade, so that industry prices are the same and no longer appear in (3.25):¹⁴ it follows that the determinants of the output shares are productivities and endowments. It is the inclusion of the productivity variables in (3.25) that distinguishes this approach from Leamer (1984) and Harrigan (1995), and this is an important generalization from earlier research.¹⁵

Harrigan estimates (3.25) over a panel dataset consisting of 10 OECD countries over about twenty years (from 1970 to 1988–90). He distinguishes seven major manufacturing sectors (food, apparel, papers, chemicals, glass, metals, and machinery), which consist of aggregates from the International Standard Industrial Classification (ISIC). For endowments he uses six factors: two types of capital (producer durables and nonresidential construction), three types of labor (high-educated with some post-secondary schooling, medium-educated with some high school, and low-educated without high-school), and arable land. Because he has data only on factor

¹⁴The prices can be normalized at unity in the first year, and then Harrigan assumes that their values in ensuing years can be captured by a simple time trend in each share equation.

¹⁵The model proposed by Harrigan (1997) has also been used in work by Redding (2002), which is recommended for further reading.

TABLE 3.1: Estimates of Rybczynski Effects from Harrigan (1997)

	Producer Durables	Non-residential Constr'n	High-education Workers	Medium-education Workers	Low-education Workers	Arable Land
Food	1.31 (0.19)	-0.20 (0.29)	-0.17 (0.13)	0.68 (0.20)	-0.02 (0.14)	-1.60 (0.30)
Apparel	0.94 (0.14)	-0.35 (0.20)	-0.66 (0.09)	0.69 (0.14)	0.10 (0.10)	-0.71 (0.23)
Paper	-0.02 (0.11)	0.16 (0.17)	-0.22 (0.07)	-0.04 (0.11)	-0.15 (0.08)	-0.26 (0.70)
Chemicals	1.19 (0.21)	-1.53 (0.29)	-0.002 (0.10)	-0.89 (0.20)	-0.40 (0.15)	1.63 (0.32)
Glass	0.36 (0.09)	-0.24 (0.14)	-0.19 (0.06)	0.38 (0.09)	-0.10 (0.07)	-0.20 (0.15)
Metals	0.19 (0.20)	-0.07 (0.28)	-0.50 (0.13)	-0.21 (0.19)	-0.22 (0.14)	0.81 (0.31)
Machinery	0.91 (0.48)	-1.75 (0.72)	-2.11 (0.32)	1.01 (0.48)	1.82 (0.35)	0.12 (0.88)

Source: Harrigan (1997, 488).

Notes: Standard errors are in parentheses.

quantities but not prices, he does not estimate the corresponding system of factor-share equations. The output shares are measured relative to GDP, so these sum to less than unity over the seven manufacturing sectors, and there is no need to drop one of the equations. He therefore has seven equations with about two hundred observations for each, or some fourteen thousand observations to estimate a total of seventy-seven parameters.¹⁶

The calculation of the Rybczynski elasticities in (3.20) requires both the output shares and factor shares. Since the latter are not available, we simply report the estimated coefficients φ_{ik} from the share equations (3.25), in table 3.1. These Rybczynski effects have signs that match our expectations. Thus, producer durable investment has a positive impact on the share of each of the seven manufacturing sectors, but nonresidential construction has a negative coefficient in most cases. This is consistent with nontraded services being intensive in nonresidential construction and drawing resources away from manufacturing. Only two of the sectors have a positive and significant Rybczynski effect with arable land, and these are chemicals and metals, which depend on natural resources. With the exception of these two, most other sectors have a positive Rybczynski effect with either medium-educated or low-educated labor, but a negative effect with highly educated labor, which is consistent with highly educated workers being used intensively in nontraded services. These results are broadly consistent with, and build upon, the findings of Leamer (1984) and Harrigan (1995).

¹⁶Along with seven constant terms α_i in the share equations (3.18), Harrigan also includes seven time trends. In addition, the seven-by-seven matrix $[\gamma_{ij}]$ has twenty-eight free parameters after taking into account symmetry $\gamma_{ij} = \gamma_{ji}$, and the seven-by-six matrix $[\varphi_{ik}]$ has thirty-five free parameters after taking into account (3.16).

MORE GOODS THAN FACTORS

We now consider a model with more goods than factors, and, for convenience, suppose that there are *three* goods and *two* factors (labor and capital). We assume that there is perfect factor mobility between the industries. The zero-profit conditions (3.1) then become three equations in two unknowns, so we cannot find a solution except at special values of the prices (p_1, p_2, p_3). This situation is illustrated in figure 3.6. If we consider one of these special prices (p_1, p_2, p_3) at which all three goods are produced, then the full-employment conditions (3.2) give two equations in three unknowns, so these are underdetermined: there are *many solutions* for (y_1, y_2, y_3) . What is the explanation for these unusual results?

Using the result that there are multiple solutions for (y_1, y_2, y_3) , we can graph these as a *line* of outputs on the production possibility frontier (PPF) in figure 3.7. Therefore,

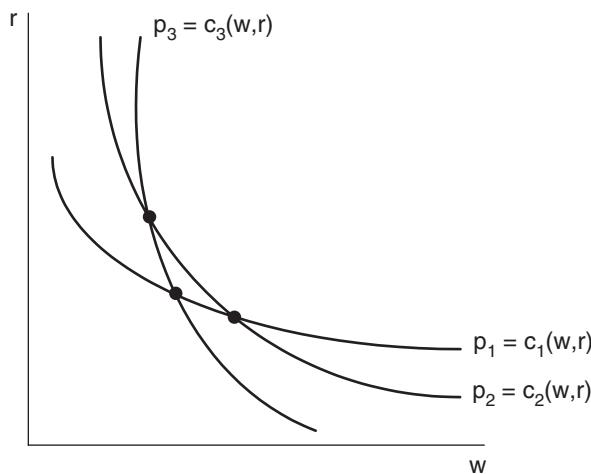


Figure 3.6

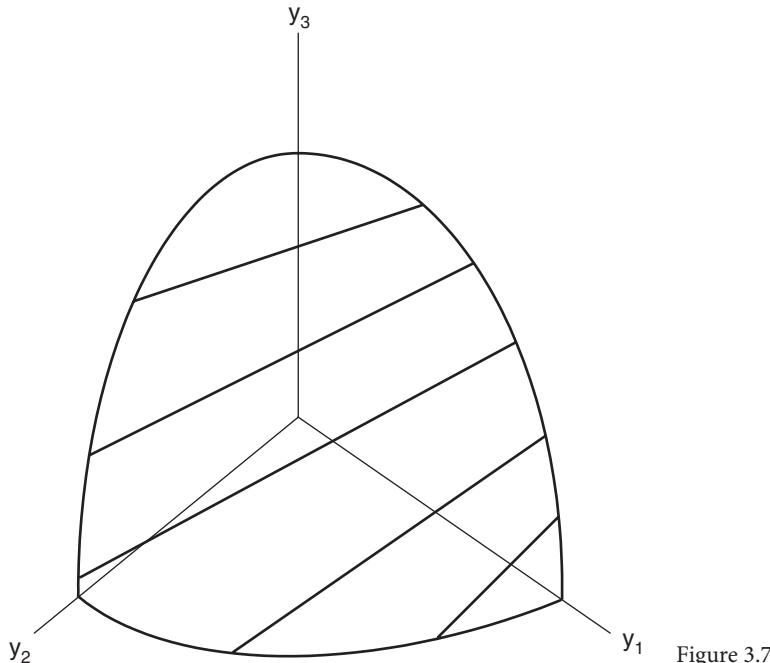


Figure 3.7

the PPF has “ruled segments” on it, as illustrated. Any price vector $p = (p_1, p_2, p_3)$ that allows for zero profits in all three industries will correspond to a *line of outputs* along the PPF. At the same time, a slight movement in prices away from this point will very likely lead to a *corner solution*, where only two of the goods are produced. Thus, for most price vectors there will be specialization in a subset of goods, equal to the number of factors. Conversely, if the price vector just happens to allow for zero profits in all goods, then there will be multiple solutions for the outputs.¹⁷

With these initial observations, what can we say about factor price equalization? At first glance, the problem seems intractable: the set of goods produced in a country will be extremely sensitive to the product prices. But let us recall that the prices being considered must be the *equilibrium prices* with free trade. Let us make the strong assumption that the countries have identical technologies. Since all goods must be produced in equilibrium, it follows that for a wide range of factor endowments the equilibrium prices will indeed be one of the special vectors (p_1, p_2, p_3) that is consistent with all goods being produced in each country. In other words, it will not be unusual to find that the plane with normal vector (p_1, p_2, p_3) is tangent to one of the ruled segments on the PPF for each country, with indeterminate outputs in each country, but adding up to world demand in total.

To demonstrate these claims, we will construct the factor price equalization (FPE) set for two countries, similar to that constructed in chapter 1 but now using multiple goods. Recall from chapter 1 that we construct the FPE set by first considering an “integrated world equilibrium” where labor and capital are free to move between countries, ensuring that their factor prices are equalized. In this equilibrium we establish product prices that are consistent with all goods being produced. These can be used to construct factor prices (the same in both countries) and the factor intensities a_{ij} . Then the question is whether we can reproduce the same equilibrium but having fixed endowments in each country, without factor mobility. The range of endowment over which we can obtain the same equilibrium is the FPE set.

To determine this range let us rank the goods in terms of increasing labor/capital ratio, so that $a_{1L}/a_{1K} \geq a_{2L}/a_{2K} \geq \dots \geq a_{NL}/a_{NK}$. Then using the world demands D_i^w for each good in the integrated equilibrium, compute the labor and capital demands in each industry, $(a_{iL}, a_{iK})D_i^w$. Summing these, we must obtain the world endowments of labor and capital,

$$\sum_{i=1}^N (a_{iL}, a_{iK}) D_i^w = (L^w, K^w). \quad (3.26)$$

This full-employment condition is illustrated in figure 3.8, where the axes measure the world endowments of labor and capital. Starting at the origin 0 for the home country, we sum the usage of labor and capital in each industry as in (3.26), obtaining the consecutive points shown in figure 3.8 that reach the opposite origin 0^* . Conversely, starting at the foreign origin 0^* , we can sum the usage of labor and capital in each industry to arrive at the home origin 0.

Now consider any point B inside this region in figure 3.8. We leave it to the reader to confirm that there is a positive linear combination of the vectors (a_{iL}, a_{iK}) that sum to point B from either origin. Indeed, there are most likely *many positive combinations* of the vectors (a_{iL}, a_{iK}) that sum to point B , which illustrates the indeterminacy

¹⁷This means that the GDP function is not differentiable in the product prices.

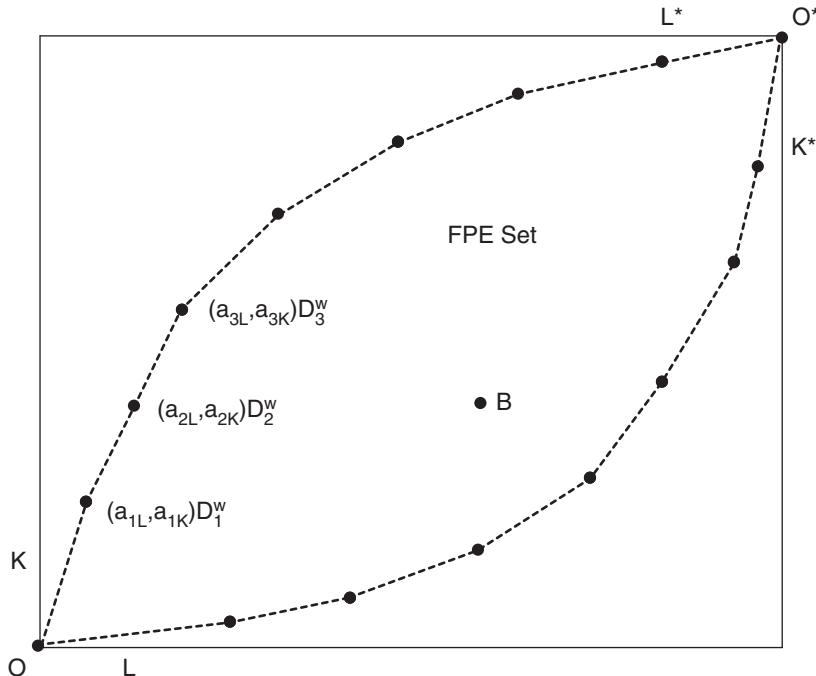


Figure 3.8

of production in each country. But this indeterminacy does not really matter, because the world equilibrium is maintained at the same product and factor prices as in our initial hypothetical equilibrium. Thus, all points inside the region labeled the FPE set in figure 3.8 lead to factor price equalization, with equilibrium product prices that are consistent with zero profits for all goods in both countries.¹⁸

An empirical test for the indeterminacy of production is proposed by Bernstein and Weinstein (2002). They note that the full-employment condition (3.2') holds regardless of the number of goods and factors, where the $M \times N$ matrix A is the same across countries. If $N = M$ and A is nonsingular, then we could invert (3.2') to obtain the regression (3.10), which is estimated by Harrigan (1995). Denoting the coefficients of that regression by the matrix $A^{-1} = B = [\beta_{jk}]$, it follows that

$$AB = I_M, \quad (3.27)$$

where I_M is an identity matrix of dimension M . Thus, a test for the indeterminacy of production proceeds by estimating the regression (3.10) as in Harrigan, and using the coefficients to test the linear restriction in (3.27). If these fail, we conclude that production is indeterminate.

Bernstein and Weinstein use data across forty-seven Japanese prefectures around 1985, with twenty-nine sectors and three factors (college-educated workers, other workers, and capital). They first check whether the full-employment condition

¹⁸This construction of the FPE set is due to Dixit and Norman (1980, 114–21), and we will have an opportunity to use it again in chapter 11, in our discussion of multinationals.

$AY^i = V^i$ holds for each prefecture $i = 1, \dots, 47$. In this test, the matrix A is held fixed at its national value for Japan. This means that the national full-employment condition $A \sum_{i=1}^{47} Y^i = \sum_{i=1}^{47} V^i$ holds by construction, and the first test is to see whether this condition also holds for each prefecture. This first test passes easily. However, Bernstein and Weinstein strongly reject the second test, which consists of estimating regression (3.10) using the cross-prefecture data and testing restriction (3.27). The implication is that the location of production across the prefectures in Japan is not simply explained by the endowments in each region: some other features must be at work.¹⁹

Is there any way to resolve production indeterminacy when there are more goods than factors? Bhagwati (1972), responding to Jones (1956–57), suggested that a determinate production structure would occur when factor prices were unequal across countries: in that case, *every commodity* exported by a country with high labor/capital endowment would need to have a higher labor/capital ratio than the exports of the other country. A proof of this “chain proposition” is provided by Deardorff (1979). Bhagwati also suggests that the existence of transportation costs would be enough to make factor prices unequal, and therefore resolve the production indeterminacy. Deardorff argues that this is true when there is only trade in final goods (and no subsidies to trade), but it breaks down when there is also trade in intermediate inputs. Setting aside this concern, are there any other cases where production indeterminacy can be resolved? Instead of relying on transportation costs to generate factor price differences, it would seem that we could instead rely on *endowment differences* between the countries, large enough to move them *outside* of the factor price equalization set. This turns out to be a fruitful line of inquiry, and to model it most elegantly, it will be useful to introduce a continuum of products into the Heckscher-Ohlin model. This is done by Dornbusch, Fischer, and Samuelson (1980), who also introduce a continuum of goods into the Ricardian model (Dornbusch, Fischer, and Samuelson, 1977). We will cover the Ricardian model with a continuum of goods next, drawing heavily on the presentation in Matsuyama (2008). Following that presentation we briefly discuss the Heckscher-Ohlin model with a continuum of goods, which is used to complete our treatment of Davis and Weinstein (2001) that we began in the previous chapter.

CONTINUUM OF GOODS: RICARDIAN MODEL

Dornbusch, Fischer, and Samuelson (1977) suppose that there is continuum of industries indexed by $z \in [0, 1]$, operating under perfect competition. We let $a(z)$ and $a^*(z)$ denote the labor required to produce a single unit of a good at home and abroad, respectively. The home country’s relative efficiency is defined as $A(z) \equiv a^*(z)/a(z)$. Without loss of generality, goods can be relabeled so that $A(z)$ is non-increasing function of z . In other words, goods that are relatively more productive at home, with lower values of $a(z)/a^*(z)$ and so higher values of $a^*(z)/a(z)$, appear first on the continuum $z \in [0, 1]$, as shown by the curve $A(z)$ in figure 3.9.

¹⁹Other features are suggested by Davis and Weinstein (2001b), which is recommend for further reading. Bernstein and Weinstein (2002) go on to compare their results for Japan to those obtained with international data. They find that even the full-employment condition $AY^i = V^i$ fails when using the Japanese national data for A , but this is not the case for (3.27) when using other countries data for Y^i and V^i . This is the familiar finding that production techniques differ across countries.

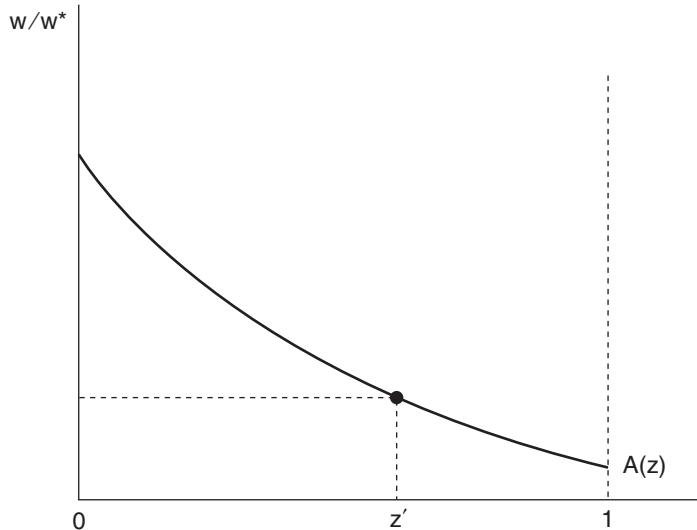


Figure 3.9

Wages at home and abroad are denoted by w and w^* . Then prices at autarky are equal to $p(z) = wa(z)$ and $p^*(z) = w^*a^*(z)$. Under free trade, the country with the lowest price for a particular good produces this good for both countries. It follows that under free trade we have $p(z) = p^*(z) = \min\{wa(z), w^*a^*(z)\}$. In equilibrium there must exist a marginal good z' for which the prices are equalized in the two countries, which means that

$$\frac{w}{w^*} = A(z'). \quad (3.28)$$

This condition is illustrated in figure 3.9. With wages at the level shown, the goods with indices $[0, z']$ are produced in the home country and goods with indices $[z', 1]$ are produced abroad. Then the prices are

$$p(z) = p^*(z) = wa(z), z \in [0, z']; p(z) = p^*(z) = w^*a^*(z), z \in [z', 1]. \quad (3.29)$$

To determine the equilibrium wages, let L and L^* denote number of workers/consumers at home and abroad, respectively. Each consumer is assumed to have Cobb-Douglas preferences: $U = \int_0^1 \ln[c(z)] dz$ and $U^* = \int_0^1 \ln[c^*(z)] dz$. World income and world expenditure are then equal to $wL + w^*L^*$ and world expenditure on domestic goods is equal to $z'(wL + w^*L^*)$, because all goods in the range $[0, z']$ are exported from the home country. With balanced trade, expenditure on domestic goods must equal to home income, which is wL . Simplifying the condition $z'(wL + w^*L^*) = wL$, we obtain

$$\frac{w}{w^*} = \frac{z'}{1 - z'} \left(\frac{L^*}{L} \right) \equiv B(z'). \quad (3.30)$$

The right-hand side of (3.30) is denoted by $B(z')$, and is drawn as the curve $B(z)$ in figure 3.10. This curve is upward sloping because as a larger share of goods is produced

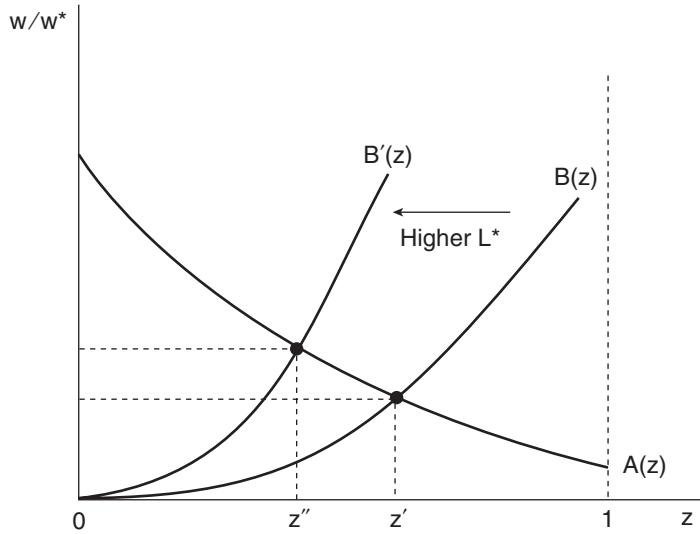


Figure 3.10

at home, the greater demand for these goods leads to a higher relative wage at home. The intersection of the curves $A(z)$ and $B(z)$ determines the equilibrium relative wage and shares of goods produced at home and abroad.

Gains from Trade

To explore the gains from trade, we follow Matsuyama (2008) and solve for consumption in the home and foreign countries as $c(z) = w/p(z)$ and $c^*(z) = w^*/p^*(z)$, so that welfare of the representative consumer at home and abroad is $U = \int_0^1 \ln [w/p(z)] dz$ and $U^* = \int_0^1 \ln [w^*/p^*(z)] dz$. In autarky we use the prices $p(z) = wa(z)$ and $p^*(z) = w^*a^*(z)$ to determine welfare

$$U_A = - \int_0^1 \ln [a(z)] dz \quad \text{and} \quad U_A^* = - \int_0^1 \ln [a^*(z)] dz. \quad (3.31)$$

Under free trade we use the prices in (3.29) to solve for welfare as

$$U_T = - \int_0^{z'} \ln [a(z)] dz + \int_{z'}^1 \ln \left[\frac{w}{w^* a^*(z)} \right] dz, \quad \text{and} \quad (3.32a)$$

$$U_T^* = \int_0^{z'} \ln \left[\frac{w^*}{wa(z)} \right] dz - \int_{z'}^1 \ln [a^*(z)] dz. \quad (3.32b)$$

The gains from trade are then equal to the difference between expressions (3.32) and (3.31) for the home and foreign countries, respectively, and using (3.28) we obtain

$$U_T - U_A = \int_{z'}^1 \ln \left[\frac{A(z')}{A(z)} \right] dz \quad \text{and} \quad U_T^* - U_A^* = \int_0^{z'} \ln \left[\frac{A(z)}{A(z')} \right] dz. \quad (3.33)$$

We see that both countries gain from trade under the weak conditions that $0 < z' < 1$ (otherwise there is no trade at all) and $A(0) > A(z') > A(1)$, which means that the relative cost of producing different goods is not constant for all z . If it is not the case, then countries would not have goods they are relatively efficient in producing and thus there is no reason to trade. These conditions allow for gains from trade even if one of the countries has absolute advantage (meaning lower labor costs) in producing all goods. Such a country would have a higher wage than abroad, so that an internal marginal good z' still exists provided that $A(0) > A(1)$.

Beyond confirming the gains from trade, we are interested in how these gains change when there is growth in one country. From (3.30), we see that a rise in the foreign size L^* shifts up or to the left the $B(z)$ curve, as illustrated in figure 3.10. As a result, the critical value z' falls to z'' and the home relative wage rises (provided that the $A(z)$ is not flat). Then from (3.32), the change in welfare as a result of trade is computed as²⁰

$$\begin{aligned} dU_T &= \int_{z'}^1 \left(\frac{dw}{w} - \frac{dw^*}{w^*} \right) dz = (1 - z') \left(\frac{dw}{w} - \frac{dw^*}{w^*} \right) > 0, \text{ and} \\ dU_T^* &= \int_0^{z'} \left(\frac{dw^*}{w^*} - \frac{dw}{w} \right) dz = z' \left(\frac{dw^*}{w^*} - \frac{dw}{w} \right) < 0. \end{aligned}$$

We can think of w/w^* as the terms of trade for the home country, which is rising, and conversely the terms of trade w^*/w for the foreign country is falling, leading to the rise in home welfare and fall in foreign welfare due to foreign growth.

While this result follows from the model, it does not necessarily capture the effect of growth in, say, China relative to that in the United States or Europe. Chinese growth is not from the size of population itself, but rather, from the effectiveness with which its labor resources can be transformed into goods. In other words, we should allow for reductions in the labor coefficients $a(z)$ and $a^*(z)$. To this end, we again follow Matsuyama (2008) and let $g(z) \equiv -d \ln [a(z)] > 0$ and $g^*(z) \equiv -d \ln [a^*(z)] > 0$ denote the rate of productivity growth in industry z at home and abroad. Totally differentiating (3.32) once again, we obtain

$$dU_T = \int_0^{z'} g(z) dz + \int_{z'}^1 g^*(z) dz + (1 - z') \left(\frac{dw}{w} - \frac{dw^*}{w^*} \right), \text{ and} \quad (3.34a)$$

$$dU_T^* = \int_0^{z'} g(z) dz + \int_{z'}^1 g^*(z) dz + z' \left(\frac{dw^*}{w^*} - \frac{dw}{w} \right). \quad (3.34b)$$

The first terms $\int_0^{z'} g(z) dz + \int_{z'}^1 g^*(z) dz$ in these expressions are positive, reflecting the direct beneficial effect of productivity growth, but once again relative wages will change and this creates an opposite effect in the two countries. To solve for the change in relative wages, we first use $\ln A(z') = \ln B(z')$ to solve for the change in the marginal good as

$$\left[\eta(z') + \frac{1}{z'(1-z')} \right] dz' = [g(z') - g^*(z')],$$

²⁰Note that the change in the marginal good z' , which appears in the limits of integration in (3.32), cancels out when we differentiate by using (3.28).

where $\eta(z) \equiv -d\ln[A(z)]/dz > 0$. When the foreign country experiences higher productivity growth for the marginal good than at home, $g^*(z') > g(z')$, then z' will decline and the foreign country will export an expanded range of goods $[z', 1]$. We now use (3.28) to solve for the change in relative wages

$$\left(\frac{dw}{w} - \frac{dw^*}{w^*}\right) = [g(z') - g^*(z')] - \eta(z') dz' = \left[\frac{g(z') - g^*(z')}{z'(1-z') \eta(z') + 1}\right].$$

It is apparent from this expression that relative wages in each country depend positively on their own productivity growth, but with a coefficient $1/[z'(1-z') \eta(z') + 1]$ that is less than unity. Substituting this expression into (3.34), we obtain the changes in welfare due to productivity growth

$$dU_T = \int_0^{z'} g(z) dz + \int_{z'}^1 g^*(z) dz + (1-z') \left[\frac{g(z') - g^*(z')}{z'(1-z') \eta(z') + 1} \right], \text{ and} \quad (3.35a)$$

$$dU_T^* = \int_{z'}^1 g^*(z) dz + \int_0^{z'} g(z) dz + z' \left[\frac{g^*(z') - g(z')}{z'(1-z') \eta(z') + 1} \right]. \quad (3.35b)$$

We can see that productivity growth in one's *own* country always raises welfare: there is a positive contribution from the first term in each expression and another positive contribution from the final term due to the improvement in own-country relative wages (and a terms of trade gain for that reason). But productivity growth in the *partner* country has offsetting effects: it raises welfare from the second term in each expression, but lowers welfare from the final term as there is a rise in the partner's relative wages (and a terms of trade loss at home for that reason). This is the situation applying to the United States or Europe, for example, in its trade with China: productivity growth in China lowers the prices of its export and benefits the rest of the world, but rising wages increase those prices and offset that gain. The question is whether productivity growth abroad can ever be *harmful* to the home country.

To answer this question, we first consider a uniform rate of foreign productivity growth, $g^*(z) = g^* > 0$. We can substitute this into (3.35a) and simplify to obtain

$$dU_T = \int_0^{z'} g(z) dz + (1-z') g^* + (1-z') \left[\frac{g(z') - g^*}{z'(1-z') \eta(z') + 1} \right] > 0.$$

It is readily apparent that this expression is positive because $[z'(1-z') \eta(z') + 1] > 1$, so that the harmful effect of rising foreign wages (in the final term) can never offset the beneficial effect of the foreign productivity growth (in the second term). But this conclusion is special, and a counterexample using non-uniform productivity growth can be constructed. Matsuyama (2008) uses an example from Jones (1979) that has three sectors.²¹ Within each sector identical goods are produced, with relative labor requirements $A(z) = A_1 = a_1^*/a_1$ for $z \in [0, \alpha_1]$ in sector 1, $A(z) = A_2 = a_2^*/a_2$ for $z \in (\alpha_1, \alpha_1 + \alpha_2]$ in sector 2, and $A(z) = A_3 = a_3^*/a_3$ for $z \in (\alpha_1 + \alpha_2, 1]$ in sector 3, with $A_1 > A_2 > A_3$, as illustrated in figure 3.11.

We assume that in the initial equilibrium, the curve $B(z)$ intersects $A(z)$ in its middle segment, where the relative wage is $w/w^* = A_2$. Notice that the home country

²¹Krugman (1994) also presents this three sector example and attributes it to Sir W. Arthur Lewis.

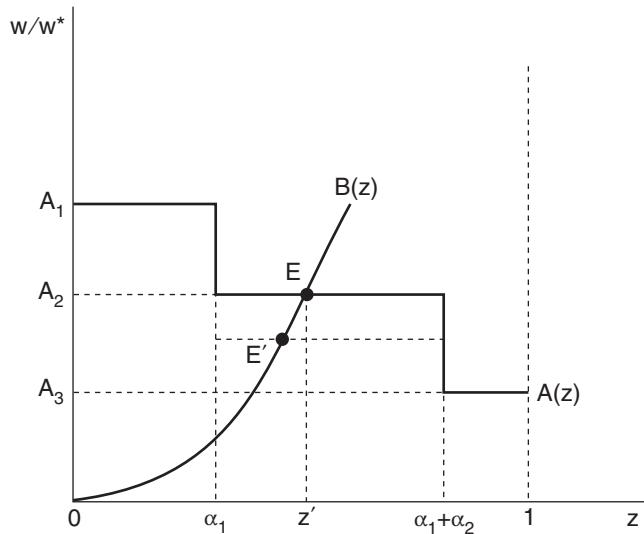


Figure 3.11

exports all of the goods in sector 1 and some of the goods in sector 2, within the interval $[\alpha_1, z']$; while the foreign country exports the rest of the goods in sector 2, within the interval $(z', \alpha_1 + \alpha_2]$, and all of the goods in sector 3. Now suppose the foreign country improves its productivity for producing goods in sector 2 only, so that a_2^* declines and the equilibrium moves from point E to point E' in figure 3.11. Because the home country is still exporting some goods in sector 2 (though over a reduced range), it is still the case that the world price is $p_2 = wa_2 = w^*a_2^*$. If normalizing the home wage at unity, $w=1$, then $p_2 = a_2$ and so it is apparent that the foreign wage w^* must *rise by the full amount* of the productivity improvement, $dw^*/w^* = -da_2^*/a_2^*$. The increase in the foreign wage is passed through to an *increased price for all goods in sector 3*, imported by the home country. The home country therefore experiences a terms of trade loss and a decline in welfare, caused by the increase in the price of goods in sector 3.

This example shows how the effects of foreign productivity growth on the home country are quite subtle, especially in a model with more than two goods so that the two countries are competing “head-to-head” in some markets. Despite the possibility of loss that we have illustrated, there are still gains from trade overall (i.e., as compared to autarky) for the home country. Matsuyama (2008) considers other examples, such as the foreign, lagging country catching up to close a technological gap, or the possibility of “offshoring” labor services abroad (as we examine formally in the next chapter). Once again, it is possible for the home country to lose due to these changes, though it still gains from trade overall.

CONTINUUM OF GOODS: HECKSCHER-OHLIN MODEL

We now consider the case with two factors of production. When there is continuum of goods, then the FPE set as shown in figure 3.8, which had piecewise linear edges, instead has smooth edges as shown in figure 3.12. If endowments of the two countries lie strictly inside the FPE set then factor prices will be equalized and there will be many combinations of outputs of the goods that can be produced in each country and satisfy

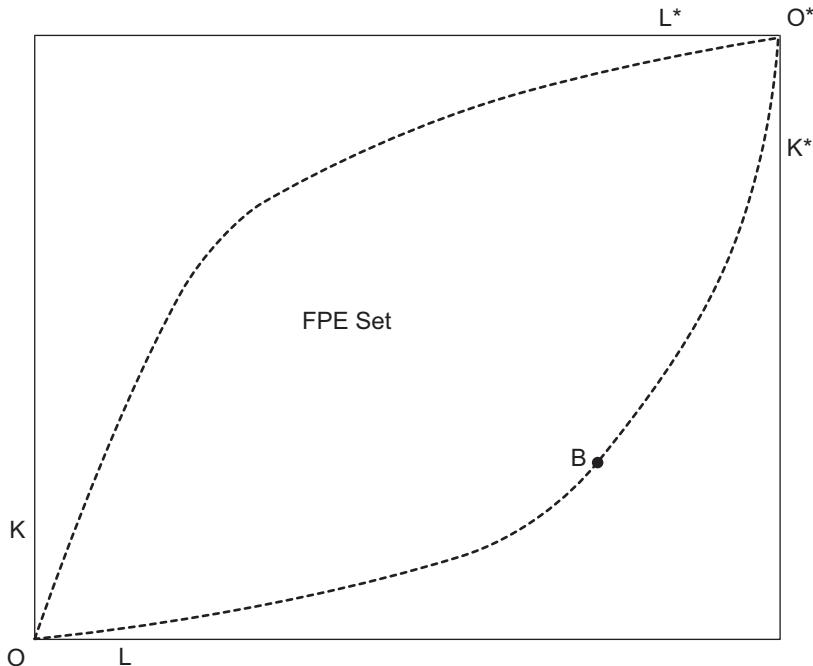


Figure 3.12

full employment. In other words, production indeterminacy still holds when there is factor price equalization. If factor endowments are sufficiently different, however, and lie outside the FPE set, then factor prices will differ in the two countries. A borderline case is shown by point B in figure 3.12, right on the boundary of the FPE set, which would allow for factor price equalization; but for endowments just slightly to the right of point B we would be outside the set and factor prices would differ.

Outside the FPE set the indeterminacy of outputs no longer applies. To demonstrate this result, we again use the index $z \in [0, 1]$ to denote the continuum of goods, and let $c(w, r, z) = wa_L(z) + ra_K(z)$ denote the unit-cost function for good z using the home factor prices (w, r) . We are assuming for simplicity a Leontief production function where $a_L(z)$ and $a_K(z)$ are the amounts of labor and capital needed for one unit of production of good z . Furthermore, without loss of generality we index goods so that $a_L(z)/a_K(z)$ is non-decreasing in z , meaning that the most capital-intensive goods have the lowest index z and the most labor-intensive goods have the highest index z . Following the Heckscher-Ohlin model, we assume the same production and cost function in the foreign country, so that with different factor prices the unit-cost abroad are denoted by $c(w^*, r^*, z) = w^*a_L(z) + r^*a_K(z)$.

With this framework, each good z will be produced in the country that has the lower unit-costs, so that the equilibrium prices will be determined by

$$p(z) = p^*(z) = \min \{c(w, r, z), c(w^*, r^*, z)\}.$$

To fix ideas, we will suppose that the home country is relatively abundant in labor, and has a lower wage/rental ratio than that abroad, $(w/r) < (w^*/r^*)$. With this assumption,

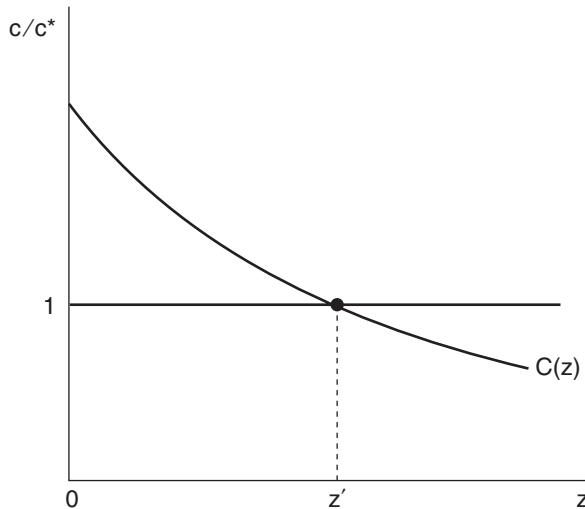


Figure 3.13

we can graphically illustrate the problem of choosing the minimum cost location for each good. For convenience, we will graph the ratio of unit-costs at home to those abroad, or

$$\frac{c(w, r, z)}{c(w^*, r^*, z)} = \left\{ \frac{(w/r)[a_L(z)/a_K(z)] + 1}{(w^*/r^*)[a_L(z)/a_K(z)] + 1} \right\} \frac{r}{r^*}. \quad (3.36)$$

From our assumptions that $(w/r) < (w^*/r^*)$ and that $a_L(z)/a_K(z)$ is non-decreasing in z , it follows that the ratio of unit-costs in (3.36) is non-increasing in z .²² This shape is shown by the line $C(z)$ in figure 3.11, where for convenience we illustrate the ratio of unit-costs as strictly decreasing in z . The reader will recognize that the curve $C(z)$ in figure 3.13 for the Heckscher-Ohlin model looks very similar to $A(z)$ in figure 3.9 for the Ricardian model, which illustrates the underlying similarity of these two models once we assume a continuum of goods.

We have labeled the point z' where the unit-costs are just equal in the two countries, and by construction we see that $c(w, r, z) > c(w^*, r^*, z)$ for $z < z'$ and $c(w, r, z) < c(w^*, r^*, z)$ for $z > z'$. In other words, the foreign country has comparative advantage (lower relative costs) in, and will produce and export, the lower z or capital-intensive goods, whereas the home country has comparative advantage in, and will produce and export, the higher z or labor-intensive goods. This pattern fits with our assumption that the home country is labor abundant, and shows that production indeterminacy in the Heckscher-Ohlin model *no longer holds* when factor prices are not equalized. We now make use of this result and return to the estimation of the Heckscher-Ohlin-Vanek (HOV) model that we began in the previous chapter, describing the results of Davis and Weinstein (2001).

²²To show this result, assume that $d[a_L(z)/a_K(z)]/dz > 0$ in (3.36). Then using our assumption that $(w/r) < (w^*/r^*)$, differentiate (3.36) to show that $d[c(w, r, z)/c(w^*, r^*, z)]/dz < 0$.

ESTIMATING THE HOV MODEL ONCE AGAIN

Recall from the previous chapter that the key issue in estimating the HOV model is to allow for some technological differences across countries, while avoiding the introduction of so many free parameters that the HOV equation holds as an equality (as in Trefler 1993a; Trefler and Zhu 2010). In other words, we want to model the technological differences in a parsimonious manner.

Our discussion of the Dornbusch, Fischer, and Samuelson (1980) model above immediately suggests a simple empirical technique. If factor prices are not equal, then the capital/labor ratios of traded goods must be systematically related to each country's endowments: a country with a capital/labor endowment higher than its neighbor's must have higher capital/labor intensities in *all of its* traded goods. Let us model this relation by

$$\ln a_{jk}^i = \alpha^i + \beta_{jk} + \gamma_k \left(\frac{K^i}{L^i} \right) + \varepsilon_{jk}^i, \quad (3.37)$$

where $i=1,\dots,C$ denotes countries, $j=1,\dots,N$ denotes traded goods, $k=1,\dots,M$ denotes factors, (K^i/L^i) is the relative capital/labor endowment, and ε_{jk}^i is a random error.

Davis and Weinstein use a sample of twenty OECD countries (plus the rest of the world), thirty-four sectors, and two factors (labor and capital) around 1985. The regression in (3.37) is run over countries, goods, and factors, where the estimated right-hand side is a measure of "predicted" factor requirements. We can think of the parameters β_{jk} as providing *average* estimates of the factor requirements across countries. The predicted factor requirements in *each* country differ from this average depending on the parameters α^i , which measure uniform technology difference across countries as in chapter 2,²³ and depending on $\gamma_k(K^i/L^i)$, which measures the impact of factor endowments. Thus, this simple regression permits a parsimonious description of technology differences across countries.

As noted by Harrigan (2003), there is considerable prior evidence that a specification like (3.37) will fit the data quite well: Dollar, Wolff, and Baumol (1988, table 2.3), for example, find that capital per worker in individual industries is highly correlated with capital per worker in overall manufacturing. Davis and Weinstein (2001) likewise find that the simple linear relationship (3.37) fits well, and that the parameters γ_k linking factor requirements to country endowments are highly significant. Using several variants of this regression, they construct the estimated technology matrix \hat{A}^i for each country using the predictions from (3.37). This estimated matrix is then used to construct the factor content of trade and to evaluate how well the constructed factor contents match the relative endowments of countries. In other words, we can proceed much as in the earlier testing of the HOV model, but now using technology matrices \hat{A}^i , which differ across countries.

As discussed near the end of the last chapter, when the technology matrices differ across countries, we should use the technology of the *exporting* country to construct

²³In chapter 2 we used the parameter δ^i , which was unity for the most efficient country (the United States) and less than unity otherwise. This can be related to α^i by $\delta^i = \exp(-\alpha^i)$, where now δ^i will be unity for the *average* country.

the factor content of trade. Davis and Weinstein follow this approach, and so the factor content of exports from country i to country j is

$$\hat{F}^{ij} \equiv \hat{A}^i X^{ij}, \quad (3.38)$$

where X^{ij} is the vector of gross exports of goods from country i to j . Davis and Weinstein sometimes use an additional assumption, whereby these exports are themselves estimated *in proportion to the purchasing country's GDP*. In that case, the factor content of export is

$$\hat{F}^{ij} \equiv \hat{A}^i \hat{X}^{ij}. \quad (3.39)$$

In either case, the prediction of the HOV model is that the factor content of trade reflects a country's relative endowments, or that

$$V^i - s^i \left(\sum_j V^j \right) = \left(\sum_{j \neq i} \hat{F}^{ij} \right) - \left(\sum_{j \neq i} \hat{F}^{ji} \right). \quad (3.40)$$

In the sign test, for example, if we assume that all OECD countries use the same *average* technology matrix, then we include only the parameters β_{jk} in regression (3.37) and construct the predicted technology matrix \hat{A} , which is the same across countries. In that case, Davis and Weinstein initially find that the right and left sides of (3.40) have the same sign in only 46 percent of the cases.²⁴ This is no better than a coin toss, much as we found in chapter 2. If instead we allow for uniform productivity differences between countries, so the regression (3.37) is estimated with α^i in addition to β_{jk} , then we find that the sign test is satisfied in 50 percent of the cases. If we also include $\gamma_k(K^i/L^i)$ in regression (3.37), so that the estimated technologies differ systematically with countries' factor endowments, then the sign test is satisfied in fully 86 percent of the cases. Finally, if we model exports as proportional to the purchasing countries GDP, so factor contents are measured with (3.39) rather than (3.38), the sign test is satisfied in 92 percent of the cases!

Clearly, we are getting closer to a 100 percent match between relative endowments and constructed factor contents, on the left and right side of (3.40). The theorem of Trefler and Zhu (2010) reviewed in the previous chapter tells us that this is no coincidence: if regression (3.37) fit perfectly, and assuming that every good is exported in proportion to the purchasing country's GDP, then (3.40) would automatically hold. What is remarkable is that even the simple specification in (3.37) is enough to greatly improve the fit of the HOV model. The contribution of Davis and Weinstein is to systematically show the importance of each modeling assumption to "closing the gap" between factor contents and endowments.

CONCLUSIONS

We have covered a lot of ground in this chapter. As we leave the "even" structure of the two-by-two or the $N \times N$ model, many of the results we found in chapter 1 are lost. With more factors than goods, the "factor price insensitivity" lemma no longer

²⁴We average the results reported for labor and capital in Davis and Weinstein (2001, table 4).

applies. This can be tested by estimating a GDP function for a country (or a group of countries), and determining whether it is linear in the factor endowments, that is, whether the functional form in (3.7) holds globally or the elasticities in (3.22) are locally insignificantly different from zero. The results of Kohli (1990a, 1993a) for the United States indicate that the elasticities in (3.22) are nonzero, but only weakly so. This can be interpreted as evidence in favor of “more factors than goods.” One reason for this is that capital might be slow to move between sectors, so with N sectors there are N fixed factors (capital in each sector) plus the additional mobile factors (types of labor). Kohli (1993b) develops a test for this production structure and finds some evidence to support it using annual data for the United States.

With more factors than goods, the GDP function is generally well behaved: it will be differentiable in prices and endowments provided that industries do not have the same factor intensities. With more goods than factors, however, the GDP function is poorly behaved: the fact that the production possibility frontier (PPF) has “ruled segments,” as illustrated in figure 3.7, indicates that the GDP function is not differentiable in prices. A test for “more goods than factors” is developed by Bernstein and Weinstein (2002), and they find evidence to support it using data from prefectures in Japan. This may seem inconsistent with the results of Kohli (1993b) for the United States, but note that Kohli uses time-series data whereas Bernstein and Weinstein use cross-sectional data. We should not be surprised to find differing conclusions when comparing the time series of national data versus the cross-section of provincial data. In any case, the results of both Kohli and Bernstein and Weinstein support the idea that we are *not* in an “even” model, with equal numbers of goods and factors. This structure had been assumed by Leamer (1984) and Harrigan (1995), for example. The more general approach of Harrigan (1997), using a GDP function estimated across OECD countries, could also be used to test “factor price insensitivity” by evaluating the elasticities in (3.22), though this was not done due to data limitations.

We have also reviewed the Ricardian and Heckscher-Ohlin models with a continuum of goods, due to Dornbusch, Fischer, and Samuelson (1977, 1980). By ranking goods according to their comparative advantage in one country, and identifying the borderline commodity with equal costs at home and abroad, the Ricardian model neatly shows how the trade pattern between countries is determined. We showed how this model can be used to illustrate the effect of technological change on welfare, but this is only one example and there are many other important applications to international trade and international macroeconomics. Matsuyama (2008) gives a valuable overview of results. Matsuyama (2013) further shows how this model—combined with increasing returns to scale as we shall cover in chapter 5—can be used to generate *endogenous productivity differences* across countries. In other words, Matsuyama (2013) provides an explanation for the technology differences across countries that we concluded in the previous chapter were so important.

The Heckscher-Ohlin model with a continuum of goods was used to motivate the empirical work of Davis and Weinstein (2001). Using the idea that industry factor intensities should depend systematically on economy-wide factor endowments, Davis and Weinstein estimate technology matrices across OECD countries and use these to test the HOV equation. We already know from the theorem of Trefler and Zhu (2010), discussed at the end of the last chapter, that using *actual* technology matrices (and an assumption on exports) can lead to a perfect fit for this equation. What is surprising is that Davis and Weinstein (2001) are able to obtain a very close fit even with a parsimonious specification of the technology matrices across countries. In case the reader

thinks that this is the last word on the HOV model, however, we note that there are still important issues left unresolved by Davis and Weinstein.

First, Trefler and Zhu (2000) have criticized Davis and Weinstein's approach for not adequately distinguishing between trade in final goods and trade in intermediate inputs. All of the theory we have developed for the HOV equation assumes trade in final goods, and yet, a great deal of actual trade is in steel, chemicals, textiles, and other intermediate inputs. Indeed, component parts can cross borders multiple times before being incorporated into a finished good. This makes the construction of the factor content of trade very difficult: if Germany imports steel from Korea, and Korea relies on coal from Australia to produce the steel, and Australia uses imported mining equipment from Germany, then whose factor requirements should we use to construct the factor content of steel trade? These types of flows tend to cancel out when we consider only *net* exports of each country, as in the original formulation of the HOV model. However, once the technology matrices differ across countries, then totaling up the *gross* exports as on the right side of (3.40), can lead to considerably larger values for trade, much of which is due to intermediate inputs. It is hard to see, then, how achieving near equality of (3.40) amounts to a validation of the HOV model, when that model has so little to say about trade in intermediate inputs. Trefler and Zhu (2010) and Davis and Weinstein (2003) concur that modeling trade in intermediate inputs is an important area for further research, as we shall investigate in the next chapter.

Second, since technological differences across countries are so important, then rather than working with the HOV model we could instead go back to the Ricardian model—but with a continuum of goods—and ask how well technological differences match up with actual trade patterns. That task is made possible by adding a random element to productivity differences across countries, as done by Eaton and Kortum (2002) and discussed in chapter 6. Costinot, Donaldson, and Komunjer (2012) develop a many-sector, many-country version of the Eaton-Kortum model and use it to test Ricardo's ideas, going well beyond the confines of the simple two-country Ricardian model. We can expect that extending that work to include many factors of production will be a topic for future research, and one that we begin to investigate in the next chapter.

PROBLEMS

- 3.1 In the specific factors model, suppose that the home and foreign countries have identical labor endowments, and identical capital in sector 2, but the *home* country has more capital in sector 1. Technologies and tastes are the same across countries.
 - (a) Can we predict the trade pattern between the countries? What factor(s) at home benefit from the opening of trade, and which factor(s) lose?
 - (b) Repeat part (a), but now assume that the endowments of capital are identical across the countries, and that the *home* country has more *labor*. Can we predict the trade pattern between the countries? Try to use assumptions on the factor cost-shares to help determine this.
- 3.2 Consider the translog GDP function $\ln G(p, V)$ in (3.15).
 - (a) Show that $G(p, V)$ is homogeneous of degree one in prices when (3.16) and $\gamma_{ij} = \gamma_{ji}$ hold.

- (b) Show that $G(p, V)$ is homogeneous of degree one in endowments when (3.17) and $\delta_{k\ell} = \delta_{\ell k}$ hold.
- 3.3. Suppose that there are two outputs, nontraded good 1 and traded good 2, and three inputs: labor, capital, and imported inputs. Assume that the cost-share of labor is higher in good 1 and the cost-share of capital is higher in good 2. The prices of both outputs are treated as fixed initially, but the price of the imported input increases.
- Write down the zero-profit conditions and totally differentiate these.
 - Determine the impact of the increase in the imported input price on the wage and rental, assuming that the traded good has a higher cost-share of imports.
 - If the price of the nontraded good changes endogenously due to the increase in the price of imported inputs, how will this affect your conclusions in (b)?

4

Trade in Intermediate Inputs and Wages

The first “golden age” of trade lasted from about 1890 to the beginning of World War I. That period saw dramatic improvements in transportation, such as the steamship and railroads, as well as wireless communication, which greatly facilitated increases in international trade. These observations led Eli Heckscher and Bertil Ohlin to modify the assumptions of the Ricardian model of trade, and suppose that technologies would spread quickly among countries while resource endowments were the domestic constraint. The world is now arguably in a second “golden age” of trade, which like the first, has relied on declines in costs of transportation, such as the container ship, as well as communication. These costs have now fallen so much that it is possible to break apart the production process, with various stages occurring in different countries. This is the idea of “fragmentation” of the production process (Jones 2000). There are many examples of this fragmentation, which is alternatively described as “foreign outsourcing,” or simply “offshoring,” the term that we shall use.¹

This new feature of globalization means that the spread of technology is even more rapid than in the time of Heckscher and Ohlin. The ability to utilize labor in other countries suggests that domestic resources are no longer the binding constraint on international trade. We might expect, therefore, that a new paradigm is needed to describe this second golden age of trade. Or is it? Is the Heckscher-Ohlin (HO) model—in all its manifestations—sufficiently rich to guide our understanding of offshoring, or does it leave out some critical elements? That will be the first question we shall address in this chapter.

We begin by reviewing evidence from the United States and other countries on the changes in the wages of low-skilled versus high-skilled workers since the 1980s.² During that decade relative wages shifted toward more-skilled workers, so that a

¹ Alternatively referred to as outsourcing (Katz and Murphy 1992; Feenstra and Hanson 1996), delocalization (Leamer 1996), fragmentation (Jones 2000; Arndt and Kierzkowski 2001; Marjit and Acharyya 2003), intraproduct specialization (Arndt 1998a, 1998b), intramediate trade (Antweiler and Trefler 2002), vertical specialization (Hummels, Ishii, and Yi 2001), and slicing the value chain (Krugman 1995), this phenomena refers to the geographic separation of activities involved in producing a good (or service) across two or more countries.

² See the surveys by Feenstra (1998), Feenstra and Hanson (2003), Freeman (1995), Johnson and Stafford (1999), Katz and Autor (1999), Richardson (1995), and Wood (1995), and the volumes by Bhagwati and Kosters (1994), Cheng and Kierzkowski (2001), Collins (1998), and Feenstra (2000).

“wage gap” developed between those with higher and lower skills. This observation led to a good deal of debate among scholars as the cause of the change in wages, and in particular, whether *international trade* or *technological change* was the proximate cause. It is fair to say that many writers looking at the 1980s rejected international trade as the cause of the change in wages, for reasons we shall discuss, while favoring instead technological change. But as we shall argue, this rejection of international trade was really a rejection of a simple version of the HO model, relying on two goods and two factors. In fact, once we move beyond that simplest version of the HO model, it is quite possible for international trade (in conjunction with technological change) to explain the global pattern of wage movements.

Feenstra and Hanson (1996, 1997, 1999) adopt a Heckscher-Ohlin structure with a continuum of goods, though the goods represent intermediate inputs *within* an industry that are then bundled into a final product. In that case, it turns out that the pattern of wage changes in the 1980s are entirely consistent with international trade, and in particular, it is possible to observe an increase in the relative wage of high-skilled labor *and* an increase in its relative demand, as occurred during the 1980s. But the story for the 1990s and later is quite different. There has continued to be an increase in the relative wage of skilled workers in U.S. manufacturing, but the relative employment of these workers has fallen. That finding is strongly suggestive of the offshoring of service activities, whereby the more routine service activities are sent overseas. This new form of offshoring can be explained by the work of Grossman and Rossi-Hansberg (2008, 2012), emphasizing what they call “trade in tasks.” Their model allows for the offshoring of tasks performed by low-skilled labor, and also allows for the offshoring of tasks that use *high-skilled labor*, like service tasks, as is needed to explain the more recent empirical observations for the United States.

EVIDENCE FROM U.S. MANUFACTURING

We begin by examining the pattern of wages over time in U.S. manufacturing. In figures 4.1 and 4.2, I use data from the manufacturing sector to measure the wage of “nonproduction” relative to “production” workers. As their name suggests, nonproduction workers are involved in service activities, while production workers are involved in the manufacture and assembly of goods. These two categories can also be called “nonmanual” versus “manual,” or “white collar” versus “blue collar.” Generally, nonproduction workers require more education, and so we will treat these workers as skilled, while production workers are less skilled.

In figure 4.1, we see that the earnings of nonproduction relative to production workers moved erratically from the late 1950s to the late 1960s, and from that point until the early 1980s, relative wages were on a downward trend. It is generally accepted that the relative wage fell during this period because of an increase in the supply of college graduates, skilled workers who moved into nonproduction jobs. Starting in the early 1980s, however, this trend reversed itself and the relative wage of nonproduction workers increased steadily to 2000, with erratic movements thereafter. The same increase in the relative wage of skilled workers has been found for other industrial and developing countries.

Turning to figure 4.2, we see that there has been a steady increase in the ratio of nonproduction to production workers through the end of the 1980s, but then a fall in the 1990s, with some recovery since 2000. The increase in the relative supply of

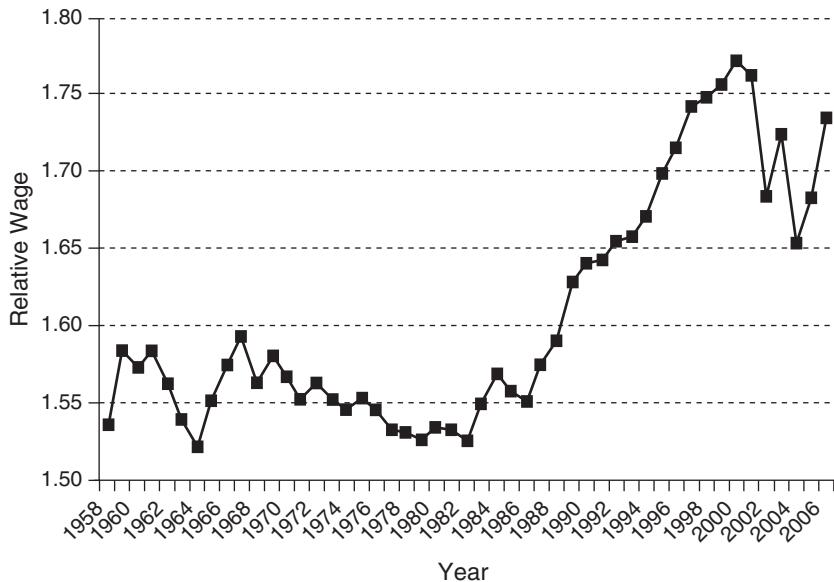


Figure 4.1: Relative Wage of Nonproduction/Production Workers, U.S. Manufacturing, 1958-2006

Source: National Bureau of Economic Research productivity database, updated after 1996 from Bureau of the Census.

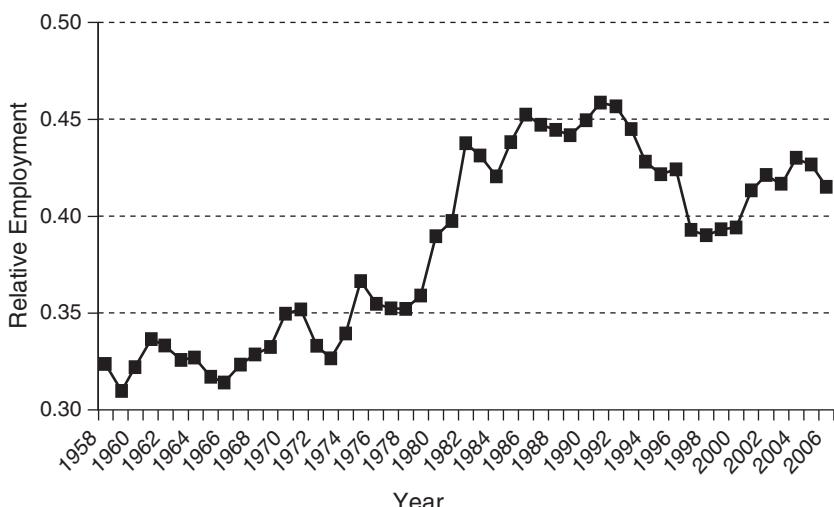


Figure 4.2: Relative Employment of Nonproduction/Production Workers, U.S. Manufacturing, 1958-2006

Source: National Bureau of Economic Research productivity database, updated after 1996 from Bureau of the Census.

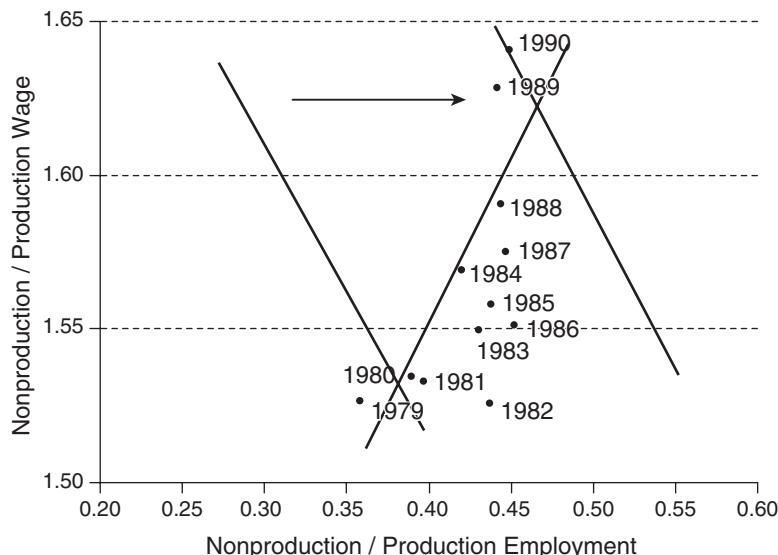


Figure 4.3: Relative Wage and Employment of Nonproduction/Production Workers in U.S. Manufacturing, 1979–90

Source: National Bureau of Economic Research productivity database.

workers through the 1970s can account for the *reduction* in the relative wage of non-production workers that decade, as shown in figure 4.1, but is at odds with the *increase* in the relative nonproduction wage during the 1980s. The rising relative wage should have led to a shift in employment *away* from skilled workers, along a demand curve, but it did not. Thus, the only explanation consistent with these facts is that there was an *outward shift* in the demand for more-skilled workers during the 1980s, leading to an increase in their relative employment and wages, as shown in figure 4.3. for the United States and which is also found for other countries.³

What factors can lead to an outward shift in the relative demand for skilled labor? Such a shift can arise from the use of computers and other high-tech equipment, or *skill-biased technological change*. Researchers such as Berman, Bound, and Griliches (1994) argue that such technological change was the dominant explanation for the rising relative wage of skilled labor in the United States and other countries. Their reason for rejecting international trade as an explanation was the finding that the majority of the increase in the manufacturing wage and employment of nonproduction workers was caused by shifts *within* industries, and not by shifts *between* industries. That is, the outward shift in relative demand being illustrated in figure 4.3 applied to many individual industries, as well as in the aggregate. In their view, that ruled out the HO

³The same decline in the relative wages of blue-collar workers during the 1980s and into the 1990s can be found for Australia, Canada, Japan, Sweden, and the United Kingdom (Anderton and Brenton 1999; Freeman and Katz 1994; Katz and Autor 1999), for Hong Kong and Mexico (Cragg and Epelbaum 1996; Hanson and Harrison 1999; Hsieh and Woo 2005; Robertson 2004), and for Denmark (Hummels et al. 2014) over the more recent period 1995–2006.

model as an explanation since in that model they expected to see a shift *between* industries instead of *within* industries.⁴

Their findings for the United States were reinforced by the work of Berman, Bound, and Machin (1998), who looked at cross-country data. They found that the same shift toward skilled workers in the United States also occurred abroad. That finding again appeared to rule out the HO model as an explanation, because in that model we expect wages to move in opposite directions between countries when comparing autarky to free trade, as factor price equalization occurs. Instead, the evidence was that wages were moving in the same direction across countries—with an increase in the relative wage of skilled workers.⁵

FACTOR BIAS VERSUS SECTOR BIAS OF TECHNOLOGICAL CHANGE

Leamer (1994, 1998) was among the first trade economists to contribute to this debate. He rejected the claim that skill-biased technological change could explain the shifts in wages for the United States because, in his view, the factor bias of technical change is not important: only the *sector bias* matters. To make this argument, he starts with the zero-profit conditions for industries $i=1,\dots,N$, which are written as

$$p_i = \sum_{j=1}^M a_{ij} w_j \quad i=1,\dots,N. \quad (4.1)$$

Differentiating these and allowing for exogenous changes in the factor requirements a_{ij} , we obtain

$$\hat{p}_i = \sum_{j=1}^M \theta_{ij} \hat{w}_j + \sum_{j=1}^M \theta_{ij} \hat{a}_{ij} \quad i=1,\dots,N, \quad (4.2)$$

where the cost-shares are $\theta_{ij} = a_{ij} w_j / p_i$. The second term above is the negative of total-factor productivity growth, which is defined as

$$\text{TFP}_i \equiv \hat{y}_i - \sum_{j=1}^M \theta_{ij} \hat{x}_{ij} = -\sum_{j=1}^M \theta_{ij} \hat{a}_{ij}. \quad (4.3)$$

⁴Bernard and Jensen (1997) responded to Berman, Bound, and Griliches (1994) by using plant-level rather than just firm-level data for the United States, and showed that a majority of the shift toward nonproduction workers in the 1980s occurred *between* plants. In particular, they found that the plants experiencing the greatest increase in relative nonproduction employment and earnings were precisely those that were engaged in exporting. The results of Bernard and Jensen therefore provide *prima facie* evidence that international trade has had an impact on factor demand and wages, through shifting the demand for labor between plants but within industries.

⁵The empirical technique initiated by Berman, Bound, and Griliches (1994) was a regression of the change in the nonproduction share of the wage bill on explanatory variables. Feenstra and Hanson (1999) later added offshoring and the use of high-technology capital as explanatory variables to this regression, showing that both made a significant contribution to the rise in the nonproduction share for the United States. Evidence supporting the link between offshoring or import penetration and wages or employment is also available for Austria (Dell'mour et al. 2000), Germany (Geishecker 2004), Hong Kong (Hsieh and Woo 2005), Japan (Head and Ries 2001), Mexico (Feenstra and Hanson 1997); and the United Kingdom (Anderson and Brenton 1999; Görg, Hijzen, and Hine 2003, 2005). More recent evidence for the United States is provided by Voigtlander (2014).

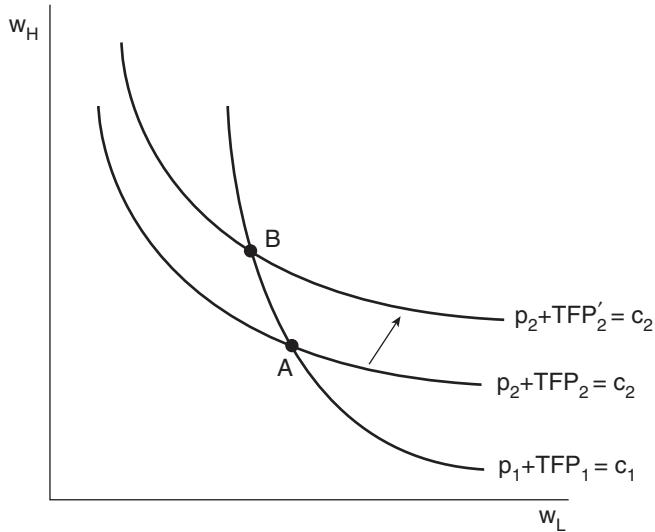


Figure 4.4

The second equality follows because $a_{ij} = x_{ij}/y_i$, and so $\hat{a}_{ij} = \hat{x}_{ij} - \hat{y}_i$, and we take the sum over j . Therefore, the differentiated zero-profit conditions are stated as

$$\hat{p}_i + \text{TFP}_i = \sum_{j=1}^M \theta_{ij} \hat{w}_j, \quad i=1, \dots, N. \quad (4.4)$$

Now suppose that the country is small, so that prices do not change. Then it is immediate from this equation that the sector-bias of technological change, or TFP_i , will determine the change in factor prices. This argument can be illustrated quite simply in figure 4.4, where we graph the zero-profit condition in two industries, where low-skilled labor earns the wage of w_L and high-skilled labor earns the wage of w_H . Suppose that there is technological progress affecting *either* factor in industry 2, which is skilled-labor intensive. Then that industry can afford to pay more to both factors, so its zero-profit contour shifts up. As a result, the wage earned by high-skilled labor rises and the wage for low-skilled labor falls, regardless of whether the technological progress was biased toward one factor or the other. That is the point that Leamer is making.

This relationship between changes in prices and wages is called the “mandated wage equation,” and is estimated as a regression of the log change in industry prices on factor shares, by expressing (4.4) in discrete time:

$$\Delta \ln p_i + \text{TFP}_i = \sum_{j=1}^M \theta_{ij} \beta_j, \quad i=1, \dots, N. \quad (4.5)$$

We interpret the coefficients β_j as the change in factor prices that are “mandated by” the change in product prices, which is the dependent variable in (4.5). We hope to find that the estimated factor price changes β_j from this regression are quite close to their actual changes, and if so, we can conclude that a Stolper-Samuelson linkage between product and factor prices works empirically.⁶

⁶Baldwin and Hilton (1984) were among the first to estimate this price regression, and later applications include Baldwin and Cain (2000), Slaughter (2000), and the studies discussed in the main text.

TABLE 4.1: Log Change in Industry Price, 1979–90, as Dependent Variable

	(1)	(2)	(3)	(4)	(5)
Effective TFP				-1.0 (0.01)	-1.0 (0.001)
TFP	-1.0 (0.1)	-0.8 (0.1)			
Production labor cost-share	3.1 (1.2)	2.4 (1.2)	3.6 (1.9)	4.7 (0.02)	4.7 (0.01)
Nonproduction labor cost-share	2.3 (1.4)	4.1 (1.7)	6.2 (4.0)	5.5 (0.02)	5.4 (0.03)
Capital cost-share	7.9 (0.8)	8.1 (0.9)	9.5 (2.2)	4.0 (0.01)	4.0 (0.02)
Materials cost-share times change in materials price	1.0*	1.0*	1.2 (0.3)	1.0*	1.0 (0.002)
Energy cost-share times change in energy price	1.0*	1.0*	-0.9 (0.9)	1.0*	1.0 (0.01)
Constant	-0.7 (0.3)	-0.83 (0.29)	-1.9 (0.9)		0.01 (0.005)
R^2	0.90	0.81	0.93	0.99	0.99
N	447	446	446	447	447

Source: Feenstra and Hanson (1999) and empirical exercise 4.3.

Notes: Standard errors are in parentheses. All regressions omit three industries with missing data on materials purchases or prices (SIC 2067, 2794, 3483) and are weighted by the industry share of total manufacturing shipments, averaged over the first and last period.

In columns (1)–(3) and (5), the dependent variable is the log change in the gross industry price, and the factor cost-shares sum to one across all factors. The materials cost-share is multiplied by the log change in the materials price; the energy cost-share is treated similarly. In column (4), the dependent variable is the log change in the industry value-added price, and factor cost-shares sum to one across primary factors. Column (1) uses primal TFP as a regressor; column (2) drops the computer industry (SIC 3573) from the sample; column (3) also drops TFP as a regressor; and column (5) uses effective TFP as a regressor, where effective TFP equals primal TFP minus the change in wage differentials.

*Constrained at unity.

Estimates of (4.5) for 447 U.S. manufacturing industries, over 1979–90, are provided in table 4.1. The dependent variable is the log change in the industry output price over the period, and total factor productivity (TFP) growth is included on the right, measured in discrete time as

$$\text{TFP}_i = \Delta \ln y_i - \sum_{j=1}^M \theta_{ij} \Delta \ln x_{ij}. \quad (4.6)$$

The other independent variables in the regression are the average cost-shares for production labor, nonproduction labor, and capital; the materials cost-share times the log change in materials prices; and the energy cost-share times the log change in energy prices. Note that we do not attempt to estimate the change in materials prices and energy prices, but simply include these cost-shares times their prices as controls.

In columns (1) and (2), we constrain the coefficients on the materials-share times the materials price, and the energy-share times the energy price, to be unity. This approximates the specification in Leamer (1998). In column (1), the coefficients on the labor shares imply a *decrease* in the nonproduction/production relative wage of $2.3\% - 3.1\% = -0.8\%$ per year, which is consistent with the results in Leamer (1998). But in reality, the nonproduction-production wage gap in the United States *rose* by 0.7

percent per year, or about seven-tenths of one percent.⁷ So the regression in column (1) does *not reproduce at all* the actual factor price changes in the United States! In column (2), we follow Sachs and Shatz (1994) and drop the computer industry (SIC 3573), which *reverses* the predicted change in wage inequality. Now nonproduction wages are mandated to rise by 1.5 percent per year *more* than production wages. In column (3), we approximate Krueger's (1997) specification by dropping TFP as a regressor, while estimating coefficients on materials and energy. There is again a mandated *rise* in the nonproduction-production wage gap, but one that is *much larger* than the actual increase in relative wages.

The estimates in table 4.1 are troubling because they show that the estimates of β are quite far off the mark: they do not reflect the actual change in wages that occurred in the United States during the 1980s, and are quite sensitive to the data used and specification of the regression.⁸ While it is true that the computer industry is an outlier, the sensitivity of the results to the specification suggests that something more basic is going on. We will argue that there is a good explanation for why this regression does not work as well as expected, as explained later in the chapter, but first we describe the response that Krugman gives to Leamer's arguments.

Krugman (2000) argues that Leamer's conclusions hold only in a small-country model with fixed world prices. But Krugman goes on to make an important theoretical point: if world prices are endogenously determined, and under the simplifying assumption of Cobb-Douglas preferences, then the sector bias of technological changes completely cancels out and has no impact at all on factor prices. Instead, he argues, *only* the factor bias matters.

To make this point, Krugman considers a closed economy with two sectors and two factors, high-skilled and low-skilled labor, where we assume for convenience that preferences are Cobb-Douglas. Now suppose that either one of the sectors has Hicks-neutral technological progress: how does that affect the relative demand for labor? The answer is not at all: Hicks-neutral progress lowers the price of that good and raises its demand by just the amount needed to leave factor demands unchanged.⁹ Since factor-demands are unchanged, then the relative wage is also unchanged by neutral technological progress. Furthermore, this result continues to hold in a two-country HO model with factor price equalization, provided that the Hicks-neutral technological shift is worldwide.

So in strong contrast to Leamer's small-country case, in the world economy the sector bias of Hicks-neutral technological change does not matter at all. But the factor bias clearly does matter: if either sector has skilled-biased technological change, for example, then the demand for skilled labor shifts out, raising its relative wage. So the large-country case puts the focus squarely back onto the skill-bias of technological change. The problem with Krugman's argument, however, is that it leaves little scope for international trade to play a role in influencing relative wages.¹⁰ So for different

⁷After being weighted by each industry's share of manufacturing shipments, the average nominal wages of nonproduction workers rose by 5.4 percent per year over 1979–90, and the nominal wages of production workers rose by 4.7 percent per year. The difference between these, 0.7 percent, is the increasing “wage gap” that needs to be explained.

⁸Nevertheless, Leamer still refers to the 1970s as the “Stolper-Samuelson decade,” in the sense that that it is changes in product prices, and not TFP growth, that are principally responsible for the change in wages.

⁹See exercise 4.1.

¹⁰Krugman (2000) argues that computing the factor content of labor in U.S. imports is still the correct way to measure the pressure on U.S. wages from those imports. A limitation with this argument, however, is that

reasons, neither Leamer nor Krugman arrives at a satisfactory framework for comparing the impact of trade versus technological change on wages. That is the goal for the rest of the chapter.

TRADE IN INTERMEDIATE INPUTS

We have suggested that a simple two-good, two-factor HO model cannot explain the shift in relative labor demand toward skilled labor that has occurred in the United States and other countries. But a more general specification of the HO model can explain this shift. Feenstra and Hanson (1996, 1997) present a model of an industry in which there are many activities undertaken in the creation, production, and delivery of a product to the consumer, denoted by the index $z \in [0, 1]$. Rather than listing these activities in their *temporal* order, we will instead list them in *increasing order* of high-skilled/low-skilled labor, where, for example, the least skill-intensive activity is assembly and the most skill-intensive activity is R&D. Letting $x(z)$ denote the quantity produced of each one of these inputs, we let $a_H(z)$ and $a_L(z)$ denote the high-skilled and low-skilled labor, respectively, needed to produce one unit of $x(z)$. We order the activities z so that $a_H(z)/a_L(z)$ is non-decreasing in z .

We will suppose that there are two countries, with the foreign country denoted by an asterisk. The production functions for producing the two inputs are assumed to be the same up to a Hicks-neutral productivity parameter in each country:

$$x(z) = A \left[\min \left\{ \frac{L(z)}{a_L(z)}, \frac{H(z)}{a_H(z)} \right\} \right]^\theta K^{1-\theta}, \quad z \in [0, 1]. \quad (4.7)$$

Thus, the amount of high-skilled and low-skilled labor $H(z)$ and $L(z)$ are used in the home country to produce input z , using a Leontief technology between these two types of labor, and then a Cobb-Douglas technology between labor overall and capital K . The parameter θ denotes the share of labor in the costs of producing each input. The foreign production function is the same, except with a different productivity parameter A^* .

Rather than working with the production functions for the inputs in (4.7), it will be convenient to work with the dual unit-cost function:

$$c(w_L, w_H, r, z) = B[w_L a_L(z) + w_H a_H(z)]^\theta r^{1-\theta}, \quad (4.8)$$

where $c(w_L, w_H, r, z)$ denotes the costs of producing one unit of $x(z)$ in the home country, given the wage of low-skilled labor w_L , the wage of high-skilled labor w_H , and the rental on capital r . The inputs can be produced in *either country*, and are then combined into the production of a final product. The production function for the final good is assumed to be a Cobb-Douglas function over all the inputs:

$$\ln Y = \int_0^1 \alpha(z) x(z) dz, \quad \text{with} \quad \int_0^1 \alpha(z) dz = 1. \quad (4.9)$$

factor content calculations generally results in rather small implied changes in wages. Feenstra and Hanson (2000) and Feenstra (2010a) suggest that those calculations might be biased downward.

Note that in (4.9) we have not included labor as an input, so the final good is assembled “costlessly” from all the individual inputs $z \in [0, 1]$. This means that we do not need to keep track of which country the assembly takes place in, because there is zero value added (and zero demand for labor) in this activity.

In general, firms doing the assembly will wish to source the inputs from the minimum-cost location. To determine this, we will make the following assumption on factor prices:

$$\frac{w_H}{w_L} < \frac{w_H^*}{w_L^*}, \quad \text{and} \quad r < r^*. \quad (4.10)$$

The first assumption in (4.10) states that the relative wage of skilled labor is *lower* at home than abroad, which is realistic if the home country is skilled-labor abundant, like the United States: despite the increase in the relative wage of skilled labor in the United States during the past two decades, it is still much lower than in Mexico. The second assumption states that the rental on capital is lower at home, so that if capital is able to move, it will want to relocate abroad. This is also realistic if the home country is capital abundant, and will provide the basis for some comparative statics we shall consider.

Then just like the HO model with a continuum of goods, in a trade equilibrium we will find that countries specialize in different portions of the skill continuum. Under our assumption that the relative wage of skilled labor is higher abroad, and that goods are arranged in increasing order of their skill intensity, then the ratio of the home to foreign unit-costs is downward sloping, as shown by the schedule c/c^* in figure 4.5. Foreign production—or offshoring—occurs where the relative costs at home are greater than unity, in the range $[0, z']$, whereas home production occurs where the relative costs at home are less than unity, in the range $(z', 1]$. The borderline activity z' is determined by equal unit-costs in the two countries

$$\frac{c(w_L, w_{H^*}, r, z')}{c^*(w_L^*, w_{H^*}^*, r^*, z')} = 1. \quad (4.11)$$

Using this unique borderline activity z' , we can then calculate the demand for labor in each country. At home, for example, the relative demand for high-skilled/low-skilled labor is

$$D(z') = \frac{\int_{z'}^1 \frac{\partial c}{\partial w_H} x(z) dz}{\int_{z'}^1 \frac{\partial c}{\partial w_L} x(z) dz}. \quad (4.12)$$

It can be shown that this schedule is a downward sloping function of the relative wage. A downward sloping relative demand curve applies to the foreign country, too, where now we integrate over the activities in $[0, z']$. In both countries, equilibrium factor prices are determined by the equality of relative demand and supply.

Suppose now that the home firm wishes to offshore more activities. The reason for this could be a capital flow from the home to the foreign country, reducing the rental abroad and increasing it at home; or alternatively, technological progress abroad, neutral across all the activities, but exceeding such progress at home. In both cases, the

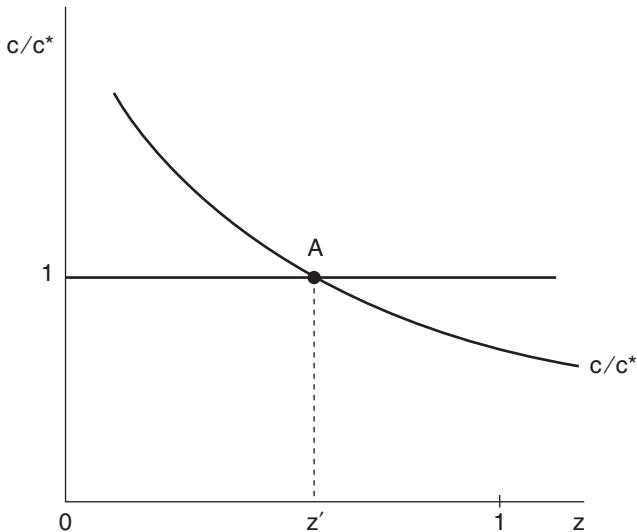


Figure 4.5

relative costs of production at home rise, which is an upward shift in the relative cost schedule, as shown in figure 4.6. As a result, the borderline between the activities performed at home and abroad therefore shifts from the point z' to the point z'' , with $z'' > z'$.

What is the impact of this increase in offshoring on the relative demand for skilled labor at home and abroad? Notice that the activities no longer performed at home (those in between z' and z'') are *less* skill-intensive than the activities still done there (those to the right of z''). This means that the range of activities now done at home are more skilled-labor intensive, on average, than the set of activities formerly done at home. For this reason, the relative demand for skilled labor at home increases, as occurred in the United States during the 1980s. That increase in demand will also increase the relative wage for skilled labor, as shown in figure 4.7.

What about in the foreign country? The activities that are newly sent offshore (those in between z' and z'') are *more* skill-intensive than the activities that were initially done in the foreign country (those to the left of z'). That means that the range of activities now done abroad is also more skilled-labor intensive, on average, than the set of activities formerly done there. For this reason, the relative demand for high-skilled labor in the foreign country also increases, just as shown in figure 4.7. With this increase in the relative demand for high-skilled labor, the relative wage of skilled labor *also* increases in the foreign country. This is a realistic description of what occurred in Mexico (Cragg and Epelbaum 1996; Hanson and Harrison 1999; Robertson 2004), Hong Kong (Hsieh and Woo 2005), and other countries. In contrast, it is difficult to generate the *same direction of movements* in relative wages across countries from the HO model, as we have found in the offshoring model.

Can we say anything about absolute rather than relative wages? Feenstra and Hanson (1996) explore this, and conclude that while the *relative* wage of low-skilled workers fall in both countries, their *real wages* need not fall. Low-skilled workers in the home country are the most disadvantaged by the offshoring, but nevertheless, it is possible that they gain due to lower prices of the final good. These results hold

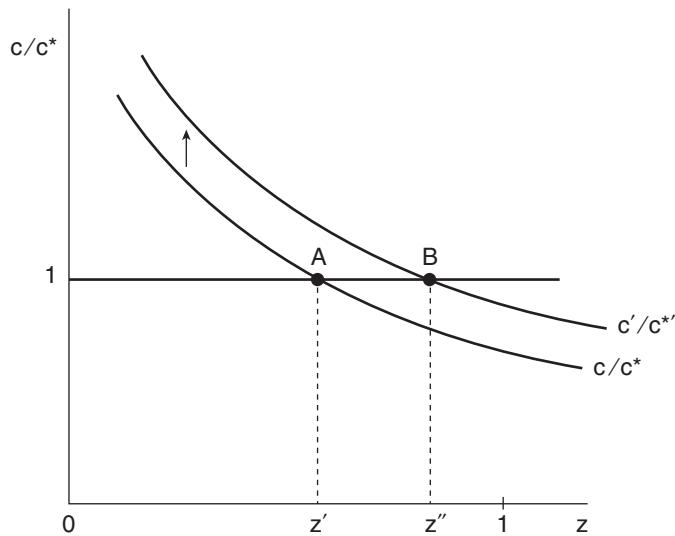


Figure 4.6

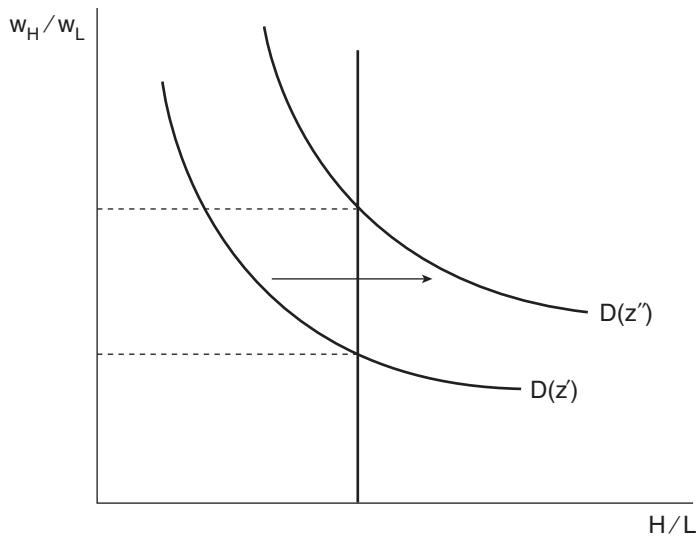


Figure 4.7

regardless of whether the increased offshoring is due to a capital flow, as we have described, or due to a variety of other causes: growth in the capital endowments abroad, at a rate exceeding that at home; or simply technological progress abroad exceeding that at home. In each of these cases, the comparative static effects are the same as we have just described.

Another question, not explored by Feenstra and Hanson, is the extent to which factor accumulation or a reduction in trade barriers increases trade between countries. With trade due to offshoring, it would appear that modest shifts in the relative cost loci in figure 4.6 could potentially lead to large changes in the borderline activity z'

and in the amount of offshoring. Yi (2003, 2005) argues that this is indeed the case in a Ricardian model with a continuum of goods, and that “vertical specialization” accounts for more than 50 percent of the growth in U.S. trade since 1962. These papers are recommended for further reading.

To summarize, this model of Feenstra and Hanson, which borrows the structure of the HO model with a continuum of goods, gives an explanation for the increase in the relative demand for skilled labor that was observed across countries during the 1980s. Of course, this explanation does not *prove* that offshoring was the source of the wage changes, since *skill-biased technological change* is equally well an explanation. So determining which of these explanations accounts for the changes observed during the 1980s is an empirical question.

Mandated Wage Equation

To address that issue, Feenstra and Hanson (1999) start again with the mandated wage equation suggested by Leamer, but modify this equation in a fundamental way. They recognize that in any dataset, the wages paid to high-skilled and low-skilled labor differ across industries. They incorporate those inter-industry wage differentials into the model by rewriting the zero-profit conditions from (4.1) as

$$p_i = \sum_{j=1}^M a_{ij} w_{ij}, i=1, \dots, N,$$

where the wages w_{ij} now differ across industries i and factor j . Differentiating these zero-profit conditions we obtain

$$\hat{p}_i + TFP_i = \sum_{j=1}^M \theta_{ij} \hat{w}_{ij}, i=1, \dots, N, \quad (4.13)$$

where the cost-shares are $\theta_{ij} = a_{ij} w_{ij}/p_i$ and TFP_i is still defined as in (4.3).

When this regression is run as a mandated wage equation, as we did earlier in the chapter, we are treating the factor price changes as common across industries, thereby ignoring the inter-industry wage differentials. That is, letting $\beta_j \equiv \hat{w}_j$ denote the *average* value of the change in factor price j across industries, we are actually running the regression

$$\hat{p}_i + TFP_i = \sum_{j=1}^M \theta_{ij} \beta_j + \varepsilon_i, i=1, \dots, N, \quad (4.14)$$

where the error term is

$$\varepsilon_i \equiv \sum_{j=1}^M \theta_{ij} (\hat{w}_{ij} - \hat{w}_j), \quad (4.15)$$

and reflects the difference between the change in the industry and average factor prices. That is, we replace the *industry* wage changes on the right of (4.13) by the *average* wage changes in (4.14), and incorporate the difference between these two into an error term. In economic terms, ε_i reflects inter-industry wage differentials: that is, the difference between wages paid in each industry and the manufacturing average. It is well known that these wage differentials vary systematically across industries, with capital-intensive industries paying higher wages (Katz and Summers 1989a,b).

Recognizing that the regression equation has an error term, we can answer the question of whether the estimates from the equation will match the actual factor price changes or not: the estimates of β_j obtained from (4.14) will be unbiased estimates of the *average actual* factor price changes only if the error term ε_i shown in (4.15) is *uncorrelated* with the cost-shares θ_{ij} . This result follows directly from the properties of ordinary least squares (OLS), whereby the independent variables need to be uncorrelated with the error term to obtain unbiased estimates. But this property is unlikely to be true in our data. Industries such as computers have both a high share of nonproduction labor (e.g., engineers), and probably the fastest-growing industry wage differential, as these workers have had very rapid wage gains. This suggests that the error term ε_i is correlated with the nonproduction labor cost-share. Indeed, this correlation likely explains why the *estimated* change in nonproduction wages is *lower* in column (1) of table 4.1, which includes the computer industry, than in columns (2) or (3), which exclude this industry.

To correct this problem, we can simply include the error term ε_i as an additional regressor in the equation, reflecting the change in inter-industry wage differentials. To make this argument carefully, let us start with a *dual* definition of total factor productivity (TFP) growth, which is

$$\text{TFP}_i \equiv \sum_{j=1}^M \theta_{ij} \Delta \ln w_{ij} - \Delta \ln p_p, i = 1, \dots, N. \quad (4.16)$$

Comparing this with (4.6), we see that instead of comparing the growth of output with a weighted average of the growth in inputs, the dual definition in (4.16) compares a weighted average of the growth in factor prices with the growth in the product price. An industry experiencing productivity growth can afford to increase its factor prices by more than the growth in its product price.

While (4.16) is the usual definition of dual TFP growth, we also consider a dual measure of *effective* TFP, using the change in economy-wide factor prices rather than industry wages

$$\text{ETFP}_i \equiv \sum_{j=1}^M \theta_{ij} \Delta \overline{\ln w_j} - \Delta \ln p_p, i = 1, \dots, N, \quad (4.17)$$

where $\Delta \overline{\ln w_j}$ refers to the economy average of the change in log wages. Then we can rearrange terms in this equation to obtain a mandated wage equation that holds as an identity:

$$\Delta \ln p_j + \text{ETFP}_i \equiv \sum_{j=1}^M \theta_{ij} \Delta \overline{\ln w_j}, i = 1, \dots, N. \quad (4.18)$$

Running this regression gives estimated coefficients on the average factor shares that exactly match the economy's average changes in factor prices. This results are shown in columns (4) and (5) of table 4.1. In column (4), we constrain the coefficients on the materials and energy shares to equal unity, while in column (5), we allow these coefficients to differ from unity. In either specification, the coefficients on the labor and capital shares are extremely close to the actual average annual percentage changes in factor prices, which are 4.7 percent for production labor, 5.4 percent for nonproduction labor, and 4.0 percent for capital. Thus, when we properly estimate (4.18), we end up with an identity!

To make further progress, Feenstra and Hanson (1999) propose a two-step estimation procedure. In the first step, we combine the variables $\Delta \ln p_i + \text{ETFP}_i$, which appear in (4.18), and regress these on some structural variables. Any structural variables that shift the production function and therefore affect costs should be included. In the model discussed above with a continuum of input, the *range of inputs* imported into each country should appear as a structural variable.¹¹ In practice, however, Feenstra and Hanson include just offshoring and the use of high-tech equipment. Denoting these structural variables by z_{1i} and z_{2i} , we run the regression

$$\Delta \ln p_i + \text{ETFP}_i = \alpha_0 + \alpha_1 \Delta z_{1i} + \alpha_2 \Delta z_{2i}. \quad (4.19)$$

In the second step, we then take the estimated coefficients $\hat{\alpha}_1$ and $\hat{\alpha}_2$, and use these to construct the dependent variables for the following regressions,

$$\hat{\alpha}_1 \Delta z_{1i} = \sum_{j=1}^M \theta_{ij} \Delta \ln \beta_{1j} \text{ and}, \quad (4.20a)$$

$$\hat{\alpha}_2 \Delta z_{2i} = \sum_{j=1}^M \theta_{ij} \Delta \ln \beta_{2j}, \quad i=1, \dots, N. \quad (4.20b)$$

That is, we use the estimated coefficients $\hat{\alpha}_1$ and $\hat{\alpha}_2$, times each structural variable as the dependent variables in (4.20), and regress these on the factor cost-shares. The coefficients obtained from the second-stage regression, β_{1j} and β_{2j} , are interpreted as *the portion of the total change in factor prices that is explained by each structural variable*. In this way, we are taking the total change in factor prices, and decomposing it into parts that are explained by each structural variable.

Feenstra and Hanson (1999) measure offshoring by the share of intermediate inputs used in each industry that are imported. This measure is constructed from input-output tables using the so-called “proportionality” assumption, for example, that the automobile industry purchases the same proportion of imported to domestic steel as does the entire economy.¹² This measure of offshoring can be constructed using either a broad definition that includes all imported inputs, or a narrow definition which focuses on imported inputs within the same two-digit industry (e.g., the automobile industry importing auto parts). High-technology equipment is also measured in two ways: either as a fraction of the total capital equipment installed in each industry; or as a fraction of new investment in capital that is devoted to computers and other high-tech devices.

In table 4.2, the results from the broader measure of offshoring are reported, including imported inputs from other industries, for the 1980s. Using the first measure of high-tech equipment (i.e., fraction of the capital stock), the results in the first row show that roughly 25 percent of the increase in the relative wage of nonproduction workers was explained by offshoring, and about 30 percent of that increase was explained by the growing use of high-tech capital. So we conclude that both offshoring and the increased use of high-tech capital are important in explaining the actual

¹¹In chapter 5 we discuss methods to directly measure the product variety of imports. Dinopoulos and Segerstrom (1999b), Zhu (2004, 2005), and Zhu and Trefler (2005) develop models that emphasize the link between new varieties of traded goods and wages. These papers are recommended for further reading.

¹²The validity of this assumption is explored by Feenstra and Jensen (2012) for the United States and by Winkler and Milberg (2012) for Germany.

TABLE 4.2: Impact on the Relative Wage of Nonproduction Labor in U.S. Manufacturing, 1979–1990

	Percent of Total Increase Explained by each Factor	
	Offshoring	High-technology Equipment
Measurement of high-tech equipment:		
As a share of the capital stock	21–27%	29–32%
As a share of capital flow (i.e., new investment)	12%	99%

Source: Feenstra and Hanson (1999).

increase in the relative wage of skilled workers. In the second row we use the other measure of high-tech equipment (i.e., fraction of new investment). In that case, the large spending on high-tech equipment in new investment can explain *nearly all* (99 percent) of the increased relative wage for nonproduction workers, leaving little room for offshoring to play much of a role (it explains only 12 percent of the increase in the relative wage). These results are lopsided enough that we might be skeptical of using new investment to measure high-tech equipment and therefore prefer the results using the capital stocks.

Summing up, both offshoring and high-tech equipment are important explanations for the shift in demand toward nonproduction workers in U.S. manufacturing, though the relative contributions of the two measures are sensitive to how we measure high-tech equipment. But the results reported so far are incomplete because we have focused only on the decade of the 1980s. Let us now look at later trends.

EVIDENCE FROM 1990–2000

In figure 4.8 we report the relative wage and employment of nonproduction/nonproduction workers in U.S. manufacturing over 1990–2000. While the picture for the 1980s is well known and has launched dozens of research studies, it is surprising that the picture for the 1990s—shown in figure 4.8—is not yet familiar. We see that from 1989–2000 there continued to be an increase in the relative wage of nonproduction/nonproduction labor in U.S. manufacturing, but in addition, there was a *decrease* in the relative employment of these workers. There are two possible explanations for this shift suggested by the literature. First, some labor economists have argued that the 1990s witnessed a changing pattern of labor demand, benefiting those in the highest and lowest-skilled occupations, at the expense of others in moderately skilled occupations. Autor, Katz, and Kearney (2008, 301) attribute this once again to technological change: “we find that these patterns may in part be explained by a richer version of the skill-biased technical change (SBTC) hypothesis in which information technology complements highly educated workers engaged in abstract tasks, substitutes for moderately educated workers performing routine tasks, and has less impact on low-skilled workers performing manual tasks.”¹³

¹³ Other labor economists disagree, however, and argue that the changing patterns from the 1980s to the 1990s suggest that skill-biased technological change was not the right explanation in the first place. Examples of this so-called revisionist school of thought are Card and DiNardo (2002) and Lemieux (2006).

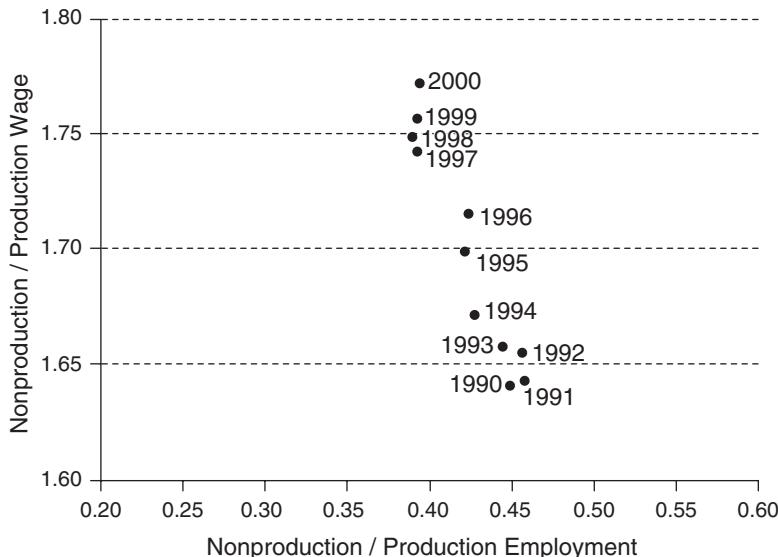


Figure 4.8: Relative Wage and Employment of Nonproduction/Production Workers in U.S. Manufacturing, 1990–2000

Source: National Bureau of Economic Research productivity database, updated after 1996 from Bureau of the Census.

A second possibility is that figure 4.8 is a “smoking gun” for service offshoring from U.S. manufacturing. To the extent that the back-office jobs being offshored from manufacturing use the lower-paid nonproduction workers, then the offshoring of those jobs could very well *raise the average wage* among nonproduction workers, while lowering their employment. So like we found for the 1980s, we once again have two different explanations for the change in wages and employment: the first drawing on technological change and the second emphasizing offshoring, but of a different type than was found in the 1980s. To choose between these explanations, we should ask whether there is any new theory to guide us.

TRADE IN TASKS

It turns out that there is new theory, due to Grossman and Rossi-Hansberg (2008). These authors prefer to think of “tasks” performed by high-skilled or low-skilled labor, rather than “activities” that combine factors, which is what Feenstra and Hanson (1996, 1997) used. They present a simple two-sector model of the economy, where in each sector and for each factor there is a continuum of tasks indexed by $i \in [0, 1]$. The production structure is Leontief in the sense that an equal amount of each task must be performed, but the firm can choose whether to do that task at home or offshore it. If the task is offshore, then the home firm uses its *own* technology abroad. While Grossman and Rossi-Hansberg do not specify whether the offshoring is done inside or outside of the firm, the fact that the home technology is transferred abroad suggests a multinational relationship in offshoring.

Focusing first on low-skilled labor, in each sector j the amount of labor a_{lj} is needed at home for every task i that is performed, where a_{lj} depends on the factor prices. If

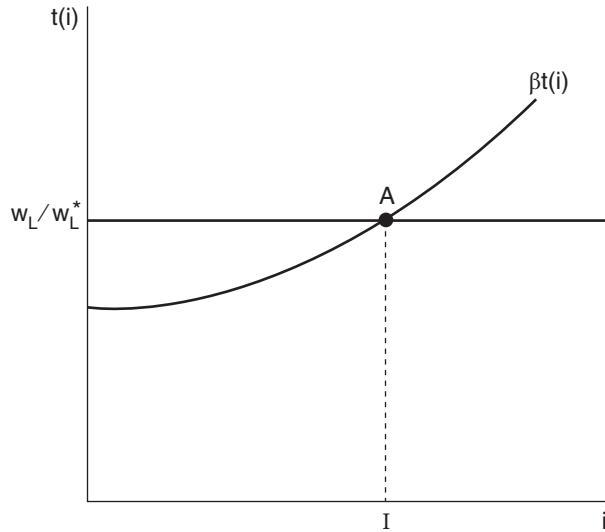


Figure 4.9

instead the task is offshored, then $\beta t(i) a_{Lj}$ units of low-skilled labor must be employed abroad. The tasks are ordered so that the function $t(i)$ is increasing, as shown in figure 4.9. The amount $[\beta t(i) - 1]$ indicates the “extra” labor that must be employed abroad to achieve the same outcome as one unit of labor at home. This formulation is similar to Samuelson’s (1952) iceberg transport costs, in the sense that it is the services of low-skilled labor themselves that are used up or “melt” in the offshoring process. We follow Grossman and Rossi-Hansberg in assuming that the offshoring costs $\beta t(i)$ are identical across the two sectors.

The equilibrium amount of offshoring is determined where the costs of performing the borderline task abroad, which is $\beta t(I) w_L^*$, equals its cost at home:

$$\beta t(I) w_L^* = w_L \Rightarrow \beta t(I) = w_L/w_L^*. \quad (4.21)$$

This equilibrium condition, shown in figure 4.9, needs to be supplemented with the zero-profit and the full-employment conditions. The zero-profit conditions are that the sum of costs of domestic and offshored labor for each unit of production equals the price, or

$$p_j = w_L a_{Lj} (1 - I) + w_L^* a_{Lj} \int_0^I \beta t(i) di + w_H a_{Hj}, \quad j=1,2, \quad (4.22)$$

where w_H is the wage of high-skilled labor, which we assume for now is not offshored. Using the equilibrium offshoring condition, zero profits are rewritten as

$$p_j = w_L a_{Lj} \Omega(I) + w_H a_{Hj}, \quad j=1,2, \quad (4.23)$$

where,

$$\Omega(I) = (1 - I) + \int_0^I t(i) di / t(I) < 1. \quad (4.24)$$

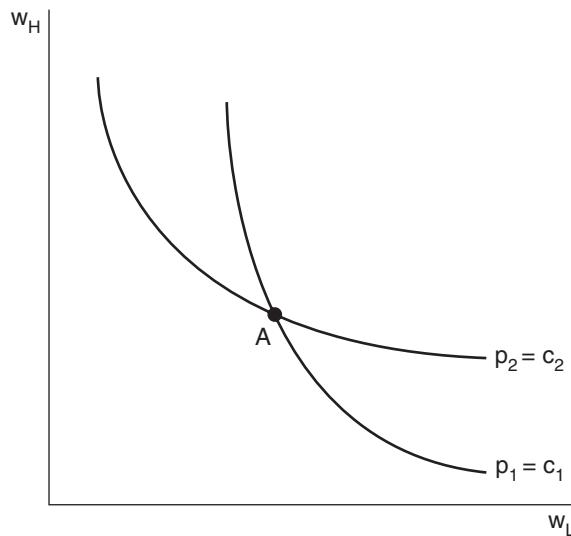


Figure 4.10

Notice that in this zero-profit condition, offshoring acts just like a low-skilled labor-saving technological innovation, or another form of skill-biased technological change. We can therefore graph two zero-profit conditions to determine the factor prices, as at point A in figure 4.10, recognizing that the iso-cost curves depend on the amount of offshoring.

Now suppose there is a reduction in the costs of offshoring, which is a fall in β . In figure 4.11 there is an increase in the amount of offshoring as I increases to I' , holding wages fixed for the moment. It is readily shown that $\Omega'(I) < 0$, and so offshoring acts like a low-skilled labor-saving innovation, which shifts both the iso-cost curves to the right horizontally in figure 4.12. The new equilibrium is established where the wage of low-skilled labor has increased, while the wage of high-skilled labor is unchanged. So low-skilled labor gains from offshoring because its productivity is enhanced: both the real wage and the relative wage of low-skilled labor rise, while high-skilled labor is not affected.

The prediction that the relative wage of low-skilled workers will rise is counter-intuitive, and it is clearly counterfactual to the experience of the United States and other countries during the 1980s and 1990s. It also contradicts the idea that skill-biased technological change—shifting relative demand towards skilled labor—should increase the relative wage of skilled workers.

To understand where this result is coming from, it is useful to go back to the debate between Leamer and Krugman discussed at the beginning of this chapter. Leamer (1994, 1998) considered a small-country model, in which case the sector-bias of technical change determines the change in wages. That is exactly what is occurring in this small-country version of the model by Grossman and Rossi-Hansberg. Offshoring acts like low-skilled labor-saving technical progress and has the greatest impact in the sector intensive in low-skilled labor, and so the real and relative wage of that factor rises.¹⁴ So these results confirm Leamer's thesis.

¹⁴Indeed, this idea was suggested earlier in the fragmentation literature, by Arndt (1997) and Jones and Kierzkowski (2001).

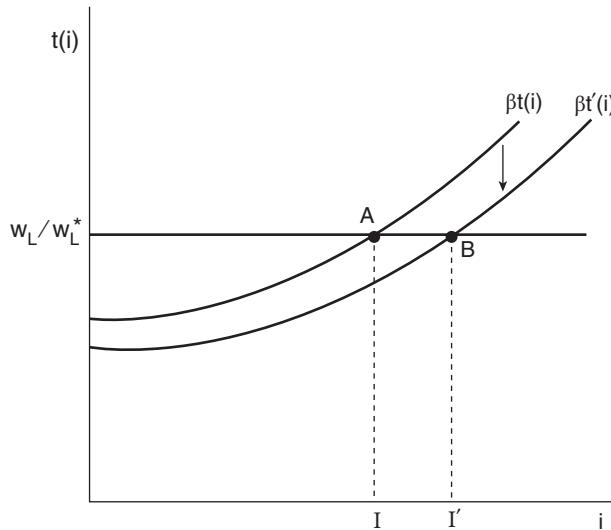


Figure 4.11

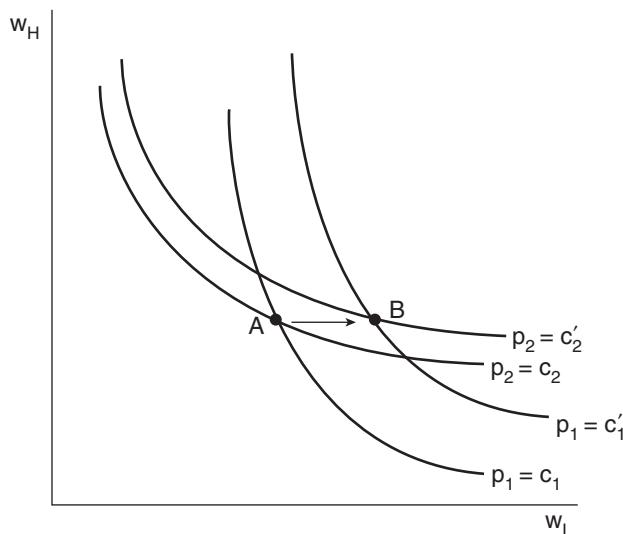


Figure 4.12

But Krugman would respond that we should instead focus on a large-country model, as Grossman and Rossi-Hansberg do next. In that case, we need to take into account how outputs change due to offshoring. The only way the low-skilled labor can remain fully employed at home in the presence of offshoring is for there to be a magnified increase in the output of the low-skilled-intensive sector. That result follows from the Rybczynski theorem, which holds in a modified version here. Since both sectors are offshoring the activities up to I , the full-employment condition for low-skilled labor is

$$\gamma_1 a_{L1}(1-I) + \gamma_2 a_{L2}(1-I) = L, \quad (4.25)$$

which is rewritten as

$$y_1 a_{L1} + y_2 a_{L2} = L/(1 - I).$$

Thus, a rise in offshoring will have the same impact on sector outputs as an effective increase in the endowment of low-skilled labor. Through the usual Rybczynski effect, this will have a magnified impact on the home output of the low-skilled-intensive good, and thereby also raise that output on world markets and lower its relative price. By the usual Stolper-Samuelson results, the rise in offshoring will *reduce* the relative wage of low-skilled labor. So the price effect works against low-skilled labor, whereas the productivity effect of offshoring works in its favor. In general, either of these effects can dominate, and so the relative wage can move in either direction.

To sharpen the results, suppose that preferences are Cobb-Douglas, as Krugman (2000) assumed. If production is also Cobb-Douglas, then any technological progress looks like Hicks-neutral, so we are almost back in Krugman's case where the price effect just offsets the productivity effect, and relative wages do not change at all. That result does not quite hold in Grossman and Rossi-Hansberg's model even with Cobb-Douglas technologies, because the productivity effect of offshoring applies only in the home country and not abroad. For that reason, we need to add more structure to obtain definite results on whether the price effect dominates the productivity effect or not.

Specifically, we follow Grossman and Rossi-Hansberg in assuming that both industries in the foreign country are uniformly less productive than at home, applying the Hicks-neutral productivity disadvantage $A^* > 1$ abroad. Note that Grossman and Rossi-Hansberg define this parameter in an *inverse* fashion to how it was used earlier in the chapter, so now if a foreign firm performs all tasks at the same factor intensities as a domestic firm, its output would be only $1/A^*$ times as great. In addition, the home country still has the low-skilled labor technological advantage of $\Omega(I) < 1$ due to offshoring, as described above. It is quite possible that there is "adjusted factor price equalization," meaning that $w_L \Omega = w_L^* A^*$ and $w_H = w_H^* A^*$, as we shall assume. The ratio of effective factor prices are $w_L \Omega / w_H = w_L^* / w_H^*$, and therefore equal across countries. It follows that the factor intensities are also equal, $a_{Lj} = a_{Lj}^*$ and $a_{Hj} = a_{Hj}^*$, $j = 1, 2$, where $A^* a_{Lj}^*$ and $A^* a_{Hj}^*$ are the foreign labor requirements per unit of output. The cost-shares are then $\theta_{Lj} \equiv w_L \Omega a_{Lj} / p_j = w_L^* A^* a_{Lj}^* / p_j$ for low-skilled labor, and $\theta_{Hj} \equiv w_H a_{Hj} / p_j = w_H^* A^* a_{Hj}^* / p_j$ for high-skilled labor, $j = 1, 2$.

With this notation, we can state the conditions under which the price effect dominates the productivity effect, so that the relative wage of high-skilled labor rises with offshoring, or when we obtain the converse result.

THEOREM (FEENSTRA 2010A)

Suppose that demand in both countries is Cobb-Douglas, with expenditure shares on the two goods of α_j , $j = 1, 2$. Then if the elasticities of substitution in production σ_j are sufficiently less than unity and the home country is sufficiently large, so that the following inequality holds

$$\sigma_j < \frac{L/\Omega}{\left(\frac{L}{\Omega} + \frac{L}{A}\right)} - \left\{ \frac{L^*/A^*}{\left(\frac{L}{\Omega} + \frac{L}{A}\right)} \left[\frac{\alpha_1 \alpha_2 (\theta_{H1} - \theta_{H2})^2}{\alpha_1 \theta_{H1} \theta_{L1} + \alpha_2 \theta_{H2} \theta_{L2}} \right] \right\}, \text{ for } j = 1, 2,$$

then the price effect dominates the productivity effect and the relative wage of high-skilled labor rises with increased offshoring. If this inequality is reversed for $j=1, 2$, then the relative wage of high-skilled labor falls.

We see that a *necessary* condition to obtain a rise in the relative wage of high-skilled labor is that the elasticities of substitution in production are less than unity (as obtained from the above inequality as $L \rightarrow \infty$). Recall that Feenstra and Hanson (1996, 1997) assumed that the elasticities of substitution in production were zero. Having a low elasticity of substitution is necessary but not sufficient, however,¹⁵ because only the home country is experiencing the effective technical progress due to offshoring. Therefore, we need the *added condition* that home is large enough as compared to the foreign country, so that the inequality in the theorem is satisfied. Under these conditions we obtain the results of Krugman (2000), that skill-biased technological change will raise the relative wage of high-skilled labor, which is most similar to the results of Feenstra and Hanson (1996, 1997). Conversely, if the inequality in the theorem is reversed, then the productivity effect necessarily dominates the price effect, and the relative wage of high-skilled labor will fall due to offshoring. That is the result in the small-country version of the model, for example.

Even without Cobb-Douglas preferences, results similar to the large-country case (with elasticities of substitution less than unity) occur if there are more factors than goods. For example, suppose there is only a single sector in both countries. That good will still be traded to compensate for the labor earnings from offshoring. In this case, Grossman and Rossi-Hansberg (2008) argue that there is a third effect at work, which they call the labor-supply effect. The effective increase in low-skilled labor due to the productivity effect cannot be absorbed by Rybczynski-like reallocation across sectors, and instead will lead to a fall in the relative wage of low-skilled labor. In this case, it turns out that if the initial amount of offshoring is small, then the labor-supply effect will definitely dominate the productivity effect, so that the relative wage of low-skilled labor falls. Once again, the real wage of low-skilled labor can move in either direction, while the real wage of high-skilled labor rises.

So we see that the small-country versus large-country assumption, and the number of sectors as compared to the number of factors, are crucial to the results.¹⁶ The large-country version has predictions that fit the facts for the United States in the 1980s. But the 1990s were different, and combined an increase in the relative wage of nonproduction workers with a *fall* in their relative demand. I have already suggested that the evidence is consistent with the offshoring of service activities from the United States, or the lower-paid of the nonproduction tasks. That outcome can also arise in the model of Grossman and Rossi-Hansberg, provided that we focus on the offshoring of *high-skilled* labor tasks rather than *low-skilled* tasks.

With offshoring of high-skilled labor, the equilibrium condition becomes

$$\beta t(I) w_H^* = w_H. \quad (4.26)$$

¹⁵The necessity of having a low elasticity of substitution follows from the results of Xu (2001): since offshoring acts like low-skilled labor-saving technical progress, this will shift relative expenditure away from that factor and towards high-skilled labor if and only if $\sigma_j < 1$.

¹⁶Analogously, Kohler (2001) has shown that the fragmentation results of Arndt (1997) no longer hold when capital is sector-specific, so there are more factors than goods. See also Kohler (2004).

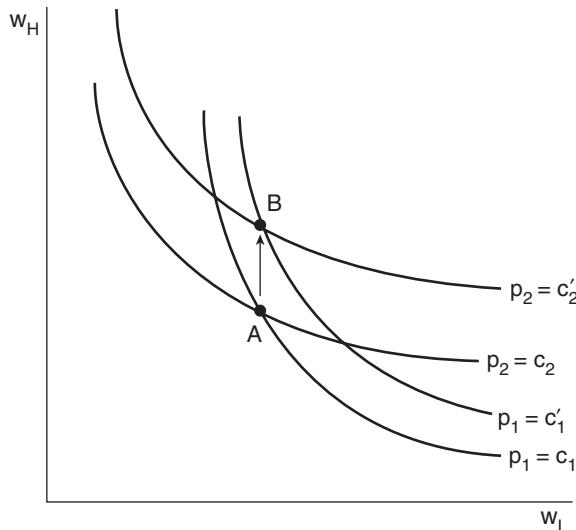


Figure 4.13

The zero-profit conditions are that the sum of the costs of domestic and offshored labor for each unit of production equal the price, or

$$p_j = w_L a_{Lj} + w_H a_{Hj}(1-I) + w_H^* a_{Hj}^* \int_0^I \beta t(i) di, j=1, 2. \quad (4.27)$$

Using the equilibrium offshoring condition, zero profits are rewritten as:

$$p_j = w_L a_{Lj} + w_H a_{Hj} \Omega(I), j=1, 2, \quad (4.28)$$

where $\Omega(I) < 1$ is defined as before.

Now offshoring acts just like a high-skilled labor-saving technological innovation. An increase in the amount of offshoring shifts the iso-cost curves vertically upward, as shown in figure 4.13. The offshoring of high-skilled tasks, which we are thinking of as service activities, leads to an increase in the relative wage of high-skilled labor and no change in the relative wage of low-skilled labor. Furthermore, such offshoring will reduce demand for skilled labor, at given industry outputs, as we saw in the United States during the 1990s. So an important contribution Grossman and Rossi-Hansberg's model is that it gives us a robust way to model this service offshoring in addition to the low-skilled offshoring of the 1980s. The rich specification of offshoring costs that are built into the model allow for a wide array of outcomes, and go beyond the Heckscher-Ohlin structure.

Evidence for the United States

There are several studies that document the growing importance of service offshoring for the United States. One study separates the impact of offshoring on production and nonproduction workers in U.S. manufacturing for the 1990s (Sitchinava 2008).

TABLE 4.3: Impact on the Relative Wage of Nonproduction Labor in U.S. Manufacturing, 1989–96

	Percent of Total Increase Explained by Each Factor		
	Materials Offshoring	Service Offshoring	High-technology Equipment
Relative wage of nonproduction labor	7%	15%	95%

Source: Nino Sitchinava (2008). Computed from the first-difference results in table 8.

It applies the two-step procedure of Feenstra and Hanson, using materials offshoring, service offshoring, as well as computer capital as potential explanations. These results are summarized in table 4.3, for 1989 through 1996. While the relative wage of nonproduction workers continued to rise during this period, materials offshoring explains only 7 percent of that increase. Service offshoring is twice as important, explaining some 15 percent of the increase in the relative wage. But the increased use of computers (as a share of the capital stock) can account for nearly all of the rise in the relative wage.

Other studies do not focus on the relative wage, but on other aspects of offshoring. Amiti and Wei (2009) find that for U.S. manufacturing, imported services grew from two-tenths of one percent of total inputs used in 1992, to three-tenths by 2000. The fact that imported services are so small does not prevent them from being important for productivity, however. They find that service offshoring accounts for around 10 percent of labor productivity growth over this period, which is quite robust across specifications, while material offshoring accounts for a smaller and less-robust 5 percent of labor productivity growth.

What about employment? We saw the relative employment of nonproduction workers fall during the 1990s, in marked contrast to the 1980s. Can we attribute that fall to tasks that require skills but are more routine, allowing them to be offshored? Amiti and Wei (2005a,b) do not identify a significant impact of service offshoring on employment, possibly because they work with a single aggregate of labor. But another study of white-collar employment in the United States (Crinò 2010), for both manufacturing and services, suggests that offshoring has an impact. Crinò finds that service offshoring raises high-skilled employment and lowers medium- and low-skilled employment. Within each skill group, there is a differential response depending on whether the tasks performed are classified as “routine,” requiring “face-to-face” communication or not, and “interaction with PCs,” which Crinò combines to obtain a “tradability” index for each occupation. Service offshoring is found to penalize the tradable occupations and benefit the nontradable occupations, which is consistent with the theory. Liu and Trefler (2011) also investigate how trade in services—specifically with China and India—has impacted U.S. labor markets. They find a significant increase in the number of workers who are displaced from the service sector, and in the portion of these workers who “switch down” to lower-paying occupations. Crinò (2008, 2013) further discusses the impact of service offshoring on productivity and labor demand in Western Europe.¹⁷

¹⁷For example, Lorentowicz, Marin, and Raubold (2008) and Marin (2005) have shown that the jobs being offshored from Germany and Austria to locations in Eastern Europe are in fact high-skilled jobs.

In order to develop his “tradability” index, Crinò relied on the O*NET database, which provides information on job characteristics for 812 occupations. Ebenstein et al. (2014), Grossman and Rossi-Hansberg (2006), Liu and Trefler (2011), and Odenski (2012, 2014) all find evidence that these O*NET characteristics make a difference for measuring the effects of offshoring in the United States. Wright (2014) also makes use of the O*NET data in an empirical framework that most closely mirrors the Grossman and Rossi-Hansberg model, and so we describe that work in some detail.¹⁸

Wright (2014) measures the extent of offshoring using the Feenstra-Hanson (1999) measure of the share of intermediate inputs used in each industry that are imported. He treats that offshoring share as endogenous, and to offset that concern, he focuses on intermediate inputs imported from China: or what we call “Chinese offshoring.” As an instrument for the *growth* in Chinese offshoring in each industry (over 2001–7), he uses the initial *share* of Chinese offshoring in each industry (in 1997), interacted with the overall growth (across all U.S. manufacturing) in Chinese offshoring. In other words, following Chinese entry into the WTO in 2001, Wright is presuming that the growth of U.S. imports is strongly predicted by the comparative advantage that China held in each industry prior to entry. With this approach Wright finds that “a one percentage point increase in the extent of offshoring is associated with a 0.29 percent direct decline in low-skill worker hours” (71). This direct negative impact of offshoring on low-skill employment is offset, however, when there is an expansion in the output of industries that are offshoring, as occurs in the Grossman-Rossi-Hansberg model. Wright finds evidence to support that increase in output and, as a result, the negative effect of offshoring on low-skilled employment, is reduced: “Combining the direct employment and output effects, on *average* offshoring can explain approximately 6% of the average annual decline in production worker employment over this period. This implies that offshoring is responsible for a net loss of 69,000 production worker jobs, a relatively minor impact during a period in which 1.2 million production worker jobs were lost” (73).

Wright goes on to construct an index of offshorability from the O*NET data and test whether the extent of low-skilled (production) jobs and high-skilled (nonproduction) job loss responds differently in industries that have greater offshorability: while more low-skilled jobs are lost when the tasks done in an industry are more routine, the opposite occurs for high-skilled jobs. That hypothesis is strongly confirmed in his results, and in fact, the gain of high-skilled jobs more than offsets the loss in low-skilled jobs. Adding together these two effects, he concludes that “the aggregate effect of offshoring is estimated to have led to a cumulative increase in aggregate employment of 2.6% over the period 2001 to 2007, a relatively minor effect” (76).

Perhaps the most compelling argument supporting the idea that offshoring in the 1980s was different from that in the 1990s and 2000s is provided by Basco and Mestieri (2013). They argue that trade liberalization in the 1980s increased trade in low-skilled goods (what they call L-globalization), while reduction in information and communication costs in the 1990s and 2000s increased trade in middle- and high-skilled intensive goods (what they call C-globalization). They present a model that integrates Feenstra and Hanson (1996, 1997) with Grossman and Rossi-Hansberg (2008), allowing for both trade in intermediate inputs and trade in tasks.¹⁹ The difference between the two

¹⁸Ottaviano, Peri, and Wright (2013) further estimate the effects of offshoring and immigration on jobs in the United States.

¹⁹Baldwin and Robert-Nicoud (2014) also integrate the trade-in-goods and trade-in-task models.

types of globalizations is that the reduction of trade costs applies to different segments in the continuum of intermediate inputs/tasks. As a theoretical result, L-globalization increased the wage of middle- and high-skilled labor relative to low-skilled labor, but C-globalization has a quite different effect: while high-skilled labor still gains, the wage of middle-skilled relative to low-skilled labor can fall as the demand for middle-skilled labor is reduced due to service offshoring. This outcome accords well with the evidence on “wage polarization” for the United States, whereby the workers in the middle of the skill range have suffered (Autor, Katz, and Kearney 2008). Notably, Basco and Mesteiri find that the time at which wage polarization due to C-globalization occurs is *delayed* when there is more L-globalization. A wide range of empirical evidence is presented and this paper is highly recommended for further reading.

SUPERMODULAR PRODUCTION

We conclude this chapter by considering some of the newest theory on offshoring. To motivate this, it is useful to start with an unpublished 1996 paper by Kremer and Maskin (1996). They begin their paper with the same observation as many others at that time: wage inequality in the United States was increasing. But they propose a new feature of this rising inequality, which is the segregation of workers across firms. They use the example of shifting from companies like General Motors, which uses both high-skilled and low-skilled workers, to an economy based on Microsoft and McDonalds, which segregates workers of different skills across companies. While the segregation of workers across firms is endogenous, Kremer and Maskin suggest that it can help us to understand increasing wage inequality.

To develop a model with these features, Kremer and Maskin (1996, 4) argue that the production function of a firm should satisfy three conditions, which they describe as:

- (i) workers with different skills are imperfect substitutes for one another;
- (ii) different tasks within a firm are complementary; and
- (iii) different tasks within a firm are differentially sensitive to skill.

The second of these features—complementarity—is also called supermodularity of the production function.

While Kremer and Maskin’s analysis is for a closed economy, it seems to have a clear application to trade, and to offshoring in particular. There have been a number of papers published that address aspects of the problem that Kremer and Maskin propose. For example, one class of models has managers solving problems for employees (Garicano 2000), which leads to complementarities between these two types of workers, or supermodularity. That type of model has been applied to offshoring, with the managers and employees located in different countries.²⁰ Another set of papers analyzes trade between economies with supermodular production and more general distributions of worker skills, but without offshoring.²¹ Kremer and Maskin (2006) themselves have a paper which allows for offshoring in a two-country model, with

²⁰See Antràs, Garicano, and Rossi-Hansberg (2006a,b), Garicano and Rossi-Hansberg (2006), and Caliendo and Rossi-Hansberg (2012). Supporting empirical evidence for French firms is provided by Caliendo, Rossi-Hansberg, and Monte (2015).

²¹See Grossman and Maggi (2000), Grossman (2004), and Manasse and Turrini (2001).

only two types of workers in each country. While this illustrates some of the results from their earlier one-country model, the analysis remains rather special.

The latest paper in this line of research, by Costinot and Vogel (2010), allows for both trade and offshoring, in very general framework with a continuum of goods and continuum of factors. We might think of this framework as a substantial generalization of the HO model, but one that is possible only by using some strong simplifying assumptions. Their results can be summarized as follows. Let us work with a continuum of skill-types, denoted by $s \in [\underline{s}, \bar{s}]$, and a continuum of sectors, denoted by $z \in [0, 1]$. Let $A(s, z)$ denote the output in sector z from using one unit of labor with skill s , which is strictly increasing in s , and let $L(s, z)$ denote the units of labor with skill s that are allocated to sector z . To simplify the problem, Costinot and Vogel make the strong assumption that workers are perfect substitutes in production, so that output in sector z is

$$Y(z) = \int_{\underline{s}}^{\bar{s}} A(s, z) L(s, z) ds.$$

Note that this assumption violates the first condition proposed by Kremer and Maskin, that workers of different skills are imperfect substitutes.

The second condition—of complementarity—is maintained by assuming that the production function $A(s, z)$ is log-supermodular between the skills of workers and their sectors. That is, choosing two skills with $s > s'$, and two sectors with $z > z'$, we assume that

$$\ln A(s', z') + \ln A(s, z) > \ln A(s, z') + \ln A(s', z).$$

Rewriting this condition slightly we get

$$\frac{A(s, z)}{A(s', z)} > \frac{A(s, z')}{A(s', z')}, \text{ for } s > s', \text{ and } z > z'.$$

Thus, workers with higher skills are relatively more productive in sector with higher index z , so we have arranged sectors in increasing order of their skill-intensity. The exogenous distribution of workers is denoted by $V(s) > 0$, so we close the model by using full employment of factors

$$\int_0^1 L(s, z) dz = V(s), \text{ for all } s,$$

as well as demand equal to supply in each sector.

When workers are perfect substitutes, the firm can easily be indifferent as to whom it hires. But that ambiguity is resolved in general equilibrium because Costinot and Vogel show that there is a monotonically increasing matching function $M(s)$, mapping the interval of skills $[\underline{s}, \bar{s}]$ onto the interval of sectors $[0, 1]$. Each sector z employs only one skill-type of worker s , with $z = M(s)$. Such a matching function is illustrated in figure 4.14, which maps the skills of workers to their sectors.

That result that each sector employs only one type of worker, due to the assumption of perfect substitution, gives us an easy way to determine the wage profile $w(s)$. Denoting the equilibrium prices across sectors by $p(z) = p[M(s)]$, profit maximization

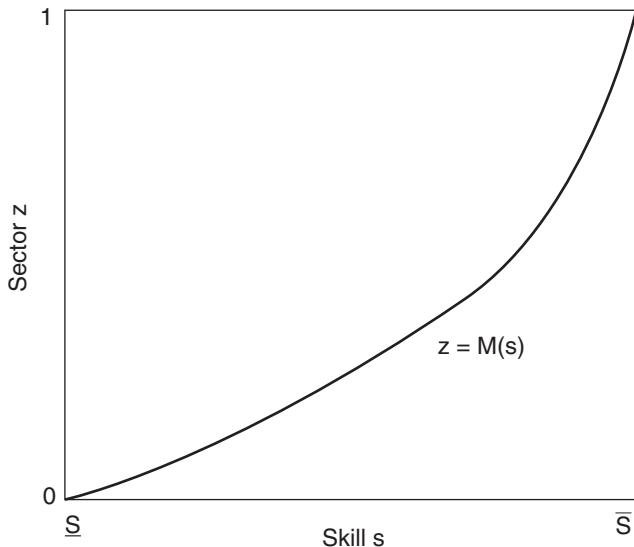


Figure 4.14

by competitive firms means that they will choose the skill level of workers to solve the problem

$$p[M(s)]A[s, M(s)] - w(s) = \max_{s'} \{p[M(s)]A[s', M(s)] - w(s')\}.$$

Differentiating this expression with respect to s and using the envelope theorem, we find that the slope of the wage schedule reflects the slope of the technology:

$$w_s(s) = p[M(s)]A_s[s, M(s)].$$

Furthermore, since wages in equilibrium equal the value of the marginal product of labor, $w(s) = p[M(s)]A[s, M(s)]$, we obtain a differential equation to determine the wage profile:

$$\frac{w_s(s)}{w(s)} = \frac{A_s[s, M(s)]}{A[s, M(s)]}.$$

Of course, solving this differential equation requires knowledge of the matching function, which can be obtained from the equality of demand and supply in each sector. Costinot and Vogel now make a second simplifying assumption, which is that demand is obtained from Leontief preferences, so demand in each sector is proportional to the exogenous taste parameter $B(z) > 0$. To equate demand and supply, let us first compute the integral of output over workers between the skill levels s_1 and s_2 . That output is

$$\int_{s_1}^{s_2} A[s, M(s)]L[s, M(s)]ds.$$

To focus on sectors rather than skills, let us make a change of variables within this integral, from s to z . The range of integration is from $z_1 = M(s_1)$ to $z_2 = M(s_2)$, and making the change of variables with $dz = M_s(s)ds$, the integral becomes

$$\int_{z_1}^{z_2} A[M^{-1}(z), z] L[M^{-1}(z), z] M_s[M^{-1}(z)] dz.$$

This expression gives us the value of output over a range of sectors. Output in a single sector, say z_2 , can be obtained by differentiating with respect to z_2 , to obtain

$$Y(z) = A[M^{-1}(z), z] L[M^{-1}(z), z] M_s[M^{-1}(z)], \text{ for } z = z_2.$$

This expression for output is rewritten for a given skill level as

$$Y[M(s)] = A[s, M(s)] L(s, z) M_s(s).$$

So while we would normally think of output as just $A(s, z) L(s, z)$ in a model with discrete goods, we now obtain the expression $A(s, z) L(s, z) M_s(s)$, which incorporates the rate at which workers are matched to sectors.

From the assumption that preferences are Leontief with the taste parameters $B(s)$, and using $L(s, z) = V(s)$, the equality of demand and supply means that

$$A[s, M(s)] V(s) M_s(s) = kB(s),$$

where $k > 0$ is a constant. This gives us a differential equation that we can use to solve for the matching function, which has the implicit solution

$$M(s) = \int_{\underline{s}}^{\bar{s}} \frac{kB(s)}{A[s, M(s)]V(s)} ds.$$

The constant k can be determined by the condition that the matching function maps the interval of skills $[\underline{s}, \bar{s}]$ onto the interval of sectors $[0, 1]$, so that $M(\bar{s}) = 1$, which implies

$$k = \left[\int_{\underline{s}}^{\bar{s}} \frac{B(s)}{A[s, M(s)]V(s)} ds \right]^{-1}.$$

This completes the solution of the model.

Now consider the effect of changing the distribution of skills, from $V(s)$ to $V'(s)$. First, suppose that the distribution shifts to the left, in the sense that

$$\frac{V(s)}{V(s')} > \frac{V'(s)}{V'(s')}, \text{ for } s > s'.$$

This condition means that there are relatively more workers with lower skills in the distribution $V'(s)$. For concreteness, we can think of the original distribution as applying to the North and the new distribution to the South. Heuristically, we can see how this shift will impact the matching function from examining the differential equation

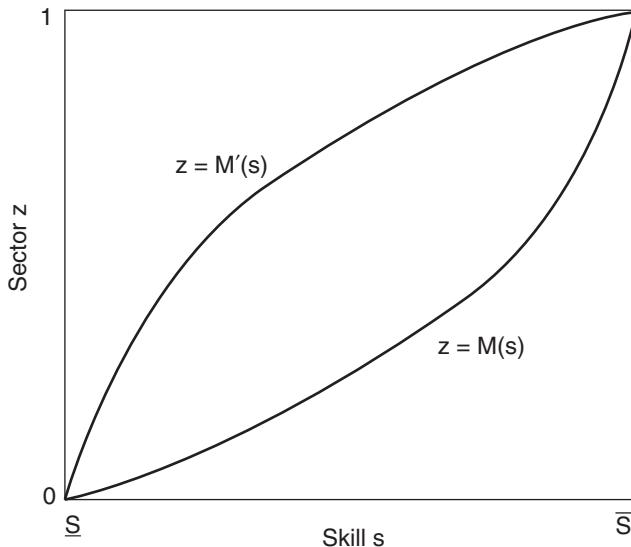


Figure 4.15

$$M(s) = \int_{\underline{s}}^{\bar{s}} \frac{kB(s)}{A[s, M(s)]V(s)} ds.$$

↑ ↓

The arrows in the above expression suggest that a fall in V should increase the matching function M , so that $M'(s) > M(s)$. This intuition is very rough, because M also enters into the right-hand side of the equation (both directly in A and through the term k). But it proves to be correct, and Costinot and Vogel show that the shift in the skill-distribution to the left leads to a rise in the matching function for all skills $\underline{s} < s < \bar{s}$, except at the end points.

This rise is illustrated in figure 4.15, with the old matching function M applying to the North and the new function M' to the South. In the South, workers of lower skill are matched to sectors with higher z , as compared to the North. From the differential equation for wages, we use the rise in the matching function to infer that the slope of the wage schedule must increase:

$$\frac{\partial \ln w(s)}{\partial s} = \frac{\partial \ln A[s, M(s)]}{\partial s}.$$

↑ ↑

The rise in the matching function, shown by the arrow on the right, suggests a steeper slope for the wage profile, as shown by the arrow on the left and as follows from the assumed log-supermodularity of the technology.²² The Southern economy, with a greater number of less-skilled workers, will therefore have a more *unequal wage distribution* than in the North.

²²Supermodularity of $\ln A(s, z)$ means that $\partial^2 \ln A / \partial s \partial z > 0$, so that $\partial \ln A / \partial s$ rises as z rises.

This type of exercise is not restricted to shifts in the mean of the distribution of skill, and Costinot and Vogel also examine the impact of changes in the spread of the skill distribution, or the diversity of skills, as might apply between two Northern countries. But rather than describing those results in detail, let us consider the impact of international trade. A quick type of Heckscher-Ohlin result is available from the case just examined. Suppose that the North and South engage in trade, with factor price equalization. Then the integrated world economy will have a distribution of skills that lies *in between* the distributions in the two countries. From the Northern point of view, wages are now determined by a distribution of skills that is shifted to the left, so that wages become more unequal. From the Southern point of view, however, the world distribution of skills lies to the right of its own distribution, so that wages become more equal. Therefore, in a standard Heckscher-Ohlin fashion, opening trade leads to changes in the wage distributions that are in *opposite directions* in the two countries.

As we have noted earlier in this chapter, relative wages did not move in opposite directions across countries during the 1980s, but in the same direction: toward greater inequality. A result of that type can be generated by considering a simple form of offshoring in this model. Specifically, suppose that the Southern technology is uniformly less productive to that in the North. This is equivalent to saying that all factors of production are uniformly less productive in the South. Then in an equilibrium with free trade, we can still observe effective factor price equalization, defined by efficiency-adjusted wages. To this equilibrium we now add the option of offshoring, which means that Northern firms can use their technology when they employ Southern workers. As a result, Northern technologies will be employed worldwide, which eliminates the productivity disadvantage in the South, and is equivalent to a uniform rise in the Southern endowments. This rise will shift the world distribution of skills to the left, since there are now more effective Southern workers. It follows that the wage distributions in *both* countries now become more unequal, due to offshoring.

While that simple prediction is consistent with the facts, at least for the 1980s, this framework leaves out a number of elements that we would want to incorporate into a model of offshoring: there is no difference in the range of products produced in the North and South, since effective factor price equalization is assumed; likewise, there is no difference in the skills of workers employed at a given industry in the North or South; and also, there is no difference in the skills of workers employed within each industry, due to the perfect substitutes assumption. Many of these assumptions can be weakened, as shown by the work of Monte (2011) and Sampson (2013, 2014), who explicitly introduce firms into the analysis, and as discussed in the review by Costinot and Vogel (2015), to which the reader is referred.

CONCLUSIONS

The first model covered in this chapter, by Feenstra and Hanson (1996, 1997), can be viewed as an extension of the HO model with a continuum of goods, treating the factors as high-skilled and low-skilled labor along with capital. The activities being offshored are those consistent with comparative advantage, that is, those relying on less-skilled labor and taking advantage of lower wages for such workers abroad. This model allows us to predict *within-industry* shifts in demand toward high-skilled labor, just as occurred in the United States and other countries in the 1980s. Because the

model relies on trade in intermediate inputs, and its predictions are consistent with the evidence from U.S. manufacturing, then we conclude that the 1980s can be characterized as a decade of *materials offshoring*.

But the evidence for the 1990s and 2000s shows that it is not just the less-skilled activities that are sent abroad; instead it may be medium- or high-skilled activities. That fact does not fit well with the comparative advantage-based rationale for offshoring, and requires a specification of the extra costs involved with offshoring. This is where the work of Grossman and Rossi-Hansberg (2008) proves so useful. By allowing for a rich structure of offshoring costs, we can predict that either the low-skilled activities are offshore, or that the high-skilled activities are sent abroad, depending on the costs of offshoring. The former case fits the 1980s, and we have shown that the predictions of Grossman and Rossi-Hansberg in that case are similar to Feenstra and Hanson, provided that attention is paid to whether we are assuming a small country or a large country. In contrast, the offshoring of high-skilled activities appears to fit the 1990s, when *services offshoring* has become more important. Grossman and Rossi-Hansberg (2008) predict that it is the lower end of high-skilled activities that will be offshore, that is, those relying on more routine activities and less face-to-face communication. This prediction appears to fit the facts for the 1990s and 2000s, where those workers have faced more job losses.

The work of Costinot and Vogel (2010) allows the second-moments of the skill distribution to influence trade and therefore factor earnings. It is not just the average amount of factors found in an economy—such as high-skilled and low-skilled individuals, or nonproduction and production workers—but also the entire distribution of skills that determines trade. Countries with more diverse endowments of skills can be expected to export goods at the extreme ends of the skill requirements (see also Grossman and Maggi 2000, and Grossman 2004). This type of model therefore gives us a rationale for North-North trade, even when average endowments are the same, in addition to North-South trade. And offshoring can be incorporated into a North-North model, too, as in the work of Grossman and Rossi-Hansberg (2012).

Many of the key articles we have discussed in this chapter are reproduced in the book of readings by Goldberg (2015), and the reader is referred to her introduction. The survey article by Hummels, Munch, and Xiang (2015) is also particularly valuable. As these readings make clear, this chapter still has not covered the most recent research on offshoring. On the theoretical side, Costinot, Vogel, and Wang (2013) model a sequential supply chain in which mistakes potentially occur at each stage in a continuum. There are many countries that differ in their probabilities of making mistakes. In equilibrium there is a matching between stages of production and countries. Baldwin and Venables (2013) call this type of sequential product chain a “snake,” and call the assembly of multiple parts at a central facility a “spider.” They provide a partial equilibrium model that illustrates the difficulties of solving for the location of stages in this framework, and also show that the assignments might be non-monotonically related to transportation costs. Closely related to the supply chain concept is the role of intermediaries, who provide services between buyers and sellers. Examples are the large trading houses such as Li and Fung in Hong Kong. Blum, Claro, and Horstmann (2010) and Antràs and Costinot (2011) have modeled these intermediation activities.

On the empirical side, there has been work on the appropriate measure of offshoring beyond the simple ratio of imported intermediate inputs, proposed by Feenstra

and Hanson (1999). The typical construction of that ratio suffers from the so-called proportionality assumption: for example, to estimate the amount of imported steel used in automobiles, we assume that the automobile industry purchases the same proportion of imported to domestic steel as does the entire economy. Other measures of offshoring beyond this simple ratio have been proposed. Johnson and Noguera (2012) propose the ratio of value added to gross exports (VAX ratio) in an industry as a measure of the intensity of production-sharing across countries. They also show that bilateral trade imbalances measured in value added differ from gross trade imbalances. For example, the U.S.–China imbalance in 2004 is 30–40 percent smaller when measured in value added. That topic is also addressed by Koopman et al. (2014), who criticize the VAX ratio proposed by Johnson and Noguera. In addition, a measure of the fragmentation of the production process has been developed by Antràs, Chor, Fally, and Hillberry (2012), and applied theoretically by Antràs and Chor (2013) and by Fally and Hillberry (2014).

Of course, offshoring can be expected to influence unemployment in addition to wages. Current research on unemployment in trade models depends on either “fair wages,” that are above the market clearing level, or on search frictions. Examples of papers that make use of “fair wages” to discuss the impact of offshoring include Amiti and Davis (2012) and Grossman and Helpman (2008).²³ An early model combining search frictions and offshoring, by Mitra and Ranjan (2010), finds that unemployment is actually reduced due to offshoring because the cost-savings for firms leads them to expand output and employment. Groizard, Ranjan, and Rodríguez-López (2014) solve for the impact of offshoring on the *extensive margin* of employment (due to the birth and death of firms) and its *intensive margin* (due to expansion of production), and identify conditions under which employment can rise.²⁴ That theoretical prediction is supported by the empirical results of Wright (2014) is discussed earlier in the chapter. More general treatments of trade and unemployment are provided in Helpman and Itskhoki (2010) and Helpman, Itskhoki, et al. (2008, 2010, 2012), who find that openness to trade may increase unemployment, but that the gains from trade are still positive. These models also allow us to explore wage inequality within sectors and among groups of workers with the same *ex ante* characteristics.

The issue of the gains from offshoring is the final, important topic that we have not addressed sufficiently. The question is not whether a country gains from trade as compared to autarky: that outcome is guaranteed in most models, as discussed in the following chapters. Rather, the question is whether the *extra* trade resulting from offshoring leads to gains for the country doing the offshoring. In our discussion of the Ricardian model with a continuum of goods in chapter 3, we argued that technological progress abroad could either help or harm the home country, depending on whether that technological progress occurs in the goods that the home country is importing (implying a terms of trade gain) or exporting (implying a terms of trade loss). Samuelson (2001, 2004) uses that observation to argue that offshoring does not necessarily

²³See also Davis and Harrigan (2011) and Kreickemeier and Nelson (2006).

²⁴Rodríguez-López (2014) introduces offshoring into the Melitz (2003) model of monopolistic competition with heterogeneous firms, discussed in chapter 6, and confronts a key limitation of that model: it predicts that only the most productive and largest firms export. In fact, while the data show that larger, more productive firms are more likely to be exporters, there are many cases of smaller firms exporting, too. Rodríguez-López assumes a lumpy decision to offshore with random adjustment costs, which allows for some smaller firms to offshore and export.

leads to “extra” gains, due to a terms of trade loss. That argument is easy to demonstrate even at an undergraduate level,²⁵ though it often has been misinterpreted in the popular press to suggest that there are not gains from trade overall (e.g., Mandel 2007).

The corresponding empirical question is whether the increased U.S. trade with China, for example, has led to a terms of trade gain or loss. Wright (2014) used the strong growth in U.S. imports from China following its 2001 entry into the WTO to instrument for the growth in overall offshoring. His results suggest that the impact of offshoring on U.S. unemployment has been rather small. On the other hand, work by Autor, Dorn, and Hanson (2013) and Acemoğlu et al. (2014) estimate that the impact of Chinese imports on U.S. wages and employment since 2001 is actually quite large. But neither of these papers tells us directly about the changes in the U.S. terms of trade. Hsieh and Ossa (2011) calculate the global effects of productivity growth in China, and find effects that are relatively modest, while Amiti (2014) obtains larger impacts. This question of the gains or losses due to offshoring, especially with China, will no doubt continue to be a topic of future research.

PROBLEMS

- 4.1 Near the beginning of this chapter we consider a thought-experiment proposed by Krugman (2000). Suppose that a two-factor, two-sector economy is in autarky, and there is Hicks’ neutral technological progress in good 1. How does this affect the factor prices?

To answer this question, we write the production function for good 1 as $y_1 = A_1 f_1(L_1, H_1)$. Alternatively, we can write the “effective” output of good 1 as $(y_1/A_1) = f_1(L_1, H_1)$. Then we let $a_{ij}(w)$ denote the factor requirements for one effective unit, $(y_1/A_1) = 1$, where $w = w_H/w_L$ denotes the relative wage of high-skilled labor. It follows that the full-employment conditions for this economy are

$$\begin{aligned} a_{1L}(w)(y_1/A_1) + a_{2L}(w)y_2 &= L, \\ a_{1H}(w)(y_1/A_1) + a_{2H}(w)y_2 &= H. \end{aligned} \quad (4.29)$$

Because the prices are endogenous, instead of working with the zero-profit conditions we instead use the condition that supply must equal demand for each good. Krugman assumes Cobb-Douglas preferences, with the share of income α going to good 1. Using the low-skilled wage as the numeraire, income is $L + wH$, and so $\alpha(L + wH)$ is the value of demand for good 1, which must equal the value of output of good 1:

$$a_{1L}(w)(y_1/A_1) + wa_{1H}(w)(y_1/A_1) = \alpha(L + wH). \quad (4.30)$$

- (a) By adding L plus wH from (4.29) and subtracting $\alpha(L + wH)$ from (4.30), show that we obtain $a_{2L}(w)y_2 + wa_{2H}(w)y_2 = (1 - \alpha)(L + wH)$, and interpret this equation.

²⁵See Feenstra and Taylor (2014, chapter 7).

- (b) Now suppose that there is an increase in A_1 , denoted by $\hat{A}_1 > 0$, which indicates Hick's neutral technological progress in good 1. Then show that the equilibrium changes $\hat{y}_1 = \hat{A}_1$ and $\hat{y}_2 = \hat{w}_1 = 0$ are consistent with the equilibrium conditions (4.29)–(4.30).
- (c) The price of good 1 and 2 can be written as equal to their unit-costs:

$$p_1 = [a_{1L}(w) + wa_{2H}(w)]/A_1,$$

$$p_2 = a_{2L}(w) + wa_{2H}(w).$$

Using the result from part (b), how does the increase in A_1 affect the price of each good? Explain why this is consistent with the changes in outputs from part (b).

- 4.2 In this question we will examine the aggregate demand curve for labor that arises in autarky, and under free trade. Make all the usual assumptions of the two-by-two economy.
 - (a) First consider the autarky economy, where there is a single representative consumer in the country, and the prices for the goods are determined by supply = demand. Suppose that the endowment of labor increases. Then verbally trace through the effects on the outputs of the goods, relative prices of the goods, and then factor prices. Graph the relationship between the labor endowment and the real wage.
 - (b) Now consider the same economy but with free trade, and *fixed* world prices. If there is an increase in the labor endowment, what is the effect on the real wage? Graph this relationship.
 - (c) Reconsider your answer to part (b), supposing that the labor endowment grows so large that the country moves outside its cone of diversification. How will further increases in the labor endowment affect the real wage? Add this to your graph.
- 4.3 Consider an economy with two industries and three factors: low-skilled labor (L), high-skilled labor (H), and capital (K).
 - (a) Suppose that industry 1 has a higher cost-share of low-skilled labor and also of capital (think of factory production). Suppose further than the product prices are *constant*, but that the rental on capital *goes up* (perhaps because capital is leaving the country). In this case, can we definitely predict what happens to the *relative wage* of low-skilled labor?
 - (b) Now add a nontraded good, which uses only low-skilled and high-skilled labor. What is the impact of the increase in the rental on capital, and the change in wages, on the price of the nontraded good? Contrast a high-skill-intensive versus low-skill-intensive nontraded good.

EMPIRICAL EXERCISES

- 4.1 Download the NBER productivity dataset at <http://www.nber.org/nberces/nbprod96.htm>, compute the relative wage and relative employment for 1958–96, and reconstruct figures 4.1 and 4.2 up to 1996. Note: Given this data, you need to first compute the wage rates in production and nonproduction sectors using the following formula:

$$\begin{aligned}
 \text{Production worker wage rate} &= \frac{\sum_i \text{production worker wage bill}_i}{\sum_i \text{production workers}_i}, \\
 \text{Non production worker wage rate} &= \frac{\sum_i \text{Non production worker wage bill}_i}{\sum_i \text{Non production workers}_i}, \\
 &= \frac{\sum_i (\text{total pay roll}_i - \text{production worker wage bill}_i)}{\sum_i (\text{total employment}_i - \text{production workers}_i)}, \\
 i &= \text{industry}
 \end{aligned}$$

- 4.2 Store the files for this chapter in the directory: c:\Empirical_Exercise\Chapter_4\. Run the STATA program “Problem_4_2a.do” to reproduce the regressions in table 4.1 (i.e., table I in Feenstra and Hanson 1999). Then run “Problem_4_2b.do” to perform the two-step regression, table IV and table V in Feenstra and Hanson 1999. Note that table V is obtained using the coefficients in the first column of table IV.
- 4.3 Run the program “Problem_4_3.do” to reproduce some of the regressions in table III in Feenstra and Hanson (1999). Interpret these regressions. Then answer this question: what weights are used in these regressions, and how are the results affected if these weights are not used?

5

Monopolistic Competition and the Gravity Equation I

In this chapter, we make a significant departure from earlier trade models by allowing for increasing returns to scale. The idea that increasing returns might be a reason for trade between countries was well recognized by Bertil Ohlin (1933) and also Frank Graham (1923), and has been the motivation for policy actions.¹ In Canada, for example, arguments were made in the 1960s to the effect that Canadian firms would benefit from unrestricted access to the U.S. market. It was believed that the Canadian market was too small to allow manufacturing industries to operate at a minimum efficient scale, and that with access to the U.S. market, firms could move down their average costs curves, which is a gain in efficiency.² Indeed, this was a principal reason that Canada entered in a free trade agreement with the United States in 1989: to give its firms free access to the large American market. As promising as this line of argument is, however, it contains a puzzle: as firms in Canada and the United States move down their average cost curves due to access to the other market, surely not *all firms* can expand output that much, since who would buy it? Thus, an expansion by some firms seems to suggest that others will need to exit the market entirely. So we need a model to sort out these various effects.

The model that is most suited to this purpose is one of *monopolistic competition*: a market with a large number of firms, each producing a unique variety of a differentiated product, with freedom of entry and exit. This model dates back to Edward Chamberlin (1936) and Joan Robinson (1933), who presented graphical analyses. The widespread use of this model had to wait for a mathematical formulation, however, which was achieved by Lancaster (1975, 1979), Spence (1976), and Dixit and Stiglitz (1977). Lancaster presented a model in which consumers differed in their “ideal variety” of a differentiated good. In contrast, Spence and Dixit and Stiglitz had a single representative consumer demanding many varieties of the differentiated good, in what is called the “love of variety” approach. Lancaster (1980) and Helpman (1981) applied the “ideal variety” approach to

¹Graham (1923) argued that trade due to increasing returns to scale might be an argument for *protection*, that is, that one country could lose from trade. Ethier (1982a) analyzed a model of this sort, relying on increasing returns that are *external* to the firm. In this chapter we will focus exclusively on increasing returns that are *internal* to the firm.

²See, for example, Eastman and Stykolt (1967).

international trade, while Krugman (1979, 1980, 1981) applied the “love of variety” approach due to Dixit and Stiglitz. The comprehensive treatment by Helpman and Krugman (1985) shows that these two approaches lead to very similar results, so we will use the simpler “love of variety” approach. We begin by describing the model of Krugman (1979).

MONOPOLISTIC COMPETITION MODEL

We will suppose that there are $i = 1, \dots, N$ product varieties, where the number N will be endogenously determined. There is a fixed number L of consumers, each of whom receives the following utility from consuming c_i of each variety:

$$U = \sum_{i=1}^N v(c_i), \quad v' > 0, \quad v'' < 0. \quad (5.1)$$

Note that this utility function is *symmetric* over the product varieties, that is, the same function $v(c_i)$ applies to the consumption of each. Each consumer receives labor income of w , and so the consumer’s budget constraint is $w = \sum_{i=1}^N p_i c_i$. They choose consumption c_i of each variety to maximize utility in (5.1), subject to this budget constraint. The first-order conditions for this problem are

$$v'(c_i) = \lambda p_i, \quad i = 1, \dots, N, \quad (5.2)$$

where λ is the Lagrange multiplier (i.e., the marginal utility of income).

The effect of a change in price on consumption can be derived by totally differentiating the system of equations in (5.2), together with the budget constraint. Normally, such a change in price would affect λ . However, it can be argued that if the number of varieties is sufficiently large, so that the budget share of each of them is small, then we can ignore the impact of a change in one price on λ . In that case, the effect of a change in price is simply

$$v''dc_i = dp_i \lambda \Rightarrow \frac{dc_i}{dp_i} = \frac{\lambda}{v''} < 0. \quad (5.3)$$

Using (5.2) and (5.3), we define the elasticity of demand for variety i as

$$\eta_i = -\frac{dc_i}{dp_i} \frac{p_i}{c_i} = -\left(\frac{v'}{c_i v''}\right) > 0. \quad (5.4)$$

While the assumptions we have made on v (increasing and concave) ensure that the elasticity is positive, we do not in general know whether it is increasing or decreasing in c_i . This will turn out to be quite important, and we shall *assume* that $d\eta_i/dc_i < 0$, so that as we move up a demand curve (consumption falling) the elasticity rises. The reader should verify that this assumption holds for a linear demand curve, and more generally, for any demand curve that is “less convex” than a constant-elasticity curve.

On the production side, labor is the only resource and each firm requires the following labor to produce output of y_i :

$$L_i = \alpha + \beta y_i, \quad (5.5)$$

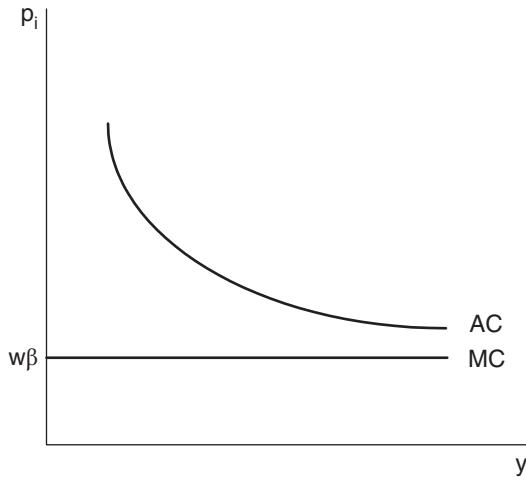


Figure 5.1

where α is the fixed labor input needed for production and β is the marginal labor input. Given the equilibrium wage w , it follows that average costs for the firms are $AC_i = wL_i/y_i = w\alpha/y_i + w\beta$, while marginal costs are simply $w\beta$. These are both graphed in figure 5.1.

Monopolistic competition has two key equilibrium conditions for firms. First, each firm maximizes its own profits, requiring that marginal revenue equals marginal cost ($MR = MC$). Second, there is free entry whenever economic profits are positive, so in the long-run equilibrium we must have zero profits, or price equal to average cost ($P = AC$). In addition to these, we will add the requirement that the equilibrium is “symmetric,” meaning that prices and quantities are identical across varieties. Dropping the subscript i , we therefore have the equilibrium conditions

$$MR = MC: p\left(1 - \frac{1}{\eta}\right) = w\beta, \text{ or } \frac{p}{w} = \beta\left(\frac{\eta}{\eta-1}\right); \quad (5.6)$$

$$P = AC: p = \left(\frac{w\alpha}{y}\right) + w\beta, \text{ or } \frac{p}{w} = \left(\frac{\alpha}{Lc}\right) + \beta. \quad (5.7)$$

Notice that in (5.7), we have replaced supply of each good, y , by the demand for each good, Lc . Equations (5.6) and (5.7) form two equations to solve for the two unknowns, (p/w) and c . The first of these equations is graphed as the line PP in figure 5.2. Our assumption that $d\eta_i/dc_i < 0$ ensures that the PP curve is *upward sloping*. Turning to equation (5.7), this is graphed as the *downward sloping* line ZZ in figure 5.2, which is simply the firm’s average cost curve. The intersection of these two curves determines the equilibrium values of (p/w) and c . Then to determine the equilibrium number of product N , we make use of full employment in the economy, which is stated as

$$L = \sum_{i=1}^N L_i = \sum_{i=1}^N (\alpha + \beta y_i) = N(\alpha + \beta y) = N(\alpha + \beta Lc), \quad (5.8)$$

from which it follows that

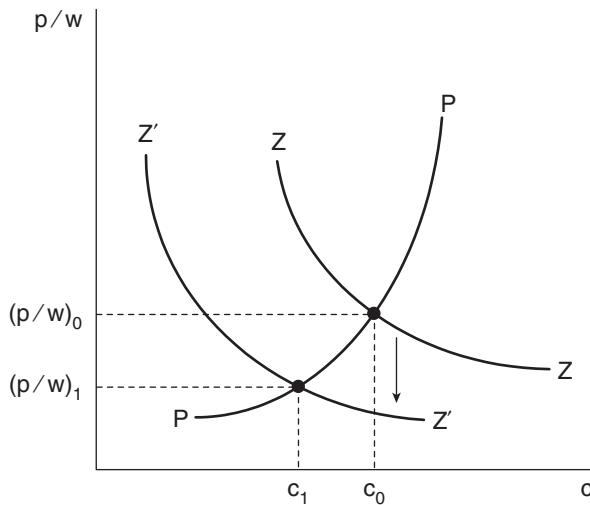


Figure 5.2

$$N = \frac{1}{[(\alpha/L) + \beta c]}. \quad (5.9)$$

The equilibrium value of c therefore determines the number of products N .

So far, we have represented the equilibrium for a single economy. But the effects of trade are easily introduced. For example, suppose that two countries of identical size move from autarky to free trade. Notice that because the economies are identical, then in the Heckscher-Ohlin model there would be absolutely no reason for trade between them. But in the monopolistic competition model, there will be a rationale for trade: because firms produce differentiated products, they will begin to export to the other country, and at the same time, face competition from firms abroad. This increase in the number of competitors can be expected to lower the equilibrium price. This is exactly what happens, as can be confirmed from the equilibrium conditions, as follows.

Having the two identical countries trading is just like doubling the population L . This has no impact on the PP curve since L does not appear in (5.6), but it does shift the ZZ curve *down* as can be seen from (5.7). Therefore, the equilibrium consumption of each variety *falls*, from c_0 to c_1 , while the real wage *rises*, from $(w/p)_0$ to $(w/p)_1$. Consumption falls because individuals are spreading their expenditures over more product varieties, and this raises the elasticity of demand, reducing the equilibrium prices and therefore raising real wages. This is a source of gain for consumers. But there is also a second source of gain from trade, because with the rise in L and fall in c , it can be verified from (5.9) that total product variety *increases*. That is, the sum of varieties from both countries under free trade exceeds the number in any single country before trade.

But notice that the *number of varieties produced in each country necessarily falls* with the opening of trade. This can be seen by noting that the fall in (p/w) , as firms move down their average cost curve, necessarily implies that output y increases. This means that any firm that produces both in autarky and under free trade will be selling more with trade. But the full-employment conditions for *each* economy is $L = N(\alpha + \beta y)$, where now L is fixed. So the increase in y , as firms exploit economies

of scale, necessarily implies a *reduction in the number of firms in each country*. This resolves the puzzle raised at the beginning of the chapter: opening trade between countries indeed implies that firms must exit in each, while the remaining firms expand their output and take advantage of scale economies.

EVIDENCE FROM THE CANADA-U.S. FREE TRADE AGREEMENT

To summarize, Krugman's model contains two predictions concerning the impact of trade on the productivity of firms: the *scale effect*, as surviving firms expand their outputs, and the *selection effect*, as some firms are forced to exit. Which one of these effects is most important in practice? This is a question of some importance since the dislocation of workers caused by the bankruptcy of firms is bound to cause unemployment. This was certainly true in Canada after it entered into the free trade agreement with the United States on January 1, 1989, because at the same time, Canada also had a tight monetary policy that led to further job loss. The early computable general equilibrium model of Harris (1984) and Cox and Harris (1985, 1986), done *before* the 1989 agreement, presumed substantial scale economies based on engineering estimates for various industries, and predicted very high efficiency gains for Canada. These estimates were influential in Canadian policy circles as the merits of the agreement with the United States were being debated. With more than two decades since the Canada-U.S. FTA began, there are now several studies we can look to that quantify the scale versus selection effect empirically.

The scale effect—that surviving firms will expand their scale of production following trade liberalization—has unfortunately not been borne out empirically. Head and Ries (1999) use plant-level data for a sample of 230 Canadian industries over 1988–94 to examine the impact on plant scale in the six years following the free-trade agreement with the United States. They find that the tariff reductions in the United States increased plant scale in Canada by some 10 percent on average, but this was largely offset by a 8.5 percent *reduction* in plant scale due to the reductions in Canadian tariffs. On net, therefore, the free-trade agreement had only a small impact on scale. This type of evidence is also found from tariff reductions in a wide range of developing countries, where the average scale of firms in import-competing industries either shrank following liberalization, or expanded only slightly.³ The only silver lining to this result is that it appears that the extent of economies of scale is not very great in the first place, so that any reduction in scale does not really cause higher costs. In addition, evidence from developing countries shows that the markup of price over marginal costs is indeed reduced following tariff reductions, and so consumers gain in that respect.⁴

Turning to the selection effect, this is another avenue through which productivity in an industry can change following liberalization. In particular, if the *least efficient* firms are the ones to exit, then *average industry productivity* will rise. Note that this is outside the framework of the original Krugman model, which makes the “symmetry” assumption that all firms are of the same size and efficiency (it follows that the exit of some would not automatically change average productivity). This assumption was

³Tybout and Westbrook (1995) obtain this result for liberalization in Mexico during 1984–90, whereas Tybout, de Melo, and Corbo (1991) find evidence of small increases in scale following liberalization in Chile.

⁴We will discuss the empirical link between tariffs and markups in chapters 8 and 11.

made for analytical convenience, but contradicts the empirical fact that every industry has a very wide range of firms operating within it. Trade theory has now caught up with this empirical fact, and there is now a monopolistic competition model that allows for a range of firms with differing productivities, due to Melitz (2003), that can predict the effects of tariff reductions. In these models, which we examine in the next chapter, the gains from the Canada-U.S. free trade area would come not from scale effects, but rather, from selection effects, whereby the least efficient firms exit after liberalization.

The empirical evidence on this comes from Trefler (2004), which is the first attempt to comprehensively assess the impact of this agreement on productivity within Canada. Trefler (2004, 870) finds strong evidence of productivity gains:

The Canada-U. S. Free Trade Agreement provides a unique window onto the effects of a reciprocal trade agreement on an industrialized economy (Canada). For industries that experienced the deepest Canadian tariff cuts, the contraction of low-productivity plants reduced employment by 12 percent while raising industry-level labor productivity by 15 percent. For industries that experienced the largest U. S. tariff cuts, plant-level labor productivity soared by 14 percent. These results highlight the conflict between those who bore the short-run adjustment costs (displaced workers and struggling plants) and those who are garnering the long-run gains (consumers and efficient plants).

But surprisingly, these gains do not come from adjustment of plant scale (887):

The results indicate that the Canadian tariff concessions led the most impacted, import-competing plants to contract by 5 percent ($t = 1.36$) while the U.S. tariff concessions led the most impacted, export oriented plants to expand by 6 percent ($t = 2.01$). These are not statistically significant results. Thus, this is not strong evidence in support of a simple scale-effects explanation of labor productivity gains

Thus, productivity in Canada was increased due to free trade with the United States from the selection effect, not the scale effect. Similar evidence is found for other countries, too. Pavcnik (2002b), for example, considers Chile's trade liberalization of the late 1970s and early 1980s. She finds that Chile's manufacturing productivity increased by 19 percent due to this trade liberalization. Fully two-thirds of these 19 percent were due to a reallocation of resources from low-productivity to high-productivity firms, whereas only one-third was due to firm-level productivity increases.

The finding that the scale of firms does not change that much following trade liberalization can be explained from several models, as discussed in Head and Ries (1999, 2001). We note here one particularly simple theoretical result: if the elasticity of demand for product varieties is *constant*, then firm scale will not change at all due to tariffs or trade liberalization. A utility function that leads to this case is the constant elasticity of substitution (CES) function, which was used by Krugman (1980):

$$U = \sum_{i=1}^N c_i^{(\sigma-1)/\sigma}, \quad (5.10)$$

The elasticity of substitution between products is equal to $\sigma > 1$, which also equals the elasticity of demand η when N is large.⁵ This implies that the markup of prices over

⁵In problem 5.2 you are asked to derive the elasticity of demand and investigate other properties of the utility function (5.10).

marginal costs is fixed, $p/(\beta w) = [\sigma/(\sigma - 1)]$. Substituting this optimal price into the expression for profits of the firm we obtain

$$\pi = py - w(\alpha + \beta y) = w \left[\left(\frac{\beta y}{\sigma - 1} \right) - \alpha \right]. \quad (5.11)$$

In order to have zero profits in equilibrium, the output of firms is therefore *fixed* at

$$\bar{y} = (\sigma - 1)\alpha/\beta. \quad (5.12)$$

So there is no scale effect in a model using (5.10), though trade will still affect the product variety available to consumers. The product variety produced in *each country* is readily solved from (5.12) and the full-employment condition (5.8), obtaining $N = L/(\alpha + \beta \bar{y})$. Thus, the number of varieties *produced* does not change due to trade in this CES model, so there is no selection effect either, but the number of varieties *consumed* will increase due to those available from imports.⁶

The key advantage of using the CES functional form is that it is *homothetic*, whereas this will not hold for any other choice of $v(c_i)$ in (5.1). For this reason the CES functional form is commonly used in the monopolistic competition literature despite its special properties, and we shall also make use of it through the rest of this chapter and in the next chapter. We note, however, that there has been a good deal of recent research on the more general additively separable utility function used by Krugman (1979) and on other functional forms that allow for changing elasticities and markups.⁷

GAINS FROM IMPORT VARIETY

Despite the fact that the CES model of Krugman (1980) has only a single source of gains from trade, that is, the expansion of product varieties, it is still of substantial interest to ask whether we can measure these gains in practice. From a technical point of view, measuring the benefits of new import varieties is equivalent to the so-called “new goods” problem: because the price for a product is not observed before it is available, we don’t know what *price drop* to use to evaluate the consumer gain from having the product available. The answer given many years ago by Hicks (1940) was that the relevant price of a product before it is available is the “reservation price” for consumers, namely, a price so high that demand is zero. Once the product appears on the market it has a lower price, determined by supply and demand. The fall in the price

⁶If firms are heterogeneous, as in the model of Melitz (2003), then a selection effect will reappear even using CES preferences as discussed in the next chapter.

⁷Dhingra and Morrow (2013), Kichko, Kokovin, and Zhelobodko (2013), and Zhelobodko et al. (2011, 2012) consider a broader class of additively separable functions than does Krugman (1979) by allowing the elasticity of demand to be increasing or decreasing in quantity or price. Other examples of the additively separable function include Behrens and Murata (2007, 2012), who use an exponential function, and Saure (2009) and Simonovska (2015), who use a logarithmic function with displaced origin. Outside of the additively separable class, the most common alternative is the quadratic utility function used by Melitz and Ottaviano (2008) and Neary (2003). These examples are all nonhomothetic utility functions. In problem 5.5, we provide an example of the translog expenditure function, used by Feenstra (2003), Bergin and Feenstra (2000, 2001, 2009), and Feenstra and Weinstein (2010), which corresponds to a utility function that is homothetic but does not require constant price elasticities of demand.

from its “reservation” level to the actual price can be used in an index number formula to obtain the consumer gains from the appearance of that new good.

For the constant elasticity of substitution (CES) utility function, we immediately run into a problem with implementing this suggestion because the reservation price for any good is infinite: the demand curve approaches the vertical axis as the price approaches infinity. But provided that the elasticity of substitution is greater than unity, then we can show that welfare gains from having the new varieties available are finite and, in fact, are not that difficult to measure. To show these results, we will work with the nonsymmetric CES function

$$U_t = U(c_t, I_t) = \left[\sum_{i \in I_t} a_{it} c_{it}^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \sigma > 1, \quad (5.13)$$

where $a_{it} > 0$ are taste parameters that can change over time (reflecting the quality of a product, for example), and I_t denotes the set of goods or varieties that are available in period t at the prices p_{it} . We will treat this set of goods as expanding due to new import varieties that are available with international trade. We can see that (5.13) is a generalization of (5.10), where we now allow for taste parameters that can differ across goods or varieties, so that the demand for each good need not be the same. The rationale for having these taste parameters possibly change over time will be explained later in the chapter.⁸

For the CES utility function in (5.13), the minimum expenditure to obtain one unit of utility is

$$e(p_t, I_t) = \left[\sum_{i \in I_t} b_{it} p_{it}^{1-\sigma} \right]^{1/(1-\sigma)}, \sigma > 1, b_{it} \equiv a_{it}^\sigma. \quad (5.14)$$

We can think of the expenditure function $e(p_t, I_t)$ as measuring the cost of living, and we shall also refer to it as the “theoretical” price index for the CES function. With labor as the only factor of production, the real wage (or utility) earned by consumers is $w_t/e(p_t, I_t)$.⁹ It follows that the rise in utility between two time periods equals the ratio $[w_t/e(p_t, I_t)]/[w_{t-1}/e(p_{t-1}, I_{t-1})]$. We can choose labor as the numeraire in the economy so that the wage equals unity, and then we see that the rise in utility is $[e(p_{t-1}, I_{t-1})/e(p_t, I_t)]$, or the inverse of the change in the cost of living, $[e(p_t, I_t)/e(p_{t-1}, I_{t-1})]$. Thus, to measure the change in utility we need to obtain the ratio of the expenditure functions over time.

For simplicity, first consider the case where $I_{t-1} = I_t = I$, so there is no change in the set of goods, and also $b_{it-1} = b_{it}$, so there is no change in tastes. We assume that the observed purchases c_{it} are optimal for the prices and utility, that is, $c_{it} = U_t(\partial e / \partial p_{it})$. Then the index number P_{SV} due to Sato (1976) and Vartia (1976b) shows us how to measure the ratio of expenditures.¹⁰

⁸See note 22 of this chapter.

⁹See problems 5.2 and 5.3 to derive (5.14) when there are two goods.

¹⁰We refer to P_{SV} as a “price index” because it can be measured with observed data on prices and quantities, as in (5.15)–(5.16). In contrast, the expenditure function in (5.14) depends on unknown parameters like b_{it} , so we refer to it as a “theoretical” price index. Sometimes we will not be careful in making this distinction, however.

THEOREM (SATO 1976; VARTIA 1976B)

If the set of goods available is fixed at $I_{t-1} = I_t = I$, taste parameters are constant, $b_{it-1} = b_{it}$, and observed quantities are optimal, then

$$\frac{e(p_t, I)}{e(p_{t-1}, I)} = P_{SV}(p_{t-1}, p_t, c_{t-1}, c_t, I) \equiv \prod_{i \in I} \left(\frac{p_{it}}{p_{it-1}} \right)^{w_i(I)}, \quad (5.15)$$

where the weights $w_i(I)$ are constructed from the expenditure-shares $s_{it}(I) \equiv p_{it}c_{it}/\sum_{i \in I} p_{it}c_{it}$ as,

$$w_i(I) \equiv \left(\frac{s_{it}(I) - s_{it-1}(I)}{\ln s_{it}(I) - \ln s_{it-1}(I)} \right) / \sum_{i \in I} \left(\frac{s_{it}(I) - s_{it-1}(I)}{\ln s_{it}(I) - \ln s_{it-1}(I)} \right). \quad (5.16)$$

The results in (5.15) show that the ratio of expenditure functions is measured by an “exact” price index P_{SV} due to Sato and Vartia, which is constructed using observed prices and quantities as a weighted geometric mean of the price ratios, using the weights shown in (5.16).¹¹ The important point of this theorem is that goods with high taste parameters b_i will also tend to have high weights, so even without knowing the true values of b_i , the exact ratio of expenditures is obtained. The numerator in (5.16) is the “logarithmic mean” of the shares $s_{it}(I)$ and $s_{it-1}(I)$, and lies in between these two shares, while the denominator ensures that the weights $w_i(I)$ sum to unity. The special formula for these weights in (5.16) is needed to precisely measure the ratio of expenditures in (5.15), but in practice the Sato-Vartia formula will give similar results to using other weights, such as either $s_{it-1}(I)$ or $s_{it}(I)$.¹²

Now consider the case where the set of goods is changing over time, but some of the goods are available in both periods, so that $I_{t-1} \cap I_t \neq \emptyset$. We again let $e(p, I)$ denote the expenditure function defined over the goods within the set I , which is a non-empty subset of those goods available in both periods, $I \subseteq I_{t-1} \cap I_t \neq \emptyset$. We sometimes refer to the set I as the “common” set of goods because they are available in both periods. Then the ratio $e(p_t, I_t)/e(p_{t-1}, I_t)$ is still measured by the Sato-Vartia index in the above theorem. Our interest is in the ratio $e(p_t, I_t)/e(p_{t-1}, I_{t-1})$, which can be measured as in the following result.

THEOREM (FEENSTRA 1994)

Assume that $b_{it-1} = b_{it}$ for $i \in I \subseteq I_{t-1} \cap I_t \neq \emptyset$, and that the observed quantities are optimal. Then for $\sigma > 1$,

$$\frac{e(p_t, I_t)}{e(p_{t-1}, I_{t-1})} = P_{SV}(p_{t-1}, p_t, c_{t-1}, c_t, I) \left(\frac{\lambda_t(I)}{\lambda_{t-1}(I)} \right)^{1/(\sigma-1)}, \quad (5.17)$$

¹¹The concept of “exact” price indexes, which equal the ratio of expenditure functions, is discussed in appendix B, where other examples are given. See also Diewert (1976).

¹²Treating $s_{it-1}(I)$ as a fixed number, it is straightforward to show using L'Hôpital's rule that as $s_{it}(I) \rightarrow s_{it-1}(I)$ then the numerator of (5.16) also approaches $s_{it-1}(I)$. Thus, as $s_{it}(I) \rightarrow s_{it-1}(I)$ for all goods i then the weights in (5.16) also approach $s_{it-1}(I)$ for all i . Provided that the shares do not change too much over time, then the weights in (5.16) will be similar to them.

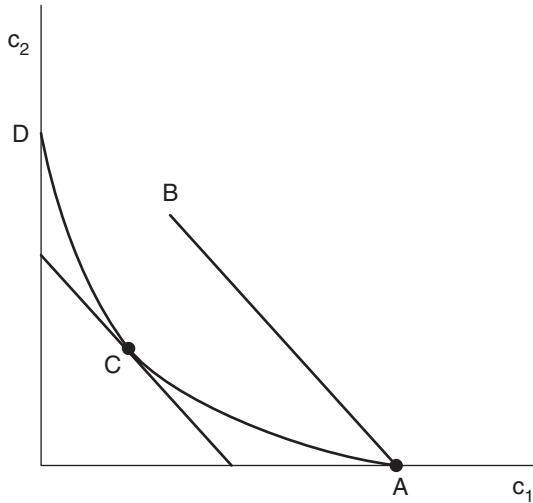


Figure 5.3

where the values $\lambda_t(I)$ and $\lambda_{t-1}(I)$ are constructed as

$$\lambda_\tau(I) = \left(\frac{\sum_{i \in I} p_{it} c_{it}}{\sum_{i \in I_\tau} p_{it} c_{it}} \right) = 1 - \left(\frac{\sum_{i \in I_\tau, i \notin I} p_{it} c_{it}}{\sum_{i \in I_\tau} p_{it} c_{it}} \right), \tau = t-1, t.$$

Each of the terms $\lambda_\tau(I) \leq 1$ can be interpreted as the *period τ expenditure on the goods in the common set I , relative to the period τ total expenditure*. Alternatively, $\lambda_t(I)$ is interpreted as *one minus the period t expenditure on new goods (not in the set I), relative to the period t total expenditure*, while $\lambda_{t-1}(I)$ is interpreted as *one minus the period $t-1$ expenditure on disappearing goods (not in the set I), relative to the period $t-1$ total expenditure*. When there is a greater expenditure share on new goods in period t than on disappearing goods in period $t-1$, then the ratio $\lambda_t(I)/\lambda_{t-1}(I)$ will be less than unity, which leads to a *fall* in the cost of living in (5.17) by an amount that depends on the elasticity of substitution.

The importance of the elasticity of substitution can be seen in figure 5.3, where we suppose that the consumer minimizes the expenditure needed to obtain utility along the indifference curve AD . If initially only good 1 is available, then the consumer chooses point A with the budget line AB . When good 2 becomes available, the same level of utility can be obtained with consumption at point C . Then the drop in the cost of living is measured by the inward movement of the budget line from AB to the line through C , and this shift depends on the convexity of the indifference curve, or the elasticity of substitution.

Gains from Product Variety in Trade

There are several ways that the product variety formula $\lambda_\tau(I)$ can be applied to international trade data. Hummels and Klenow (2002) measure the product variety of *exports* for a broad sample of 110 countries in one year, and evaluate whether differences in

the value of exports across countries are due to differences in product variety or due to differences in sales of goods that are commonly exported by countries. They find that differences in product variety are a principal explanation of cross-country export patterns, a finding confirmed by Schott (2004), who compares rich and poor countries.

An application closer to the spirit of the Krugman (1980) model, however, is to focus on the product variety of *imports*.¹³ We suppose that the utility function in (5.13) applies to the purchases of varieties of a single good from various source countries $i \in I_t$. That is, each country is exporting a different variety of the good in question (cars, shoes, etc.), which was the assumption made by Armington (1969) and Krugman (1980). Then the gains being measured in (5.17) are the gains from greater import variety of that good. These gains have been measured for imports of intermediate inputs by Indian firms by Goldberg, Khandelwal, Pavcnik, and Topalova (2010), where a greater variety of inputs contributes to firm productivity. Conversely, Gopinath and Neiman (2014) have shown the decrease in import variety during times of crises leads to a fall in productivity.

If we repeat the calculation of the gains from import variety over many different goods, and add up the gains across goods, we can obtain an estimate of the overall gains from import variety for a country. This calculation has been made by Broda and Weinstein (2006) for the United States. They define a *good* as a ten-digit Harmonized System (HS) category, or before 1989, as a seven-digit Tariff Schedule of the United States (TSUSA) category. The imports from various source countries are the *varieties* available for each good. The ratio $(\lambda_t / \lambda_{t-1})$ as appears in (5.17) is constructed for each good, using the expenditure on new and disappearing source countries for U.S. imports as compared to a base year $t - 1$. In addition, they estimate σ for each good, using a generalized method of moments (GMM) technique from Feenstra (1994), which exploits heteroskedasticity across countries to identify this elasticity (see the appendix to this chapter).¹⁴ Combining these calculations, they measure $(\lambda_t / \lambda_{t-1})^{-1/(\sigma-1)}$ for some thirty thousand goods available in the HS and TSUSA data. For the TSUSA data they use 1972 as the base year and measure the gains from new supply countries up to 1988, and then for the HS data they use 1990 as the base year, and measure the gains from new supplying countries up to 2001.¹⁵

Two features of Broda and Weinstein's (2006) methods deserve special mention. First, by measuring the expenditure on new supplying countries relative to a base year, they are following the hypothesis of the above theorem that the "common" set of countries should be those with *constant* taste parameters. In contrast, when countries first start exporting goods, it is reasonable to expect that the demand curve in the importing country shifts out over some number of years, as consumers become informed about the product. Broda and Weinstein are allowing for such shifts for new and disappearing countries after the base year, and all such changes in demand for these countries are incorporated into the λ_τ terms in the above theorem. That is the correct way to measure the gains from new import varieties.¹⁶

¹³While we focus on the product variety of imports in this chapter, Feenstra (2010a) argues that the product variety of exports can still be given a welfare interpretation.

¹⁴See Soderbury (2010, 2012) and Feenstra, Luck, Obstfeld, and Russ (2014) for further properties of this estimator.

¹⁵The year 1989 is omitted because West and East Germany unified then, making comparisons with later years difficult.

¹⁶In addition, countries that are suspected of selling a changing range of product varieties *within* each HS good should be excluded from the set I , and instead included in the λ_τ terms.

Second, Broda and Weinstein do not incorporate any changes in the number of U.S. varieties into their calculation of the λ_t terms. This is the correct approach only under the limited case where *the number of U.S. varieties is constant*. While that is true under our assumptions in the model of Krugman (1980), it is certainly not the case in more general models: we could expect that increases in import variety would result in some reduction in domestic varieties. In that case, the gains from import varieties would be offset by the welfare loss from reduced domestic varieties. That potential loss was addressed to a limited extent by Broda and Weinstein, and we shall address it more fully in the next chapter.

Setting aside that concern, Broda and Weinstein (2006) aggregate over goods to obtain an estimate of the gains from trade for the United States due to the expansion of import varieties over 1972–2001. The gains that they obtain are sizable, growing by 1.2 percentage points per year over this period, to a total of 28 percent of import expenditure in 2001. Because imports are a small fraction of U.S. GDP, however, the gains measured relative to GDP are only about 2.6 percent in 2001, as reported in table 5.1. While this number may seem modest in size, remember that it is an estimate of the *incremental* gains from import variety from 1972 to 2001, and not the total gains from trade as compared to autarky.

To obtain the gains from trade as compared to autarky, we can again use formula (5.17) and let period $t - 1$ denote autarky. Assuming that the variety of domestic goods do not change when moving to trade, as in the Krugman (1980) model, then the set I equals all domestic varieties and λ_{t-1} equals unity (since there are no disappearing domestic varieties). All import varieties are new as compared to autarky, so λ_t equals one minus the import share. Then the drop in the cost of living due to imports is measured very simply by $\lambda_t^{1/(\sigma-1)}$, or inverting this, the rise in utility relative to autarky equals $\lambda_t^{-1/(\sigma-1)}$.

Next, let us be explicit about how we aggregate across goods. Following Ossa (2012), we assume that there is a Cobb-Douglas utility function over goods, where (5.13)—now using the subscript n rather than the subscript t —refers to the *subutility* from consuming all the varieties of that differentiated good n . We denote the Cobb-Douglas share of expenditure on each good n by α_n and we also allow the elasticity of substitution σ_n to differ across goods. We just argued that the rise in utility relative to autarky for good n equals $\lambda_n^{-1/(\sigma_n-1)}$, so taking the Cobb-Douglas function of this using the shares α_n we obtain

$$\text{Rise in utility relative to autarky} = \prod_n \lambda_n^{-\alpha_n/(\sigma_n-1)}. \quad (5.18)$$

TABLE 5.1: Estimates of Gains from Trade

Authors	Time Period	Country	Gains (% of GDP)	Interpretation
Broda and Weinstein (2006)	1972–2001	U.S.	2.6	Incremental gain from growth in import variety
Ossa (2012)	2005	U.S.	23.5	Gains as compared to autarky
Ossa (2012)	2005	Japan	7.9	Gains as compared to autarky
Ossa (2012)	2005	Hungary	117.6	Gains as compared to autarky
Ossa (2012)	2005	Average	47.1	Gains as compared to autarky
Ossa (2012)	2005	U.S.	3.8	Gains as compared to autarky, but using average elasticity

This expression is a geometric mean of the welfare gains from each good, $\lambda_n^{-1/(\sigma-1)}$, where λ_n equals one minus the import share of that good, or the *domestic* share. Making this explicit, if we let M_n denote the value of imports and let D_n denote the value of demand from domestic sources, so that $C_n \equiv M_n + D_n$ denotes the value of total consumption, then $\lambda_n = D_n/C_n = 1 - (M_n/C_n)$. So the data needed to measure (5.18) are simply imports and total consumption, which also give us the consumption shares $\alpha_n = C_n/(\sum_n C_n)$, along with the elasticities of substitution σ_n . Ossa (2012) uses the elasticities estimated by Broda and Weinstein (2006) and calculates (5.18) for the United States and forty-nine other countries for the year 2005. As in Broda and Weinstein, a further adjustment is also made to reflect the fact that traded goods are only a fraction of GDP for any country.

The results of Ossa's calculations are shown in table 5.1. For the United States the gains from trade are 23.5 percent of GDP, or nine times bigger than those obtained by Broda and Weinstein (2006). This much higher value occurs because Ossa is computing the gains as compared to autarky, whereas Broda and Weinstein consider only the incremental gains from increased variety.¹⁷ The gains that Ossa finds for the United States are not even that high as compared to those for other countries in his sample. In table 5.1 we report the country with the lowest gains in his calculations (Japan, with 7.9 percent of GDP), the country with the highest gains (Hungary, with 117.6 percent of GDP), and the average over all countries (47.1 percent of GDP), which is still twice as big as what is found for the United States. In summary, these estimates of the gains from trade are quite large indeed.

Ossa points that the formula in (5.18) looks quite different if, instead of using the domestic shares for the individual goods n , we instead compute the domestic share for the overall economy, or $\lambda \equiv (\sum_n D_n)/(\sum_n C_n)$. It turns out that (5.18) is equivalently written as¹⁸

$$\text{Rise in utility relative to autarky} = \lambda^{-\sum_n (\ln \lambda_n / \ln \lambda) \alpha_n / (\sigma_n - 1)}. \quad (5.18')$$

Whenever λ is close to unity then we can use the approximation $\ln \lambda \approx \lambda - 1$. It follows that $(\ln \lambda_n / \ln \lambda) \approx (1 - \lambda_n)/(1 - \lambda)$, which appears in the exponent of (5.18') and is interpreted as the import share for good n relative to the import share of the entire economy. It is not hard to show that $\sum_n \alpha_n (1 - \lambda_n)/(1 - \lambda) = \sum_n (M_n / \sum_i M_i) = 1$.¹⁹ It follows that the exponent of (5.18') is approximately equal to a weighted average of the inverse elasticities minus one, $1/(\sigma_n - 1)$. We can further interpret this expression by noting that the *weighted harmonic mean* of the elasticities minus one is

$$\text{Harmonic mean} = H \equiv \left[\sum_n \frac{\alpha_n (1 - \lambda_n)/(1 - \lambda)}{(\sigma_n - 1)} \right]^{-1}.$$

Therefore, the gains from trade in (5.18') are approximately equal to

$$\text{Rise in utility relative to autarky} \approx \lambda^{-1/H}.$$

¹⁷ As noted, both Broda and Weinstein and Ossa make an adjustment for nontraded goods in the economy before computing the gains relative to GDP. The form of this adjustment differs in the two papers, however, Ossa uses parameters that apply on average across many countries, whereas Broda and Weinstein use the low import share of the United States. This difference may partially explain their low estimate relative to U.S. GDP.

¹⁸ See problem 5.6 to derive (5.18').

¹⁹ See problem 5.6.

Suppose that instead we use the direct weighted average of the elasticities minus one, $\sigma - 1 \equiv \sum_n \alpha_n (1 - \lambda_n) / (1 - \lambda) (\sigma_n - 1) = \sum_n (M_n / \sum_i M_i) (\sigma_n - 1)$. A feature of the harmonic mean H is that it approaches zero whenever there is one good with an elasticity $\sigma_n \rightarrow 1$, whereas that is not true for the arithmetic mean $\sigma - 1$. In other words, whenever there is one good with an elasticity close to unity (and a weight that is not close to zero), then the gains from trade $\lambda^{-1/H}$ will become large. That is not true, however, for the gains $\lambda^{-1/(\sigma-1)}$ computed with the average elasticity. More generally, it is known that *the weighted harmonic mean is less than or equal to the weighted arithmetic mean*, with equality only in the case where the elasticities are the same across goods.²⁰ We have therefore established the following result.

THEOREM (OSSA 2012)

The gains from trade in (5.18) approximately equal those obtained by using the economy-wide domestic share λ and the weighted harmonic mean of the elasticities minus one, $\lambda^{-1/H}$. If the elasticities of substitution differ across goods, then these gains exceed that obtained from $\lambda^{-1/(\sigma-1)}$, using the import-weighted average elasticity $\sigma - 1 \equiv \sum_n (M_n / \sum_i M_i) (\sigma_n - 1)$.

This result is illustrated in the last row of table 5.1, where we report the gains from trade in the United States computed using the arithmetic average of the elasticities of substitution. In contrast to the calculation in the second row we now find that the gains are quite small, only 3.8 percent of GDP. The difference between the results in the second and last rows reflects the very wide range of values obtained for the estimated elasticities across sectors, resulting in the downward bias of the gains when we incorrectly use the average elasticity instead.

THE GRAVITY EQUATION

In the monopolistic competition model, each country will be exporting varieties of the differentiated product to another. While firms in different countries may produce the same product varieties in autarky, we assume they can costlessly leave one variety and produce another, so that with trade it is profit maximizing to produce *different* varieties. In other words, the countries are completely specialized in different product varieties. Trade in these product varieties is referred to as *intra-industry trade*. In this case, it turns out that trade patterns can be described by a remarkably simple equation called the “gravity equation.”

In its simplest form, the gravity equation states that the bilateral trade between two countries is directly proportional to the product of the countries’ GDPs. Thus, larger countries will tend to trade more with each other, and countries that are more similar in their relative sizes will also trade more. The importance of countries’ economic size in explaining trade patterns was first recognized by Tinbergen (1962). Tinbergen proposed that the volume of trade between countries would be similar to the force of

²⁰More generally, the weighted harmonic mean is less than or equal to the weighted geometric mean, which is less than or equal to the weighted arithmetic mean; see Hardy, Littlewood, and Polya (1934).

gravity between objects. Suppose that two objects each have mass M_1 and M_2 , and they are located distance d apart. Then, according to Newton's universal law of gravitation, the force of gravity F between these two objects is

$$F = G \frac{M_1 M_2}{d^2},$$

where G is the gravitational constant, M_1 and M_2 the mass of the two objects, and d is the distance between them. The larger each of the objects is, or the closer that they are to each other, then the greater the force of gravity between them.

The equation proposed by Tinbergen to explain trade between countries is very similar to Newton's law of gravity, except that instead of the mass of two objects we use the gross domestic product (GDP) of two countries, and instead of predicting the force of gravity we are predicting the amount of trade between them. So the gravity equation in trade is

$$X = A \frac{Y_1 Y_2}{d^\rho}, \quad (5.19)$$

where X trade is the amount of trade (that is, imports, exports, or their sum total) between two countries, Y_1 and Y_2 their gross domestic products, d the distance between them, and A is a constant. We use the exponent ρ on distance rather than d^2 as in Newton's law of gravity, because we are not sure of the economic relationship between distance and trade.

While Tinbergen (1962) showed that the gravity equation worked well empirically, it was some years before the theoretical foundation of this equation was known. The earliest contribution is Anderson (1979), followed by Bergstrand (1985, 1989) and Helpman (1987). All these authors suppose that countries produce different goods. That result might be due to the underlying assumption of monopolistic competition, in which case every firm produces a different product variety from every other firm. It follows that firms in different countries also produce different product varieties (which we label as different goods). But that is not the only model where countries produce different goods. That result also applies in a perfectly competitive model where there are more goods than factors and some differences in factor prices or technologies across countries (Bhagwati 1972; Davis 1995). For example, in the Heckscher-Ohlin (HO) model with a *continuum* of goods, we have complete specialization in different products when factor prices are unequal, as described in chapter 3. The common feature of the monopolistic competition model and the HO model with a continuum of goods is that they both have *many more goods than factors*: that is what allows for complete specialization in different product varieties across countries.

Trade with and between Canada and the United States

An early application of the gravity model, and one that has stimulated a large amount of research, came from comparing *intranational* trade between Canadian provinces to *international* trade between Canadian provinces and U.S. states. This was the question posed by McCallum (1995) in a study using 1988 data, just before the Canada-U.S. FTA was signed. He estimated a gravity model like (5.19), where the dependent variable is exports from each Canadian province to other provinces or to U.S. states.

Exports depend on the province or state GDPs, and so the regression estimated by McCallum is

$$\ln X^{ij} = \alpha + \beta_1 \ln Y^i + \beta_2 \ln Y^j + \gamma \delta^{ij} + \rho \ln d^{ij} + \varepsilon_{ij}, \quad (5.20)$$

where X^{ij} denotes the exports from province or state i to province or state j , δ^{ij} is an indicator variable that equals unity for trade between two Canadian provinces and zero otherwise, and d^{ij} is the distance between any two provinces or states. This common form of the gravity equation comes from taking natural logarithms of (5.19), while adding the coefficients β_1 and β_2 to the GDP terms as well as adding the indicator variable δ^{ij} .

The results from estimating (5.20) using 1988 data are shown in column (1) of table 5.2. The same regression using 1993 data is shown in column (2). Note that McCallum's dataset includes trade between Canadian provinces, and provinces and U.S. states, but not trade between the U.S. states. This is added in column (3) using 1993 data, in which case we also include an indicator variable that equals 1 when trade is between two U.S. states and zero otherwise.

The results in columns (1)–(3) show coefficients on provincial or state GDP close to unity, and a strong negative relationship between distance and trade. This is no surprise. What is unexpected is the very large coefficient on cross-provincial trade: ranging from 3.09 in column (1) to 2.75 in column (3). Taking the exponents of these, we obtain the estimates on “Canada trade” shown at the bottom of the table, indicating the cross-provincial trade is some twenty-two times larger than cross-border trade in 1988, and 15.7 times larger in 1993. These numbers are extraordinarily high! They are meant to capture any and all factors that might impede trade between the United States and Canada, or what we might call “border effects.” It seems nearly unbelievable, however, that these factors would lead to sixteen or twenty-two times more internal trade in Canada than external trade.

How is it that the Canada-U.S. border can apparently have an impact this large? The answer from work of Anderson and van Wincoop (2003) is that border effects have an *asymmetric effect* on countries of different size, and in particular, have a larger effect on small countries. To understand their argument, we can run through a numerical example using Canada-U.S. trade. Let us make the realistic assumption that the United States is ten times bigger economically. It follows that

- With frictionless trade, Canada exports 90 percent of GDP to the United States, so it sells 10 percent internally
- Suppose the border effect reduces cross-border trade by a factor of one-half
- Then, Canada exports 45 percent of its GDP to the United States, and so it must trade 55 percent internally
- So internal trade went up by 5.5 times (from 10 to 55 percent), cross-border down by one-half (from 90 to 45 percent), and so internal trade has increased by eleven times more than cross-border trade
- If we ask what has happened in the United States, it used to export 10 percent of its output to Canada, and now it exports only 5 percent. So internal trade has risen from 90 to 95 percent of output—a modest change—while external trade has fallen in half. We conclude that cross-state trade has increased by slightly more than two times cross-border trade.

TABLE 5.2: Gravity Equation for Canada-U.S. Trade

	Dependent Variable: Value of Exports for Province/State Pair				
	McCallum (1995) and other samples			Anderson, van Wincoop (2001)	With Fixed Effects ^a
	(1)	(2)	(3)	(4)	(5)
Regions included:				US-US	US-US
CA-CA	CA-CA	CA-CA	CA-CA	CA-CA	CA-CA
CA-US	CA-US	CA-US	CA-US	CA-US	CA-US
Year of data:	1988	1993	1993	1993	1993
ln Y^i	1.21 (0.03)	1.22 (0.04)	1.13 (0.02)	1	1
ln Y^j	1.06 (0.03)	0.98 (0.03)	0.97 (0.02)	1	1
ln d^{ij}	-1.42 (0.06)	-1.35 (0.07)	-1.11 (0.03)	-0.79 (0.03)	-1.25 (0.04)
Indicator-Canada	3.09 (0.13)	2.80 (0.12)	2.75 (0.11)		
Indicator-U.S.			0.40 (0.05)		
Indicator-Border				-1.65 (0.08)	-1.55 (0.06)
Border effect Canada ^b	22.0 (2.9)	16.4 (2.0)	15.7 (1.9)	10.5 (1.2)	
Border effect US ^b			1.5 (0.1)	2.6 (0.1)	
Border effect Average ^c			4.8 (0.3)	5.2 (0.4)	4.7 (0.3)
R^2	0.81	0.76	0.85	n.a.	0.66
Observations	683	679	1511	1511	1511

Source: McCallum (1995); Anderson and van Wincoop (2003), table 2; Feenstra (2002) and empirical exercises 5.1 and 5.2.

Notes: Standard errors are in parentheses.

a. Including fixed effects for source and destination provinces or states.

b. Computed as the exponent of the Canada or U.S. indicator variable, except for the calculation in column (4), which is explained in the text.

c. Computed as the geometric mean of the Canada and U.S. border effects in columns (3)–(4), and as the exponent of the (absolute value of the) coefficient on the border indicator in columns (4)–(5).

We can see from this example that comparing cross-state trade to cross-border trade for the United States (a large country) gives a reasonable estimate of the true border effect (which was a factor of one-half), but that comparing cross-provincial trade with cross-border trade for Canada (a small country) gives a really exaggerated estimate of the border effect. To avoid this bias, we need to rederive the gravity equation while introducing trade barriers (such as transport costs or tariffs) right from the start. This means that prices differ across countries. Anderson (1979) was the first to derive the gravity equation while taking into account these “price effects.” Estimating the resulting equation still presents a challenge, however, and we shall discuss

three approaches: the use of *price indexes* to measure the price effects in the gravity equation, as in Bergstrand (1985, 1989) and Baier and Bergstrand (2001); the use of *estimated border effects* to measure the price effects, as in Anderson and van Wincoop (2003); and the use of *fixed effects* to account for the price effects, as in Redding and Venables (2004) and Rose and van Wincoop (2001). Our derivation of the gravity equation follows Redding and Venables.

BORDER EFFECTS IN THE GRAVITY MODEL

When there are border effects, such as transport costs or tariffs, then it is no longer the case that prices are equalized across countries, so the pattern of trade is more complex than in the gravity equation (5.20). The only way to sort this out is to assume a specific utility function, and we shall adopt the CES specification. We let c_k^{ij} denote the exports from country i to country j of good k . Because each country produces unique product varieties, the exports of good k from country i to j are identical to the consumption of good k in country j . In contrast to Krugman's model, we now let c_k^{ij} denote total consumption of good k in country j , rather than *per capita* consumption. We will suppose that country $i=1,\dots,C$ produces N^i products. Then utility for country j is

$$U^j = \sum_{i=1}^C \sum_{k=1}^{N^i} (c_k^{ij})^{(\sigma-1)/\sigma}, \quad (5.21)$$

The triple-index on consumption is a bit unwieldy, but this can be simplified by assuming that all products exported by country i sell for the same price p^{ij} in country j . These prices are inclusive of transport costs from country i to j , on a c.i.f. (cost, insurance, freight) basis. In contrast, the local prices p^i for goods produced in country i are net of any transport costs, on a f.o.b. (free on board) basis. Let us model the relationship between these as $p^{ij} = \tau^{ij} p^i$, where $\tau^{ii}=1$ and $\tau^{ij} \geq 1$, indicating that τ^{ij} units of the product must be shipped to country j in order for one unit to arrive; the amount ($\tau^{ij}-1$) "melts" along the way. This formulation is called "iceberg" transport costs and was introduced by Samuelson (1952).

With equal prices p^{ij} across varieties, then consumption in country j is also equal over all the products $k=1,\dots,N^i$ sold by country i , so that $c_k^{ij} = c^j$. Then the utility function is simplified as

$$U^j = \sum_{i=1}^C N^i (c^j)^{(\sigma-1)/\sigma} \quad (5.22)$$

where c^j now denotes the consumption of any product sent from country i to country j .

The representative consumer in country j maximizes (5.22) subject to the budget constraint

$$Y^j = \sum_{i=1}^C N^i p^{ij} c^j, \quad (5.23)$$

where Y^j is aggregate expenditure and income in country j (we will assume balanced trade).

Maximizing (5.22) subject to (5.23), we can derive the following expression for the demand for each product c^{ij} :²¹

$$c^{ij} = \left(\frac{p^{ij}}{P^j} \right)^{-\sigma} \left(\frac{Y^j}{P^j} \right), \quad (5.24)$$

where P^j refers to country j 's overall price index, defined as²²

$$P^j = \left(\sum_{i=1}^C N^i (p^{ij})^{1-\sigma} \right)^{1/(1-\sigma)}. \quad (5.25)$$

To compare this with our earlier derivation of gravity equation in (5.15), let us compute the total value of exports from country i to j . This will be $X^{ij} \equiv N^i p^{ij} c^{ij}$, or from (5.24) and (5.25),

$$X^{ij} = N^i Y^j \left(\frac{p^{ij}}{P^j} \right)^{1-\sigma}. \quad (5.26)$$

ESTIMATING THE GRAVITY EQUATION

Using Price Index data

In order to estimate (5.26), we should recognize that the “true” number of products N^i in each country is unobservable. One approach, then, is to solve for this using the zero-profit conditions. Assuming that labor is the only factor of production and using the same production functions as in Krugman (1979), it follows that firm output is fixed as in (5.12). Then GDP in country i is $Y^i = N^i p^i \bar{y}$, and substituting this into (5.26), we obtain

$$X^{ij} = \frac{Y^i Y^j}{p^i \bar{y}} \left(\frac{p^{ij}}{P^j} \right)^{1-\sigma} = \frac{Y^i Y^j}{p^{io} \bar{y}} \left(\frac{\tau^{ij}}{P^j} \right)^{1-\sigma}, \quad (5.26')$$

where we use $p^{ij} = \tau^{ij} p^i$ to obtain the last expression. The approach of Bergstrand (1985,1989) and Baier and Bergstrand (2001) is to estimate (5.26') directly, where the variables on the right side are country GDPs, τ^{ij} broken into tariffs and transport costs (measured with International Financial Statistics data), and country prices p^i or P^j are measured with GDP deflators.

Specifically, taking the logs and first differences of the variables in (5.26'), we obtain the estimating equation

$$\Delta \ln X^{ij} = \Delta \ln (Y^i Y^j) + (1 - \sigma) \Delta \ln \tau^{ij} - \sigma \Delta \ln p^i + (\sigma - 1) \Delta \ln P^j. \quad (5.27)$$

²¹See problem 5.2 to derive (5.24) and (5.25) when there are two goods.

²²Notice that this theoretical price index is the same as the expenditure function for one unit of utility in (5.14), where the coefficients b_u play the role of N^i in (5.25).

The term $(Y^i Y^j)$ appearing in (5.27) can be further decomposed as $(Y^i + Y^j)^2(s^i s^j)$, where s^i and s^j here denote the shares of each country's GDP relative to their sum, that is, $s^i = Y^i/(Y^i + Y^j)$, and similarly for s^j . Therefore, we can rewrite (5.27) as

$$\Delta \ln X^{ij} = 2\Delta \ln(Y^i + Y^j) + \Delta \ln(s^i s^j) + (1 - \sigma)\Delta \ln \tau^{ij} - \sigma \Delta \ln p^i + (\sigma - 1)\Delta \ln P^j. \quad (5.27')$$

In this formulation, we see that the growth of trade depends on changes in transport costs, changes in the sum of GDP, changes in relative country size measured by $(s^i s^j)$, and changes in the prices of each country, measured with GDP deflators. Because $s^i + s^j = 1$, then by squaring both sides we obtain $s^i s^j = [1 - (s^i)^2 - (s^j)^2]/2$, which measures the "size dispersion" of countries and becomes larger as the countries are *more similar* in size.

Using data for sixteen OECD countries, and taking differences between the averages in 1958–60 and 1986–88, Baier and Bergstrand (2001, 19) estimate the following regression (with standard errors in parentheses):

$$\begin{aligned} \Delta \ln X^{ij} &= 0.05 + 2.37 \Delta \ln(Y^i + Y^j) + 0.60 \Delta \ln(s^i s^j) - 4.49 \Delta \ln(1 + t^{ij}) - 3.19 \Delta \ln(1 + a^{ij}) \\ &\quad (0.56) \quad (0.38) \quad (0.34) \quad (1.00) \quad (0.37) \\ &\quad - 0.68 \Delta \ln Y^j - 0.25 \Delta \ln(p^i/P^j) - 0.08 \ln X_0^{ij}, R^2 = 0.40, N = 240. \\ &\quad (0.24) \quad (0.09) \quad (0.03) \end{aligned} \quad (5.28)$$

The first variables appearing in the right side of (5.28) follow directly from the specification in (5.27'), where the term τ^{ij} has been broken up into tariffs (t^{ij}) and transport costs (a^{ij}). Both of these enter with negative and highly significant coefficients. The log sum of country GDPs has a coefficient that is close to 2, as expected from (5.27'), while the product of shares ($s^i s^j$) has a coefficient that is insignificantly different from unity, so that trade increases as the countries are more similar in size.

The final terms appearing on the ride side of (5.28) are slightly different from those in (5.27'). First, Baier and Bergstrand's modeling of supply is somewhat more general than ours, which leads to the extra term $\ln Y^j$ in (5.28), which has a negative coefficient. Second, the prices that appear separately in (5.27') are combined as a ratio in (5.28), and measured with GDP price deflators.²³ Third, they include the *initial* amount of trade in 1958–60, $\ln X_0^{ij}$, as a regressor in (5.28) to allow for the lagged adjustment of trade flows.

Overall, with an R^2 of 0.40, this gravity equation explains nearly one-half of the changes in bilateral trade flows for the OECD countries. Furthermore, we can use the first variables that appear in the ride side of (5.28) to explain the growth in trade between 1958–60 and 1986–88. Over these three decades, bilateral trade between the sixteen OECD countries used in the sample grew by 150 percent. Of this, 100 percent or *two-thirds of the total*, was explained by the growth in GDP (i.e., the first variable on the right of (5.28) after the constant, times its coefficient). Next, the actual reductions in tariffs, times its coefficient in (5.28), explains another 38 percent increase in trade, or roughly *one-quarter of the total*. Third, the actual reduction in transport costs, times its coefficient in (5.28), explains a further 12 percent increase in trade, or *one-twelfth of*

²³Actually, all of the variables in (5.27') are measured by Baier and Bergstrand (2001) in real rather than the nominal terms as we have used. When that modification is made, the prices p^i and P^j should indeed appear as a ratio.

the total. So we conclude that the reduction in tariffs is about three times as important as the reduction in transport costs in explaining increased OECD trade.

Notice that these three variables (sum of GDPs, tariffs, and transport costs) explain the *entire increase* in trade, leaving no role at all for the convergence in relative country size, which is the term $\Delta \ln(s^i s^j)$ in (5.28). While this term is positive on average across all OECD countries for 1958–60 to 1986–88, it is actually negative but very small (−3 percent) for the sixteen countries used in the sample. In other words, there is a slight *divergence* in size of real GDP among the sixteen OECD countries used, which explains a *slight reduction* in trade (of 2 percent) rather than an increase in trade. Note that this occurs despite the fact that the coefficient on the GDP shares $\Delta \ln(s^i s^j)$ is insignificantly different from unity in (5.28). So while this size dispersion variable passed the statistical test of having a statistically significant coefficient, it fails to be economically important in explaining the growth in trade among the OECD countries.

A similar method to decompose the growth of trade in a historical period is undertaken by Estevadeordal, Frantz, and Taylor (2003). Finally, we note that Baier and Bergstrand (2009) have developed another method to estimate the gravity equation while taking into account the multilateral resistance terms, and that is by taking a linear approximation of those terms.

Using Estimated Border Effects

An objection to using published price indexes to measure p^i and P^j is that these indexes may not accurately reflect the “true” border effects. That is, the myriad of costs (in money, time, and currency risk) involved in making transactions across the border are probably not reflected in aggregate price indexes.²⁴ So instead of using data to measure prices, we might want to model the c.i.f. prices p^{ij} as differing from the f.o.b. prices p^i due to distance and other factors, as with

$$\ln \tau^{ij} = \delta^{ij} + \rho \ln d^{ij} + \varepsilon_{ij} \quad (5.29)$$

where d^{ij} is the distance between country i and country j , δ^{ij} is any other border effect (possibly an indicator variable) associated with selling from country i to country j , and ε_{ij} is a random error. We should think of ρ and δ^{ij} as being estimated, though we will have to explain how this is done. Generally, τ^{ij} is meant to include *all effects* limiting trade between countries i and j , so referring to it as “iceberg” transportation costs is something of a misuse of language.

Substituting (5.29) back into the gravity equation (5.26) to eliminate the terms $p^{ij} = \tau^{ij} p^i$, we obtain a set of equations where the exports between country i and country j depend on the parameters δ^{ij} and ρ . This set of equations is highly nonlinear, however, so it is difficult to estimate these parameters directly. Instead, the approach of Anderson and van Wincoop (2003) is to further simplify the gravity equation making use of the market clearing conditions, as follows.

Notice that with the iceberg transportation costs, the output of the firm exceeds the net amount received by consumer, since these are related by

²⁴Another problem with using price indexes is that they are nearly always measured relative to an arbitrary base period. This makes it impossible to compare the “level” of prices in a Canadian province and U.S. state, when the base period for each index differs. Engel and Rogers (1996) neatly avoid this problem by measuring the correlation of prices across locations, rather than their levels.

$$y^i = \sum_{j=1}^C c^{ij} \tau^{ij}. \quad (5.30)$$

Multiplying this by the price p^i , we obtain $p^i y^i = \sum_{j=1}^C p^j c^{ij}$, which is an equality between the value of output of the firm (using the f.o.b. prices but before the output has “melted”) and the expenditure of consumers (using the c.i.f. prices but after the quantity delivered has “melted”). In principle, we can use the market clearing conditions (5.30) to solve for the unknown prices p^i . Rather than obtain an *explicit* solution for prices, Anderson and van Wincoop (2003) make use of a convenient *implicit* solution. To develop their result, we again denote country i ’s GDP by $Y^i = N^i p^i y^i$, world GDP by $Y^w = \sum_{j=1}^C Y^j$, and country i ’s share by $s^i = Y^i / Y^w$.

THEOREM (ANDERSON AND VAN WINCOOP 2003)

Suppose that the transportation costs are symmetric, $\tau^{ij} = \tau^{ji}$. Then an implicit solution to the market clearing conditions (5.30) is

$$\tilde{p}^i \equiv (s^i / N^i)^{1/(1-\sigma)} / \tilde{P}^i, \quad (5.31)$$

in which case the price indexes are solved as

$$(\tilde{P}^j)^{1-\sigma} = \sum_{i=1}^C s^i (\tau^{ij} / \tilde{P}^i)^{1-\sigma}. \quad (5.32)$$

Proof:

We will show that substituting (5.31) into the price index (5.25), and substituting (5.31) into the market clearing conditions (5.30), both lead to equation (5.32). First, substituting $\tilde{p}^{ij} = \tilde{p}^i \tau^{ij} = (s^i / N^i)^{1/(1-\sigma)} \tau^{ij} / \tilde{P}^i$ into (5.25), we immediately obtain (5.32). Second, rewrite the market clearing conditions (5.30) as

$$Y^i = N^i p^i y^i = N^i p^i \sum_{j=1}^C c^{ij} \tau^{ij} = N^i \sum_{j=1}^C c^{ij} p^{ij} = N^i \sum_{j=1}^C (p^i \tau^{ij} / \tilde{P}^j)^{1-\sigma} Y^j, \quad (5.33)$$

where the final equality makes use of (5.24) and $p^{ij} = \tau^{ij} p^i$. Then substituting (5.31) into the final expression of (5.33), we obtain

$$Y^i = s^i \sum_{j=1}^C (\tau^{ij} / \tilde{P}^i \tilde{P}^j)^{1-\sigma} Y^j. \quad (5.34)$$

Summing this over $i = 1, \dots, C$, we have $Y^w = \sum_{j=1}^C (\tau^{ij} / \tilde{P}^i \tilde{P}^j)^{1-\sigma} Y^j$, which is identical to (5.32) provided that $\tau^{ij} = \tau^{ji}$. QED

To see the usefulness of this result, substitute (5.31) into the gravity equation (5.26) to obtain

$$X^{ij} = s^i Y^j \left(\frac{\tau^{ij}}{\tilde{P}^i \tilde{P}^j} \right)^{1-\sigma} = \left(\frac{Y^i Y^j}{Y^w} \right) \left(\frac{\tau^{ij}}{\tilde{P}^i \tilde{P}^j} \right)^{1-\sigma}. \quad (5.35)$$

This is a remarkably simple equation, whereby bilateral trade between countries depends on their GDPs and also their implicit price indexes. Anderson and van Wincoop call \tilde{P}^i “indexes of multilateral resistance,” because they depend on the transport costs τ^{ij} in (5.32). These indexes are unobserved, but Anderson and van Wincoop argue that we can solve for them by using equation (5.32) in combination with a formula for the transportation costs such as (5.29').²⁵

Specifically, the estimation strategy of Anderson and van Wincoop is to move the GDP terms from the right to the left side of (5.35), take logs, and substitute (5.29) for the transportation costs, obtaining (without the constant term Y^w)

$$\ln(X^{ij}/Y^i Y^j) = \rho(1-\sigma) \ln d^{ij} + (1-\sigma) \delta^{ij} + \ln(\tilde{P}^i)^{\sigma-1} + \ln(\tilde{P}^j)^{\sigma-1} + (1-\sigma) \varepsilon_{ij}. \quad (5.36)$$

The dependent variable on the left side is bilateral trade relative to the product of GDPs. On the right side we have distance between regions i and j , followed by all other border effects $(1-\sigma) \delta^{ij}$, and then the multilateral resistance terms $(\tilde{P}^i)^{\sigma-1}$. These multilateral resistance terms can be solved from (5.32) once we know the iceberg costs $\tau^{ij} = \tau^{ji}$. These transport costs, in turn, are obtained from (5.29) using the estimated value of $\rho(1-\sigma) \ln d^{ij} + (1-\sigma) \delta^{ij}$, which comes from (5.36). The estimation of this system must be custom programmed to minimize the sum of squared residuals in (5.36), while simultaneously using (5.29) to obtain the values of $\tau^{ij} = \tau^{ji}$ at each iteration, and with these, solving for the multilateral resistance terms $(\tilde{P}^i)^{\sigma-1}$ from (5.32).

To perform this estimation, we need to be more specific about the form of the border effects $(1-\sigma) \delta^{ij}$, in (5.36). Recall that in McCallum's (1995) gravity equation in (5.20), we introduced an indicator variable δ^{ij} that equaled unity for trade between two Canadian provinces, and zero otherwise. Anderson and van Wincoop instead work with an indicator variable that is $(1-\delta^{ij})$, or *unity for trade between the United States and Canada, and zero otherwise*. Introducing the coefficient γ on this variable, we replace $(1-\sigma) \delta^{ij}$ with $\gamma(1-\delta^{ij})$ in (5.36) and also use the coefficient $\alpha = \rho(1-\sigma)$ on distance, to obtain

$$\ln(X^{ij}/Y^i Y^j) = \alpha \ln d^{ij} + \gamma(1-\delta^{ij}) + \ln(\tilde{P}^i)^{\sigma-1} + \ln(\tilde{P}^j)^{\sigma-1} + (1-\sigma) \varepsilon_{ij}. \quad (5.37)$$

Their estimates of (5.37) are shown in column (4) of table 5.2. Notice that the provincial and states GDP terms have their coefficients constrained at unity, since they have been brought to the left side of (5.37). The coefficient on the indicator variable on cross-border trade in column (4) is estimated at $\hat{\gamma} = -1.65$. This can be compared to the estimates on the indicator variables for intra-Canada trade and intra-U.S. trade, in column (3), of 2.75 and 0.40, respectively. We certainly expect the indicator variables in columns (3) and (4) to have coefficients that are opposite in sign, since in the one case we are measuring *intrnational* trade and in the other case measuring *international* trade. It is noteworthy, however, that Anderson and van Wincoop's estimate in column (4) is roughly midway between the conventional gravity estimates in column (3), in absolute value. This appears to occur because Anderson and van Wincoop allow only a single indicator variable to measure cross-border trade, rather than distinguishing two

²⁵Notice that (5.32) determines \tilde{P}^i as a function of the transport costs τ^{ij} . We can write (5.32) alternatively, $1 = \sum_{i=1}^C s^i (\tau^{ij}/\tilde{P}^i \tilde{P}^j)^{1-\sigma}$, which shows that a weighted average of the terms $(\tau^{ij}/\tilde{P}^i \tilde{P}^j)^{1-\sigma}$ sums to unity, implying that these terms themselves are centered around unity. When $\tau^{ij} = 1$ for all i, j , for example, then the solution to (5.32) is $\tilde{P}^i = 1$ for all i . Note that in general the solution to (5.32) involves a normalization on the absolute level of prices, and therefore, on the absolute level of the implicit price indexes.

indicator variables (i.e., one for Canadian exports to the United States, and another for American exports to Canada). The use of a single indicator variable is related to their assumption that transport costs are symmetric, $\tau^{ij} = \tau^{ji}$.

How should we interpret the estimate $\hat{\gamma} = -1.65$? We would like to turn it into an estimate of how much more trade there is *within* Canada as compared to *across* the border. For regressions (1)–(3) in table 5.2, we simply took the exponent of the indicator coefficient, as reported near the bottom of the table. That approach is no longer appropriate, however, because if the border did not exist, then the multilateral resistance terms in (5.37) would also be affected. Let us denote by $(\tilde{P}^i)^{\sigma-1}$ the multilateral resistance terms in the *absence* of the border effect, that is, what we obtain from formula (5.32), but now using *only* distance (times its estimated coefficient) to compute T^{ij} in (5.29). Taking exponents of (5.37), and comparing this equation *with* and *without* border effects, the ratio of trade in these two cases is, therefore,²⁶

$$\frac{X^{ij}}{\tilde{X}^{ij}} = [e^{\hat{\gamma}(1-\delta^{ij})}] \frac{(\tilde{P}^i)^{\sigma-1} (\tilde{P}^j)^{\sigma-1}}{(\tilde{P}^i)^{\sigma-1} (\tilde{P}^j)^{\sigma-1}}. \quad (5.38)$$

For example, consider intra-Canadian trade, where $\delta^{ij} = 1$, so the first term on the right side of (5.38) vanishes. Anderson and van Wincoop find that (5.38) equals 4.3, meaning that intra-Canadian trade is *4.3 times larger with the border effects than without*.²⁷ In addition, for the United States, Anderson and van Wincoop find that intra-U.S. trade is *1.05 times larger with the border effects than without*. Finally, they find that cross-border trade is *0.41 times smaller with the border effects than without*. All these numbers are computed from the ratio on the right side of (5.38), averaged across the provinces or states in Canada or the United States, as appropriate. With intra-Canadian trade being 4.3 times higher due to the border effect, and cross-border trade being 0.41 times smaller, it is immediately clear that *intra-Canadian trade is $4.3/0.41 = 10.5$ times higher than cross-border trade*. This estimate is shown near the bottom of column (4) in table 5.2, along with its standard error. The analogous calculation for the United States shows that *intra-American trade is $1.05/0.41 = 2.6$ times higher than cross-border trade*, which is again shown at the bottom of column (4).

These estimates show that small economies, such as Canada, experience a much larger impact of the border effects. This is consistent with our numerical example earlier in the chapter, and indeed, the estimates of border effects obtained by Anderson and van Wincoop (10.5 and 2.6) are nearly the same as in our simple numerical example. Furthermore, these estimates have the following special property. The geometric mean of the border effects is $(10.5 \cdot 2.6)^{1/2} = 5.2$. Notice that this is the same as what we obtain by taking the exponent of the coefficient on the cross-border indicator variable (in absolute value), $e^{1.65} = 5.2$. So the geometric mean of the Canada and U.S. border effects, computing using formula (5.38), turns out to be identical to what we obtain by just using the cross-border indicator variable! This is no coincidence, and we provide a proof of this simple relation in Feenstra (2002). This result means that using cross-border indicator variable is a completely valid way to infer the *average* impact of the border on intranational relative to international trade.

²⁶We treat the province or state GDPs Y^i as unaffected by border effects, so that (5.38) follows by taking the exponent of (5.37) and its ratio *with* and *without* border effects.

²⁷This calculation requires the use of σ in (5.38). Anderson and van Wincoop use $\sigma = 5$, but find that the implied borders effects are not sensitive to this choice.

Using Fixed Effects

A drawback to the estimation strategy of Anderson and van Wincoop is that it requires custom programing to perform the constrained minimization (and obtain standard errors). A third and final approach to estimating the gravity equation, while using ordinary least squares (OLS), is to use fixed effects to take account of the unobserved price indexes. Fixed effects have been introduced into the gravity equation by a number of authors, including Harrigan (1996), Hummels (1999), Redding and Venables (2004), and Rose and van Wincoop (2001). We will summarize the results of the latter authors, as well as use fixed effects in estimating the Canada-U.S. gravity equation in table 5.2.

Since the multilateral indexes in (5.37) are unobserved, rather than calculating them according to (5.32), we could instead measure them as the coefficients of source and destination region fixed effects. That is, let δ_1^i denote an indicator variable that is unity if region i is the exporter, and zero otherwise; and let δ_2^j denote an indicator variable that is unity if region j is the importer, and zero otherwise. Then the gravity equation in (5.37) can be rewritten as

$$\ln(X^{ij}/Y^i Y^j) = \alpha \ln d^{ij} + \gamma(1 - \delta^{ij}) + \beta_1^i \delta_1^i + \beta_2^j \delta_2^j + (1 - \sigma) \varepsilon_{ij}, \quad (5.39)$$

where the coefficients $\beta_1^i = \ln(\tilde{P}^i)^{\sigma-1}$ and $\beta_2^j = \ln(\tilde{P}^j)^{\sigma-1}$ on the source and destination indicator variables estimate the multilateral indexes.

An estimate of (5.39) for Canada-U.S. intranational and international trade is shown in column (5) of table 5.2. Rather than the border estimate $\hat{\gamma} = -1.65$ as in Anderson and van Wincoop (2003), we now obtain $\hat{\gamma} = -1.55$. Taking its exponential we obtain $e^{-1.55} = 4.7$, as reported near the bottom of column (5). The value 4.7 is a consistent estimate of the *average* impact of the border barrier on Canada and U.S. trade relative to cross-border trade (as shown in the appendix to this chapter). Notice that this estimate is nearly the same as the average border effect of 5.2 obtained by Anderson and van Wincoop (2003), who explicitly introduced the multilateral resistance term computed from (5.32) into the estimation. In contrast, the fixed-effect approach estimates this term as part of the regression without relying on formula (5.32). Both approaches give consistent estimates of the average border effect. While using the explicit multilateral resistance terms should result in more efficient estimates, this benefit seems to be relatively small compared to the computational simplicity of the fixed-effect approach. Since the fixed-effects method produces consistent estimates of the *average* border effect across countries, and is easy to implement, it might be considered to be the preferred empirical method.

It is also interesting that the average border effect of 4.7 is nearly the same as the *average* effect obtained by a conventional gravity equation (like McCallum 1995): in column (3) of table 5.2, we compute the geometric mean of the border effect for Canada and the United States as 4.8. In contrast to columns (4) and (5), the estimates in column (1)–(3) are *not consistent*, because these regressions do not incorporate the price indexes. This appears to have the effect of *overstating* the border effect for Canada (15.7 in column (3) as compared to 10.5 in column (4)), and *understating* the effect for the United States (1.5 in column (3) as compared to 2.6 in column (4)). Still, for this dataset, the geometric mean of these inconsistent estimates gives an average border effect that is very close to the consistent estimates in columns (4) and (5).

The fixed-effect approach has been used by Rose and van Wincoop (2001) to estimate the impact of monetary unions on international trade. They include an indicator variable that is unity if the two countries belong to a monetary union, and interpret its coefficient of $\hat{\gamma} = 0.86$ as a consistent estimate of the average impact of monetary unions on trade. Surprisingly, they find that monetary unions increase trade by $e^{\hat{\gamma}} = 2.36$, or more than doubling trade on average between union members relative to non-union members. The actual mechanism by which monetary unions lead to such a large increase in trade remains quite unclear. Results like this have led Obstfeld and Rogoff (2000) to identify large border effects from the gravity equations as the cause of the “six major puzzles in international macroeconomics.”

The fixed effect method has also been used by Redding and Venables (2004) to determine *wages* across countries. They do not rely on symmetric transportation costs, or the above theorem of Anderson and van Wincoop, but instead work directly from the gravity equation in (5.26). In their approach, the differentiated product is used as both consumer goods and as intermediate inputs to production. Thus, proximity to trading partners affects a country’s ability to export the differentiated good and import the differentiated inputs: both of these activities will impact wages. Redding and Venables show that the fixed effects of the gravity equation are directly related to equilibrium wages. In other words, the economic geography of a country—measured by its distance from and access to trading partners—determines its wages and hence its standard of living.

In the next section, we turn to another question of economic geography that has to do with the *location* of firms. Specifically, as we consider regions or countries of differing size, will the higher demand in large areas serve to attract more firms? Since each product is produced by only one firm in our model, asking about the location of firms is the same as asking about how many products N^i each country produces. We address this in the next section, using a simplified framework where wages are constant and equal across countries.

THE HOME MARKET EFFECT

We will make the same assumptions on the differentiated good as in Krugman’s original model: labor is the only input, and output y^i of the typical firm in country i requires the labor input $\alpha + \beta y^i$. In addition, we will suppose that there is a homogeneous good, and one unit of that good requires one unit of labor. There are no transport costs in this good, so its price is equalized across countries.²⁸ Choosing the homogeneous good as the numeraire, and provided that each country produces that good under trade, wages are therefore *unity* in all countries. Having wages fixed will simplify our determination of N^i .

On the demand side, we assume that a fixed share of income γ is spent on the differentiated product, which provides utility given by the CES function in (5.21) or (5.22).²⁹ Therefore, demand for each differentiated product is still given by (5.25), but

²⁸The assumption of zero transport costs on the numeraire good is important. Davis (1998) argues that we need higher transport costs on the differentiated good than in the homogeneous product to obtain the home market effect, as derived below. See also Fujita, Krugman, and Venables (1999, chapter 7).

²⁹Denoting consumption of the homogeneous good by c_0 , the utility function is $U = c_0^{1-\gamma} (\sum_{i=1}^N c_i^{(\sigma-1)/\sigma})^{\gamma\sigma/(\sigma-1)}$.

replacing income Y^j by the amount γL^j actually spent on the differentiated product, we obtain

$$c^{ij} = (p^{ij}/P^j)^{-\sigma} \gamma L^j/P^j. \quad (5.40)$$

The output of firms is still given by (5.30), $y^i = \sum_{j=1}^C c^{ij} \tau^{ij}$. Because of the CES utility function, this output will be fixed in a zero-profit equilibrium, $\bar{y} = (\sigma - 1) \alpha / \beta$, as obtained in (5.12). So in principle, the number of products in each country can be derived from the market clearing conditions

$$\bar{y} = \sum_{j=1}^C c^{ij} \tau^{ij}, \quad i = 1, \dots, C, \quad (5.41)$$

where consumption depends on the price indexes P^j given by (5.25), which depend on the number of products.

Rather than solving for the number of products in each country, we will instead solve for the *change* in these as country size L^j varies. The fact that the firm outputs are fixed on the left-hand side of (5.41) implies that on the right-hand side some combination of the consumptions in each country must also be fixed. As a guess, let us suppose that *consumption of every variety in each country is fixed*. Under our framework where wages are equalized across countries, the prices p^i and p^{ij} are also fixed. Then we can totally differentiate (5.40), holding c^{ij} fixed, to obtain

$$0 = (\sigma - 1) \hat{P}^j + \hat{L}^j \Rightarrow \hat{P}^j = -\hat{L}^j / (\sigma - 1), \quad (5.42)$$

where $\hat{z} = dz/z$ for any variable. This equation states that the price indexes in each country j will fall in direct proportion to the rise in country GDP. Recall from (5.25) that the price indexes depend on the number of products in each country. Thus, provided that there is a change in the number of products consistent with \hat{P}^j in (5.42), then our initial assumption that the consumption c^{ij} is fixed is indeed correct.

To determine the change in the number of products, let us differentiate (5.25) to obtain

$$\hat{P}^j = \left(\frac{1}{1-\sigma} \right) \left(\sum_{i=1}^C \hat{N}^i \lambda^{ij} \right), \quad (5.43)$$

where $\lambda^{ij} = N^{ij} (p^{ij}/P^j)^{1-\sigma}$ denotes the share of products from country i in the differentiated goods purchased by country j , with $\sum_{i=1}^C \lambda^{ij} = 1$. Combining (5.42) and (5.43), we obtain an extremely simple relationship between the change in GDP of each country and the change in the number of products,

$$\hat{L}^j = \sum_{i=1}^C \hat{N}^i \lambda^{ij}. \quad (5.44)$$

We can use (5.44) to solve for the change in number of product \hat{N}^i provided that the matrix of expenditure shares $\Lambda = [\lambda^{ij}]$ is invertible. This will be true, for example, if each country i devotes a larger share of its budget to purchasing its own differentiated products than the budget share that country j uses to buy products from i , so that $\lambda^{ii} > \lambda^{ij}$ for $i \neq j$, as we shall assume.

When there are just two countries, we can express (5.44) in matrix notation as

$$(\hat{N}^1, \hat{N}^2) \begin{pmatrix} \lambda^{11} & \lambda^{12} \\ \lambda^{21} & \lambda^{22} \end{pmatrix} = (\hat{L}_1, \hat{L}_2) \Rightarrow (\hat{N}_1, \hat{N}_2) = \frac{(\hat{L}_1, \hat{L}_2)}{|\Lambda|} \begin{pmatrix} \lambda^{22} & -\lambda^{12} \\ -\lambda^{21} & \lambda^{11} \end{pmatrix}, \quad (5.45)$$

where $|\Lambda|$ denotes the determinant of the two-by-two matrix above. This determinant can be alternatively expressed as

$$\begin{aligned} |\Lambda| &= \lambda^{11}\lambda^{22} - \lambda^{12}\lambda^{21} \\ &= \lambda^{11}(1-\lambda^{12}) - \lambda^{12}(1-\lambda^{11}) \\ &= \lambda^{11} - \lambda^{12} = \lambda^{22} - \lambda^{21} \end{aligned} \quad (5.46)$$

where we have repeatedly made use of the fact that $\lambda^{ij} + \lambda^{2j} = 1$, that is, the consumption shares of products from both sources to country j sum to unity. This determinant is positive under our assumption that $\lambda^{ii} > \lambda^{ij}$ for $i \neq j$.

To fix ideas, suppose that country 1 and 2 start off as identical, with equal transport costs between them. Similar to Krugman's model at the beginning of this chapter, trade between the identical countries will lead them to export equal numbers of the differentiated product to each other: $N^1 = N^2$, and trade is balanced in the differentiated good. Now suppose that the labor endowment of country 1 grows, with no change in country 2, so that $\hat{L}^1 > 0$, and $\hat{L}^2 = 0$. Then we can solve for the changing number of products from (5.45) and (5.46) as

$$\hat{N}^1 = \frac{\lambda^{22}}{(\lambda^{22} - \lambda^{21})} \hat{L}^1 > \hat{L}^1 > 0 \quad \text{and} \quad \hat{N}^2 = \frac{-\lambda^{12}}{|\Lambda|} \hat{L}^1 < 0. \quad (5.47)$$

Thus, the number of products in the larger country will grow by *more than* the increase in country size, while those in the smaller country will *shrink*. Because consumption c^{ij} and prices p^{ij} are both fixed, exports $X^{ij} = N^i p^{ij} c^{ij}$ from each country to the other will change in proportion to the number of products. Thus, with N^1 growing and N^2 falling, country 1 will become a *net exporter* of the differentiated good to country 2, $X^{21} > X^{12}$ (where trade is balanced overall through flows of the homogeneous good). Summarizing these results, we have the following theorem.

THEOREM (KRUGMAN 1980)

With two countries trading, the larger market will produce a greater number of products and be a net exporter of the differentiated good.

This result is known as the *home market effect*, because it shows that a larger home market will attract disproportionately more firms, and therefore become a net exporter. This is quite different from what we expect in a model where the number of products N^i is constant: in that case, a larger market would have larger demand, and would therefore become a *net importer* of the good in question.³⁰ While we have

³⁰The assumption that products are differentiated by their country of origin, but the number of varieties supplied by each country is *fixed*, is known as the "Armington assumption," from Armington (1969). This was used as a convenient assumption before the introduction of monopolistic competition models. The fact that the number of varieties per country is fixed in the Armington assumption gives it quite different properties than monopolistic competition, where the number of varieties produced in each country varies due to free entry.

shown this result above for a simple two-industry, two-country model, it can be generalized to many industries and countries (Behrens et al. 2004; Hanson and Xiang 2004).

Davis and Weinstein (1996, 1999) provide a direct test of the home market effect, using industry level data for the OECD. They first measure “idiosyncratic” demand differences across countries, and then argue that having higher home demand should lead a country to be an exporter of a good if the “home market effect” operates.³¹ They test this by regressing production of industries on the estimated demand differences across countries. In their pooled sample, they obtain an estimated coefficient of 1.6 on the demand differences, indicating that having 10 percent greater demand for a product will lead to 16 percent additional production in that country, meaning that net exports will rise. This provides confirmation of the home market effect. When disaggregating across industries, they find that the impact of local demand on production exceeds unity in a majority of industries, though there are a few exceptions. When using data for Japanese industries locating in various prefectures, they find significant evidence of home market effects in about one-half of the industries, and these effects are economically large.

Head and Ries (2001) test for the home market effect between Canada and the United States. With only two countries, they need to rely on cross-industry or time-series variation in the data, rather than the cross-country demand differences used by Davis and Weinstein. For the cross-industry regression they find a weak home market effect, but for the time-series regression the home market effect is reversed (i.e., higher demand leads to imports rather than exports).

A similar reversal is found by Feenstra, Markusen, and Rose (2001), when comparing cross-country trade in *differentiated* versus *homogeneous* products.³² In the gravity equation for differentiated products, they find that the coefficient on exporter GDP exceeds unity, whereas the coefficient of the importer GDP is less than unity. Having the coefficient on exporter GDP exceed unity is like the comparative static effect reported in (5.47): the change in exports from country 1 to country 2 is $\hat{X}^{12} = \hat{N}^1 > \hat{L}^1$, so having higher country 1 income (L^1) leads to a *magnified* increase in exports. For the homogeneous products, however, this result is reversed and the exporter GDP has a coefficient smaller than importer GDP. The latter result is what we would expect from a model where the number of products in each country is fixed, so that higher demand leads to net imports rather than net exports.

CONCLUSIONS

This chapter has covered material in two topics—trade under monopolistic competition and the gravity equation—that are often linked, though they need not be. As we have mentioned, the gravity equation arises quite naturally whenever countries are specialized in different goods. Such specialization is sometimes called *national product differentiation*, and cross-border trade in different varieties of a good is referred to

³¹ Note that we could have let γ differ between the countries, and then the home market effect still operates.

³² The classification of goods according to whether they are differentiated or homogeneous is due to Rauch (1999). He has classified products at the five-digit SITC level according to whether they are (a) traded in an organized exchange, and therefore treated as homogeneous; (b) not traded in an organized exchange, but having some quoted “reference price,” such as in industry publications; (c) not having any quoted prices, and therefore treated as differentiated. See: <http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/TradeData.html#Rauch> (accessed August 21, 2014).

as *intra-industry trade*. This occurs under the monopolistic competition model, but also occurs in other contexts, for example, Dornbusch, Fischer, and Samuelson's (1977) Ricardian model with a continuum of goods, or their (1980) Heckscher-Ohlin (HO) model where factor prices are not equalized, as studied in chapter 3. An elegant extension of Ricardian differences is provided by Eaton and Kortum (2002), who consider *stochastic* differences in the technologies across countries, with the lowest-cost country becoming the exporter of each product variety. They obtain a gravity-like equation, which should include country fixed effects that are related to the probability distribution of technologies within each country, as we shall discuss in the next chapter.

So while complete specialization in goods can arise in a number of contexts, the CES monopolistic competition model proves to be a convenient way to derive the gravity equation, especially when we allow for transport costs and other trade barriers. Unlike early work on the gravity equation that assumed identical prices across countries, once we introduce transportation costs or any other border barriers, prices must differ internationally, so we need to take account of the overall price indexes in each country. We have reviewed three methods to do so: using published data on price indexes; using the computational method of Anderson and van Wincoop (2003); or using country fixed effects to measure the price indexes. The latter two methods were compared with a dataset dealing with trade between and within Canada and the United States. The fixed-effects method produces consistent estimates of the *average* border effect across countries, and is simple to implement, so it might be considered the preferred empirical method.

The fact that the gravity equation works well empirically cannot, however, be taken as evidence in support of the monopolistic competition model: it simply suggests that countries are specialized in different products, for whatever reason. To really test whether trade patterns are due to monopolistic competition, we must look for hypotheses arising from this model that would not occur otherwise. One such hypothesis is the *home market effect*, whereby firms tend to locate in the larger market due to higher profits available there. Under alternative explanations for national product differentiation, the home market effect need not occur: if the number of products N^i produced in each country is constant, then a larger market would have larger demand, and would therefore become a *net importer* of the good in question. So evidence of the home market effect, as found by Davis and Weinstein (1996, 1999) and more weakly by Feenstra, Markusen, and Rose (2001) as well as Head and Ries (2001), lends support to the idea that monopolistic competition explains national product specialization.

An important area for further research is to test other hypotheses arising from that model. One direction to look is in the area of economic geography. An expanded discussion of that topic is beyond the scope of this book, but the reader is referred to Fujita, Krugman, and Venables (1999) and Redding and Turner (2015), who describe many results building on increasing returns to scale and the monopolistic competition model. For example, this literature suggests that proximity to markets will raise wages, and Redding and Venables (2004) apply this idea to estimate wages across countries based on the gravity equation. A similar approach can be used to determine wages across cities or regions *within* a country, as done by Hanson (1997, 1998).³³ In other empirical work, Davis and Weinstein (2002a) have used the location of economic activity in Japan, going from prehistoric times to wartime dislocations to the present day, to evaluate the importance of increasing returns versus alternative explanations

³³See also the survey by Overman, Redding, and Venables (2003).

for industry location. Similarly, Redding, Sturm, and Wolf (2011) use the reunification of East and West Germany, and the subsequent relocation of Germany's air hub from Berlin to Frankfurt, to evaluate whether increasing returns to scale led to multiple steady-state equilibria.

Another direction to look for hypotheses is to loosen some of the assumptions of the monopolistic competition model. Two key assumptions are that firms are *symmetric* (their costs are all the same) and that they each produce a *single product*. Both of these assumptions are false in reality: every industry has a wide range of firms of various efficiencies, and many of them produce multiple products. The theoretical difficulty with allowing for firms with differing efficiency is that the zero-profit condition of monopolistic competition will no longer apply: if the less efficient firms earn zero profits, then the more efficient firms will not. This difficulty has been resolved in different ways by Melitz (2003) and Yeaple (2005). In Melitz's model, firms have random productivities and earn zero *expected* profits, while in Yeaple's model, the *workers* in more efficient firms receive higher wages, so profits are still zero. The Melitz model will be described in detail in the next chapter.

APPENDIX: MEASURING THE ELASTICITY OF SUBSTITUTION

We suppose that the utility function in (5.13) applies to the purchases of varieties of a single good from various source countries $i \in I_t$. That is, each country is exporting a different variety of the good in question (cars, shoes, etc.), which was the assumption made by Armington (1969) and Krugman (1980). Then the gains being measured in (5.17) are the gains from greater import variety of that good, but in order to implement it we need a reliable estimate of the elasticity of substitution across source countries for each good, which is closely related to the elasticity of demand. This parameter is challenging to estimate due to the standard "identification problem": with shifting supply and demand curves, the equilibria create a cloud of points in quantity-price space, and so it is hard to pin down the elasticities of demand and supply without knowing the variables that are shifting those curves. We normally assume that variables are orthogonal to the error terms in the supply-and-demand equations, so they become instrumental variables. But with many traded goods, it is hard to find instruments for each one. Feenstra (1994) proposed a generalized method of moments (GMM) estimator to resolve this problem that makes use of the panel nature of datasets in international trade, in other words, having time-series observations on the amount imported from multiple source-countries. This estimator can be equivalently interpreted as using instrument variables that are the fixed effects for source countries, as we shall explain.

For convenience, we will drop the superscript j denoting the destination country. Then from (5.26), the share of expenditure on each source country i is

$$s_t^i \equiv \frac{X_t^i}{Y_t} = N_t^i \left(\frac{p_t^i}{P_t} \right)^{1-\sigma}. \quad (5.48)$$

Taking natural logs and taking the difference over time, we can write this demand equation as

$$\Delta \ln s_t^i = (\sigma - 1) \Delta \ln P_t - (\sigma - 1) \Delta \ln p_t^i + \varepsilon_t^i \quad (5.49)$$

where $\varepsilon_t^i \equiv \Delta \ln N_t^i$ is an error term reflecting variety in the exporting country. To this demand curve we add a supply curve, which is assumed to be

$$\Delta \ln p_t^i = \beta \Delta \ln c_t^i + \xi_t^i \quad (5.50)$$

where ξ_t^i is the random error in supply. It is inconvenient to have the quantity appearing in the supply curve (5.50) and the share of expenditure appearing in the demand curve (5.49). To resolve this slight inconsistency, we combine (5.49) and (5.50) to eliminate the quantity from the supply curve, obtaining

$$\Delta \ln p_t^i = \psi_t + \frac{\rho \varepsilon_t^i}{(\sigma - 1)} + \delta_t^i \quad (5.51)$$

where the parameters appearing in this “reduced form” supply curve are

$$\psi_t = \frac{\beta[(\sigma - 1)\Delta \ln P_t + \Delta \ln Y_t]}{(1 + \beta\sigma)} \text{ and } \rho = \frac{\beta(\sigma - 1)}{(1 + \beta\sigma)}, \quad 0 < \rho < 1, \quad (5.52)$$

and the error term is

$$\delta_{it} = \frac{\xi_{it}}{(1 + \beta\sigma)}. \quad (5.53)$$

Notice that the error term $\varepsilon_t^i \equiv \Delta \ln N_t^i$ in the demand equation (5.49) is correlated with the price in (5.51). This illustrates the simultaneous equations problem, whereby an outward shift in the demand curve (i.e., an increase in export variety so that $\varepsilon_t^i \equiv \Delta \ln N_t^i > 0$) will lead to a rise in price along the supply curve. This fact means that OLS cannot be used to estimate the demand equation (5.49). To make headway, we shall assume that the error terms ε_t^i and $\delta_t^i, i=1, \dots, C$ are all independent, with mean zero and variances σ_{ε_i} and σ_{δ_i} , respectively. This assumption means that the error terms in the demand-and-supply equations are uncorrelated, which will give us a moment condition that we can use for estimation. To develop this condition, we begin by differencing the demand and supply equations with respect to some country k , as

$$\begin{aligned} \tilde{\varepsilon}_t^i &= \varepsilon_t^i - \varepsilon_t^k = (\Delta \ln s_t^i - \Delta \ln s_t^k) + (\sigma - 1)(\Delta \ln p_t^i - \Delta \ln p_t^k), \\ \tilde{\delta}_t^i &= \delta_t^i - \delta_t^k = (\Delta \ln p_t^i - \Delta \ln p_t^k) + \rho \tilde{\varepsilon}_t^i / (\sigma - 1) \\ &= (1 - \rho)(\Delta \ln p_t^i - \Delta \ln p_t^k) + [\rho / (\sigma - 1)](\Delta \ln s_t^i - \Delta \ln s_t^k). \end{aligned}$$

Multiply these two equations together and dividing by $(1 - \rho)(\sigma - 1)$, we obtain

$$Y_t^i = \theta_1 X_{1t}^i + \theta_2 X_{2t}^i + u_t^i, \quad (5.54)$$

where

$$Y_t^i = (\Delta \ln p_t^i - \Delta \ln p_t^k)^2 \quad (5.55)$$

$$X_{1t}^i = (\Delta \ln s_t^i - \Delta \ln s_t^k)^2 \quad (5.56)$$

$$X_{2t}^i = (\Delta \ln p_t^i - \Delta \ln p_t^k)(\Delta \ln s_t^i - \Delta \ln s_t^k), \quad (5.57)$$

$$\theta_1 = \frac{\rho}{(\sigma-1)^2(1-\rho)}, \quad \theta_2 = \frac{(2\rho-1)}{(\sigma-1)(1-\rho)}, \quad (5.58)$$

and

$$u_t^i = \frac{\tilde{\varepsilon}_t^i \tilde{\delta}_t^i}{(\sigma-1)(1-\rho)}. \quad (5.59)$$

Finally, let us average the variables appearing in (5.54)–(5.57) over time for each source country country $i=1,\dots,C$. and obtain the following equation that is estimated as a cross-country regression:

$$\bar{Y}^i = \theta_1 \bar{X}_1^i + \theta_2 \bar{X}_2^i + \bar{u}^i, \quad i=1,\dots,C. \quad (5.60)$$

Notice that the variables appearing in (5.60) are second moments of the data, or essentially the variances and covariance of log differences of prices and shares. Furthermore, as $T \rightarrow \infty$, the error term in (5.60) vanishes in its probability limit, because this error term is the cross-moment of the errors in the supply and demand curves, which are uncorrelated by assumption. Therefore, minimizing the sum of squares errors by running OLS on (5.60) is a GMM estimator that will give consistent estimates θ_1 and θ_2 provided that the two right-hand side variables in (5.60) are not colinear. Feenstra (1994) shows that this condition holds provided that there exist countries i and j such that

$$\frac{\sigma_{\varepsilon_i}^2 + \sigma_{\varepsilon_k}^2}{\sigma_{\varepsilon_j}^2 + \sigma_{\varepsilon_k}^2} \neq \frac{\sigma_{\delta_i}^2 + \sigma_{\delta_k}^2}{\sigma_{\delta_j}^2 + \sigma_{\delta_k}^2}. \quad (5.61)$$

Condition (5.61) simply states that there are some countries i and j for which the relative variance of errors in supply and demand are different. In other words, there is some heteroskedasticity in the errors terms for supply and demand across countries. That is a very plausible assumption, since the error terms in supply depend on country-specific shocks to wages, productivity and exchange rates, while the error terms in demand depend on changes to the quality of goods supplied by each country. This is an example of “identification through heteroskedasticity,” discussed in detail by Rigobon (2003), who provides examples in other fields such as finance.

Feenstra (1994) includes a constant term in (5.54) or (5.60) to control for measurement error in the prices, which are typically import unit-values. Since the dependent variable is the second moment of prices, then measurement error that has constant variance across countries (and is uncorrelated with the shares) will appear as a constant term. Feenstra (2010b, appendix 2.1) further argues that there are implicitly some instrumental variables used to estimate (5.54), and these are the source country fixed effects. Running the right-hand side variables X_{1t}^i and X_{2t}^i on the country fixed effects will give the average values \bar{X}_{1t}^i and \bar{X}_{2t}^i as predicted values, so that (5.60) is essentially the second-stage of the IV estimation.³⁴ While condition (5.61) is enough to ensure that the IV estimates are consistent, efficient estimates require that weighted

³⁴Note that the estimates obtained from (5.60) do not depend on whether or not the dependent variable Y_t^i is averaged to obtain \bar{Y}^i .

IV is used. To achieve that, an initial IV estimate is obtained, and then for each source country the variance of the predicted errors in (5.54) are constructed. Then using the inverse of this variance as weights, a second set of IV estimates is obtained. The STATA code to obtain the IV and weighted IV estimates is included in empirical exercise 5.2 of this chapter.³⁵

Given the efficient estimates $\hat{\theta}_1, \hat{\theta}_2$, a quadratic equation obtained from (5.58) is solved to obtain $\hat{\sigma}, \hat{\rho}$. So long as $\hat{\theta}_1 > 0$, we obtain real estimates of σ and ρ , which are

$$(a) \text{ if } \hat{\theta}_2 > 0, \text{ then } \hat{\rho} = \frac{1}{2} + \left[\frac{1}{4} - \frac{1}{4 + (\hat{\theta}_2^2/\hat{\theta}_1)} \right]^{1/2},$$

$$(b) \text{ if } \hat{\theta}_2 < 0, \text{ then } \hat{\rho} = \frac{1}{2} - \left[\frac{1}{4} - \frac{1}{4 + (\hat{\theta}_2^2/\hat{\theta}_1)} \right]^{1/2},$$

$$\text{and in either case, } \hat{\sigma} = 1 + \left(\frac{2\hat{\rho}-1}{1-\hat{\rho}} \right) \frac{1}{\hat{\theta}_2} > 1.$$

Note that as $\hat{\theta}_2 \rightarrow 0$, then $\hat{\rho} \rightarrow \frac{1}{2}$ and $\hat{\sigma} \rightarrow 1 + \hat{\theta}_1^{-1/2}$.

In the event that $\hat{\theta}_1$ is negative, the above formulas fail to provide estimates for σ and ρ that are both in the ranges $\hat{\sigma} > 1$ and $0 \leq \hat{\rho} < 1$. For example, if $\hat{\theta}_1$ is only slightly less than zero ($-\hat{\theta}_2^2/4 < \hat{\theta}_1 < 0$), then $\hat{\rho} \notin [0, 1]$, and $(\hat{\sigma} - 1)$ has the same sign as $-\hat{\theta}_2$. In this case it is still possible that a value of $\hat{\sigma}$ exceeding unity is obtained. However, if $\hat{\theta}_1$ is too low ($\hat{\theta}_1 < -\hat{\theta}_2^2/4$), then imaginary values for $\hat{\rho}$ and $\hat{\sigma}$ are obtained. In that case, Broda and Weinstein (2006) implement a grid-search procedure to determine the real estimates.³⁶

PROBLEMS

5.1 For the utility function in (5.1), with $v(0)=0$, $v'>0$, and $v''<0$, show that when the number of goods rises holding income fixed, the consumer is better off.

5.2 In Krugman's monopolistic competition model, suppose that the utility function takes on the CES form shown in (5.13), rewritten for simplicity with two goods as

$$U(c_1, c_2) = [a_1 c_1^\theta + a_2 c_2^\theta]^{1/\theta}, \quad 0 < \theta = (\sigma-1)/\sigma < 1.$$

Maximize this subject to the budget constraint $p_1 c_1 + p_2 c_2 \leq w$, where w equals the wage (or income) of the consumer.

- (a) Obtain an expression for the *relative demand* c_1/c_2 as a function of prices.
- (b) The *elasticity of substitution* is defined as $d \ln(c_1/c_2) / d \ln(p_2/p_1)$. What is the value of the elasticity of substitution for this utility function?
- (c) Obtain an expression for the demands c_1 and c_2 as a function of prices.

³⁵The STATA software to implement this method was prepared by Anson Soderbery.

³⁶See the software provided by Anson Soderbery in Feenstra (2010b, appendix 2.2).

- (d) Show that the elasticity of demand for good i takes the form $-\partial \ln c_i / \partial \ln p_i = \sigma + (1 - \sigma) s_i$, where $s_i = (p_i c_i / w)$ is the share of expenditure on good i .
- 5.3 In problem 5.2, substitute your solution for c_1 and c_2 into the utility function to obtain *indirect* utility as a function of prices and income, w . Then show that indirect utility equals $w/[p_1^{1-\sigma} + p_2^{1-\sigma}]^{1/(1-\sigma)}$. Note that for $a_1 = a_2 = 1$, this formula is the same as $w/e(p, I)$ for the expenditure function $e(p, I)$ shown in (5.14). Therefore, $w/e(p, I)$ is the “real wage” earned by consumers.
- 5.4 Suppose that industry 1 is monopolistically competitive, with a CES subutility function as described in problem 5.2. We denote the marginal costs by $c_1(w, r)$, and the fixed costs in the industry by $\alpha c_1(w, r)$. That is, the fixed costs use labor and capital in the same proportions as the marginal costs. Industry 2 is a competitive industry, and each industry uses labor and capital.
- Write down the relationship between the prices of goods and factor prices.
Does the Stolper-Samuelson theorem still apply?
 - Write down the full-employment conditions for the two factors. Does the Rybczynski theorem still apply in some form?
 - How are your answers to (a) and (b) affected if the fixed costs in industry 1 use different proportions of labor and capital than the marginal costs?
- 5.5 Consider the translog expenditure function

$$\ln E(p, U) = \ln U + a_0 + \sum_{i=1}^N a_i \ln p_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N b_{ij} \ln p_i \ln p_j \quad (5.51)$$

where it is convenient to impose the symmetry restrictions

$$a_i = 1/N, b_{ii} = -\gamma(N-1)/N, \text{ and } b_{ij} = \gamma/N \text{ for } i \neq j, \text{ with } i, j = 1, \dots, N, \quad (5.52)$$

and where $\gamma > 0$. Use this expenditure function to answer the following:

- Show that restrictions in (5.52) ensure that the expenditure function in (5.51) is homogeneous of degree one in prices;
 - Derive the share of expenditure on good i by differentiating (5.51) with respect to $\ln p_i$;
 - Differentiate the share of each good with respect to $\ln p_i$ to derive an expression for the elasticity of demand;
 - Use the elasticity of demand to obtain an expression for the ratio of price to marginal cost for the firm. How will each firm's price depend on its competitor's prices? Note that this symmetric form of the translog expenditure function is justified by Feenstra (2003) and Bergin and Feenstra (2009), and has been used by Bergin and Feenstra (2000, 2001) and Feenstra and Weinstein (2010) in monopolistic competition models.
- 5.6 To establish (5.18'), we set $\prod_n \lambda_n^{-\alpha_n/(\sigma_n-1)} = \lambda^x$ and take natural logs, so that

$$-\sum_n \left(\frac{\alpha_n}{\sigma_n - 1} \right) \ln \lambda_n = x \ln \lambda.$$

Now simplify this expression to show that (5.18') is obtained. Whenever λ is close to unity then we can use the approximation $\ln \lambda \approx \lambda - 1$. It follows that $\ln \lambda_n / \ln \lambda \approx (1 - \lambda_n) / (1 - \lambda)$. As in the main text, we let M_n denote the value of imports and let D_n denote the value of demand from domestic

sources, so that $C_n \equiv M_n + D_n$ denotes the value of total consumption, and then $\lambda_n = D_n/C_n = 1 - (M_n/C_n)$, while $\lambda \equiv (\sum_n D_n)/(\sum_n C_n) = 1 - [(\sum_n M_n)/(\sum_n C_n)]$. Then use $\alpha_n = C_n/(\sum_n C_n)$ to show that $\sum_n \alpha_n(1 - \lambda_n)/(1 - \lambda) = \sum_n (M_n/\sum_i M_i) = 1$.

EMPIRICAL EXERCISES

- 5.1 In this exercise, you are asked to reproduce the empirical results shown in table 5.2. There are four datasets available: “dist.csv,” which is distances; “gdp_ce_93.csv,” which is GDP in exporting location in 1993; “gdp_ci_93.csv,” which is GDP in importing location in 1993; and “trade_93.csv,” which is trade in 1993. To complete the exercise, these files should be stored in the directory: c:\Empirical_Exercise\Chapter_5\. When the files are located there, run the STATA program “data_trans.do,” which will convert these datasets to STATA files with the same name. The trade data are already converted into U.S. dollars, but GDP data are in Canadian dollars, so this is converted with the exchange rate 1 Canadian dollar = 0.775 U.S. dollars. Then do the following:
 - a) Run the program “gravity_1.do” to replicate the gravity equations in columns (1)–(3) of table 5.2.
 - b) Run the program “gravity_2.do” to replicate gravity equation using fixed effects, that is, column (5) in table 5.2. Then answer the following:
 - (i) How are these results affected if we allow the provincial and state GDPs to have coefficients different from unity?
 - (ii) What coefficients are obtained if we introduce separate indicator variables for intra-Canadian and intra-U.S. trade, rather than the border dummy?
- 5.2 In this exercise, you are asked to reproduce the elasticities of substitution estimated by Feenstra (1994) using the generalized method of moments (GMM). The appendix to this chapter describes the equations for this method. To implement it, you should download the datasets: twdata.dta (typewriters), asdata.dta (athletic shoes), ksdata.dta (knit shirts), sbdata.dta (steel bars), ssdata.dta (steel sheets), and tvdata.dta (TV receivers). To complete this exercise, these files should be stored in the directory c:\Empirical_Exercise\Chapter_5\. After the data files are in this folder, download the STATA program “Problem_5_2.do” to the same folder and run it to replicate table 2 from Feenstra (1994).

Monopolistic Competition and the Gravity Equation II

In the previous chapter we introduced the idea of increasing returns to scale and used monopolistic competition as the market structure to analyze international trade. Our treatment there covered aspects of the literature from the early models of Krugman (1979, 1980, 1981) through the treatment of the gravity equation in Anderson and van Wincoop (2003). All of those models share the common assumption that firms are identical, so that the equilibrium is symmetric across firms. But there is now a new generation of monopolistic competition models that allow for heterogeneous firms. In this chapter we discuss the work of Melitz (2003), which allows firms to differ in their productivities but maintains the assumption of constant elasticity of substitution (CES) preferences used in Krugman (1980, 1981).¹ Allowing for heterogeneous firms is consistent with the idea that any industry includes firms with a wide range of capabilities, and, as we shall see, is also consistent with the observation that often only a subset of firms in each industry are exporters. We will find that opening trade in this model gives results that are very different from those given in the CES models of Krugman (1980, 1981), because we will now find a strong *selection* effect of trade, whereby the least efficient firms exit, and also a *scale* effect whereby the most efficient firms expand production for export. So allowing for differing productivities across firms is both realistic and leads to new results.

But making this extension raises a serious conceptual problem right from the start: if firms differ in their productivities, then how can we impose a zero-profit equilibrium respecting the free entry assumption of monopolistic competition? If some firms in the industry make zero profits, then won't more efficient firms make positive profits, implying that the zero-profit condition is violated? It is these difficult questions that explain why more than twenty years passed between the original monopolistic competition and trade models, featuring identical firms, and the new generation of models allowing for heterogeneous firms. Just as the first generation of models needed a micro-founded framework (due to Dixit and Stiglitz [1977], as well as Lancaster [1975, 1979] and Spence [1976]), so too the new generation

¹ An alternative model of firm heterogeneity is presented by Yeaple (2005), who starts with firms that are homogeneous over the available technologies, which in turn require workers who are heterogeneous in their skills. In equilibrium, firms hire different types of workers and therefore become heterogeneous in their labor productivity.

of models relies on a stochastic framework due to Hopenhayn (1992). Hopenhayn presented a dynamic, general equilibrium model of firm entry and exit under perfect competition. Melitz (2003) extended that framework to allow for monopolistic competition, using the Dixit-Stiglitz model, and applied it to multiple countries to obtain a tractable model of international trade.

Instead of having just one zero-profit condition, as in the traditional monopolistic competition model with identical firms, in the Melitz (2003) model there are two conditions: the first states that any firm engaged in production must make nonnegative profits, since otherwise they will exit, but allows for the more productive firms to make positive profits; and the second free-entry condition states that the *expected value* of profits for new entrants must be zero. This second condition takes into account the fact that new firms must pay a fixed (or sunk) cost to enter the industry before they know what their future productivity will be; after paying that sunk cost they will exit if their productivity is too low. The expected value of profits is then computed over all productivity-draws that lead to production-positive per-period profits and draws that lead to exit and zero per-period profits, so that the expected value after subtracting the sunk costs of entry must be zero. The use of two zero-profit conditions will allow us to determine both the mass of firms that enter the industry and the subset of these that successfully produce (with a smaller subset successfully exporting using a third zero-profit condition, as we shall see). Our solution to these zero-profit conditions follows the exposition in Melitz and Redding (2014a), to which the reader is referred to generalize the Melitz model in many respects.

Just as the first generation of monopolistic competition models naturally led to a gravity equation, the same is true for the second generation. This point was first recognized by Chaney (2008), who derived a gravity equation from the Melitz model by adding the assumption that the distribution of productivities is Pareto: this framework is sometimes called the Melitz-Chaney model. The Pareto assumption helps us measure the gains from trade. We draw on the recent work of Arkolakis, Costinot, and Rodríguez-Clare (2012; hereafter ACR 2012) to derive the gains from trade in the Melitz-Chaney model, and link this derivation back to the analysis of Feenstra (1994) in the previous chapter. We show that the *sources* of gains from trade are quite different when there are homogeneous firms (as in the last chapter) versus heterogeneous firms (as in this chapter), despite the fact that the formula for these gains looks similar, as shown by ACR (2012). Melitz and Redding (2014a) argue that the gains are higher when there are heterogeneous firms, as we will discuss.

We complete the chapter by giving a general derivation of the gravity equation, drawing on Head and Mayer (2014). This derivation includes the gravity equation with homogeneous firms (as in the previous chapter), with heterogeneous firms (as in this chapter), and a third quite different justification for the gravity equation, which is the model of Eaton and Kortum (2002). Their model is an extension of Dornbusch, Fischer, and Samuelson (1977), discussed in chapter 3, with the added assumption that countries receive random draws of productivity for each of the goods in the continuum. For a particular assumption on the distribution of productivities—the Fréchet distribution—they find that the form of the gravity equation is very similar to that obtained from the Melitz-Chaney model. While our presentation of the Eaton-Kortum model is brief, it should be recognized that this model is an alternative framework to obtain many other results similar to the Melitz-Chaney model.

MONOPOLISTIC COMPETITION WITH HETEROGENEOUS FIRMS

In the chapter 5 discussion of the monopolistic competition model, each firm had a fixed cost of a and a marginal labor cost of β . We will continue to assume that labor is the only input, but now suppose that there are two fixed costs: one for domestic production, which we denote by f_d , and another for export production, which we denote by f_x . In addition, we allow firms to differ in their productivities. Following Melitz (2003), we denote the productivity of each firm by φ so that the marginal labor cost becomes $1/\varphi$, which we use instead of β . In the industry being considered there is a continuum of firms with varying productivity, and we suppose that the probability that any firm has a particular productivity φ is given by the density function $g(\varphi)$ with cumulative density $G(\varphi)$.

With a continuum of firms in the industry, we also want to allow for a continuum of differentiated products in the CES utility function (5.10), which is written instead as:

$$U = \left[\int_{\Omega} c(\omega)^{(\sigma-1)/\sigma} d\omega \right]^{\sigma/(\sigma-1)},$$

where we use $\omega \in \Omega$ to index the set of available products rather than $i=1, \dots, N$, as in (5.10). With each consumer earning the wage of w , the demand for each product is, as in (5.24)–(5.25),

$$c(\omega) = \left[\frac{p(\omega)}{P} \right]^{-\sigma} \frac{w}{P}, \quad (6.1)$$

where

$$P \equiv \left[\int_{\Omega} p(\omega)^{1-\sigma} d\omega \right]^{1/(1-\sigma)}, \quad (6.2)$$

which denotes the theoretical price index for the CES demand system. The elasticity of demand for each product is $-\partial \ln c(\omega) / \partial \ln p(\omega) = \sigma$.

Autarky Equilibrium

Initially we examine the autarky equilibrium, so that all firms have the fixed cost of f_d for selling in the domestic market. It will often be convenient to index variables by productivity φ rather than the generic index ω . Since marginal cost is (w/φ) , the markup of price over marginal cost is $p(\varphi)/(w/\varphi) = [\sigma/(\sigma-1)]$. For convenience, we will normalize the wage at unity, $w=1$. It follows that profits from the domestic market are written as

$$\pi_a(\varphi) = p(\varphi)y(\varphi) - \left(f_d + \frac{y(\varphi)}{\varphi} \right) = \frac{y(\varphi)}{\varphi(\sigma-1)} - f_d, \quad (6.3)$$

where $f_d + [y(\varphi)/\varphi]$ is the labor needed to produce output of $y(\varphi)$, and therefore total costs with a wage of unity. From (6.1), total demand from the home population of L consumers is $Lc(\varphi) = Lp(\varphi)^{-\sigma}/P_a^{1-\sigma}$, where P_a is the autarky price index. Market

clearing requires that $y(\varphi) = Lc(\varphi)$ under autarky. Using the prices $p(\varphi) = \sigma/[\varphi(\sigma - 1)]$, it follows that output is

$$y(\varphi) = Lc(\varphi) = \frac{L}{P_a^{1-\sigma}} \left[\frac{\sigma}{\varphi(\sigma - 1)} \right]^{-\sigma}.$$

Substituting this into the expression for profits, we obtain

$$\pi_a(\varphi) = \underbrace{\left[\frac{L\sigma^{-\sigma}}{P_a^{1-\sigma}(\sigma-1)^{1-\sigma}} \right]}_{B_a} \varphi^{\sigma-1} - f_d.$$

To summarize, using the $MR = MC$ condition to get the optimal prices for firms, and using market clearing in autarky, we obtain the above expression for profits. Since $\sigma > 1$, it is clear that this expression is increasing in the productivity of the firm, φ . Indeed, if we use the transformation of productivity $\varphi^{\sigma-1}$, then profits are linearly increasing in $\varphi^{\sigma-1}$, as illustrated in figure 6.1. For convenience in this figure we divide profits by the wage, so that profits evaluated at the lowest productivity of $\varphi^{\sigma-1} = 0$ are $-f_d$, and then rise with slope equal to B_a . We have labeled the productivity where this profits line crossed the x-axis as $\varphi_a^{\sigma-1}$ in figure 6.1, which is solved from the *zero-cutoff-profit* (ZCP) condition as

$$\text{ZCP: } \pi_a(\varphi_a) = B_a \varphi_a^{\sigma-1} - f_d = 0, \quad (6.4)$$

$$\Rightarrow \varphi_a^{\sigma-1} = \frac{f_d}{B_a} = \frac{f_d \sigma^\sigma}{L P_a^{\sigma-1} (\sigma-1)^{\sigma-1}}. \quad (6.5)$$

We refer to φ_a as the ZCP level of productivity. Firms with higher productivity earn strictly positive profits, and firms with lower productivity would earn negative profits so they do not produce and exit the market.

We see from (6.4) that the ZCP level of productivity φ_a depends on the CES price index P , which in turn depends on the range of available products. So the ZCP condition illustrated in figure 6.1 does not really pin down the level of productivity φ_a in general equilibrium: we will need an additional condition to determine the range of available products. This is where the monopolistic competition model with heterogeneous firms differs from Krugman's simpler model with homogeneous firms. With heterogeneous firms there are *two* zero-profit conditions: the first is the ZCP condition in (6.4), and the second is the *free entry* (FE) condition, which states that the expected profits from entering the market must be zero in equilibrium. Let us now develop this free entry condition.

We suppose that firms entering this market must pay a fixed cost of f_e in order to receive a draw of productivity φ from the density function $g(\varphi)$. We often refer to f_e as a "sunk cost" since it is paid just once at the beginning, whereas the fixed costs f_d (and f_x for exporting, which we have not yet discussed) are paid every period provided that the firm continues to produce. The firm will not produce whenever its productivity draw is below φ_a , so in that case profits are equal to zero. Then in a free entry (FE) equilibrium, the expected value of profits must just offset the costs of entry f_e , which means that

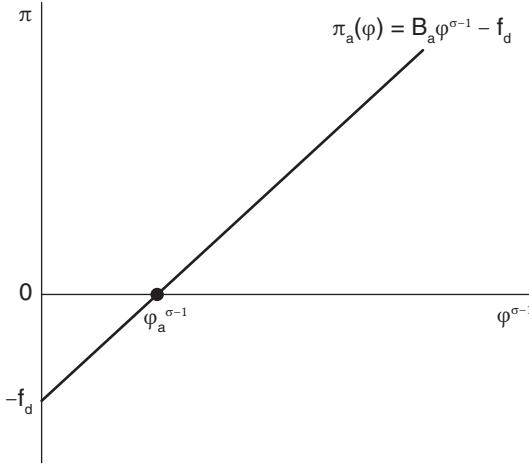


Figure 6.1

$$\text{FE: } \int_0^\infty \pi_a(\varphi) g(\varphi) d\varphi = \int_{\varphi_a}^\infty (B_a \varphi_a^{\sigma-1} - f_d) g(\varphi) d\varphi = f_e. \quad (6.6)$$

To combine the ZCP and FE conditions, we express profits with productivity $\varphi \geq \varphi_a$ as

$$\pi_a(\varphi) = B_a \varphi^{\sigma-1} - f_d = (\varphi/\varphi_a)^{\sigma-1} B_a \varphi_a^{\sigma-1} - f_d = [(\varphi/\varphi_a)^{\sigma-1} - 1] f_d,$$

where the last equality is obtained from ZCP in (6.4). Making use of this relation in (6.6) and dividing by the wage, we can solve for the expected value of real profits as

$$\begin{aligned} f_e &= \int_0^\infty \pi_a(\varphi) g(\varphi) d\varphi = \int_{\varphi_a}^\infty (B_a \varphi^{\sigma-1} - f_d) g(\varphi) d\varphi \\ &= \int_{\varphi_a}^\infty [(\varphi/\varphi_a)^{\sigma-1} f_d - f_d] g(\varphi) d\varphi \\ &= J(\varphi_a) f_d, \quad \text{for } J(\varphi_a) \equiv \int_{\varphi_a}^\infty [(\varphi/\varphi_a)^{\sigma-1} - 1] g(\varphi) d\varphi. \end{aligned} \quad (6.7)$$

The function $J(\varphi_a) > 0$ is monotonically decreasing in φ_a .² Then (6.7) proves that there is a unique value of φ_a satisfying the FE and ZCP conditions. This is the value of productivity shown in figure 6.1, and for draws below this amount the firm exits the market.

Having determined the cutoff productivity φ_a , all other variables in the model are quickly solved for. In particular, from (6.4) the shift parameter B_a governing demand must equal $B_a = f_d/\varphi_a^{\sigma-1}$. Conditional on entry with productivity above φ_a , the

²See exercise 6.1. Melitz and Redding (2014a) introduce this function to solve for the ZCP productivity.

conditional density function for remaining producers is $g(\varphi)/[1 - G(\varphi_a)]$. This conditional density is used to define the expected profits $\bar{\pi}(\varphi_a)$ from drawing a productivity above the cutoff φ_a :

$$\bar{\pi}(\varphi_a) \equiv \int_{\varphi_a}^{\infty} \pi_a(\varphi) \frac{g(\varphi)}{[1 - G(\varphi_a)]} d\varphi = \frac{f_e}{[1 - G(\varphi_a)]}, \quad (6.8)$$

where the equality follows from (6.6) and $\pi_a(\varphi) = 0$ for exiting firms with productivity $\varphi < \varphi_a$. With these solutions for the cutoff productivity φ_a and expected profits, the mass of available products can be determined from the full-employment condition, much like in Krugman's (1980) model. We conclude this section with a brief discussion of how to solve for that number, and then turn to the trade equilibrium in the next section.

Denote the mass of entering firms who pay the sunk cost of f_e each period by M_e . Of these, only the fraction $[1 - G(\varphi_a)]$ obtain a successful draw $\varphi \geq \varphi_a$ and remain in the market. It follows that the mass of surviving firms in autarky is

$$M_a = M_e [1 - G(\varphi_a)].$$

Turning to the full employment condition, it is expressed as

$$\begin{aligned} L &= M_a \int_{\varphi_a}^{\infty} \left[\frac{y(\varphi)}{\varphi} + f_d \right] \frac{g(\varphi)}{[1 - G(\varphi_a)]} d\varphi + M_e f_e \\ &= M_a \int_{\varphi_a}^{\infty} [(\sigma - 1) \pi_a(\varphi) + \sigma f_d] \frac{g(\varphi)}{[1 - G(\varphi_a)]} d\varphi + \frac{M_a f_e}{[1 - G(\varphi_a)]} \\ &= M_a [(\sigma - 1) \bar{\pi}(\varphi_a) + \sigma f_d] + M_a \bar{\pi}(\varphi_a) \\ &= M_a \sigma [\bar{\pi}(\varphi_a) + f_d], \end{aligned} \quad (6.9)$$

where the first line sums together the variable and fixed costs used to produce output, together with the fixed costs of entry; the second line uses (6.3) to rewrite the variable and fixed cost of production in terms of profits plus fixed costs, and also rewrite the fixed costs of entry using the condition $M_a = M_e [1 - G(\varphi_a)]$, the third line evaluates the integral as expected profits and uses the FE condition to simplify the last term; and then the fourth line simplifies. Given that we have solved for the average profits $\bar{\pi}(\varphi_a)$ in (6.8), it is immediate that this full employment condition determines the mass of products M_a as

$$M_a = \frac{L}{\sigma [\bar{\pi}(\varphi_a) + f_d]} = \frac{L}{\sigma \left(\frac{f_e}{[1 - G(\varphi_a)]} + f_d \right)}. \quad (6.10)$$

Trade Equilibrium

We now allow for international trade between two countries. For simplicity, we assume that these countries are identical, which is the same simplification that we used in the Krugman (1979) model. We continue to normalize the wage in both countries at unity. Unlike the Krugman model, however, we cannot treat the opening of trade as

identical to an increase in country size. The reason this trick does not work now is that we assume firms have *another* fixed cost of f_x that must be paid in order to export. This extra cost can reflect the difficulty of learning about foreign customers and setting up a distribution network abroad. In addition, we assume that firms face trade barriers when exporting, which we model as “iceberg” trade costs introduced by Samuelson (1952): in order for one unit to arrive abroad, there must be $\tau > 1$ units produced and sent, where the remainder ($\tau - 1$) “melts” along the way.

In this framework, there will still be a zero-profit-cutoff φ_d for firms serving the domestic market, where firms with lower productivity earn negative profits and exit, whereas firms with slightly higher productivity sell at home but do not find it profitable to export. There will be a second zero-profit-cutoff φ_x for firms that also find it profitable to export. Under plausible conditions we will find that $\varphi_x > \varphi_d$, so that only the more productive firms export, which is consistent with a large amount of empirical evidence.³

The domestic zero-profit condition is still given by (6.4), rewritten as

$$\text{ZCP}_d: \pi_d(\varphi_d) = B_d \varphi_d^{\sigma-1} - f_d = 0, \quad (6.4')$$

$$\Rightarrow \varphi_d^{\sigma-1} = \frac{f_d}{B_d} = \frac{f_d \sigma^\sigma}{L P_d^{\sigma-1} (\sigma-1)^{\sigma-1}}, \quad (6.5')$$

where the term $B_d \equiv L \sigma^{-\sigma} / P_d^{1-\sigma} (\sigma-1)^{1-\sigma}$ will differ from autarky due to a new value of the CES price index P_d obtained *in the presence* of trade. We will be more explicit about this price index in the next section. For now, we will presume that there are gains from trade which means that the cost of living is reduced, so that $P_d < P_a$. It follows from (6.5) and (6.5') that $\varphi_d > \varphi_a$, as we shall prove more formally below.

In the export market, the marginal cost of selling one unit abroad is (τ/φ) , so the firm charges an optimal price of $p_x(\varphi) = [\sigma/(\sigma-1)](\tau/\sigma)$. It follows that profits from the export market are

$$\pi_x(\varphi) = p_x(\varphi) y_x(\varphi) - \left(f_x + \frac{\tau y_x(\varphi)}{\varphi} \right) = \frac{\tau y_x(\varphi)}{\varphi(\sigma-1)} - f_x,$$

where $f_x + [\tau y_x(\varphi)/\varphi]$ is the labor needed to deliver export of $y_x(\varphi)$. Export demand from the foreign population of L consumers is $y_x(\varphi) = L p_x(\varphi)^{-\sigma} / P_d^{1-\sigma}$, where P_d is the same CES price index as at home due to identical countries. Using the solution for export price, foreign demand for exports is

$$y_x(\varphi) = L c_x(\varphi) = \frac{L}{P_d^{1-\sigma}} \left[\frac{\tau \sigma}{\varphi(\sigma-1)} \right]^{-\sigma}.$$

Profits in the export market are then

$$\pi_x(\varphi) = \underbrace{\left[\frac{L \tau^{1-\sigma} \sigma^{-\sigma}}{P_d^{1-\sigma} (\sigma-1)^{1-\sigma}} \right]}_{B_x} \varphi^{\sigma-1} - f_x.$$

³See, for example, the evidence presented in Bernard, Jensen, Redding, and Schott (2007).

It follows that the ZCP condition in the export market is

$$\text{ZCP}_x: \pi_x(\varphi_x) = B_x \varphi_x^{\sigma-1} - f_x = 0, \quad (6.11)$$

$$\Rightarrow \varphi_x^{\sigma-1} = \frac{f_x}{B_x} = \frac{f_x \tau^{\sigma-1} \sigma^\sigma}{L P_d^{\sigma-1} (\sigma-1)^{\sigma-1}}. \quad (6.12)$$

Dividing (6.12) by (6.5'), we immediately obtain

$$\left(\frac{\varphi_x}{\varphi_d} \right)^{\sigma-1} = \frac{f_x \tau^{\sigma-1}}{f_d}. \quad (6.13)$$

It follows that a sufficient condition to have $\varphi_x > \varphi_d$ is that $f_x \tau^{\sigma-1} > f_d$, as we shall assume. In this case, only a *subset* of the producing domestic firms—those that receive a productivity draw above φ_d —will also have productivity above φ_x and will therefore export. It follows that the profit line for exporters in figure 6.2 intersects the x-axis at a higher productivity, regardless of whether its intercept f_x is above or below f_d .

To pin down the cutoff productivities, we again use the FE condition, now extended to allow for both domestic production and exporting:

$$\begin{aligned} \text{FE: } & \int_0^\infty [\pi_d(\varphi) + \pi_x(\varphi)] g(\varphi) d\varphi \\ &= \int_{\varphi_d}^\infty (B_d \varphi^{\sigma-1} - f_d) g(\varphi) d\varphi + \int_{\varphi_x}^\infty (B_x \varphi^{\sigma-1} - f_x) g(\varphi) d\varphi = f_e. \end{aligned}$$

We can simplify this expression in the same manner as in the autarky equilibrium, to obtain

$$J(\varphi_d)f_d + J(\varphi_x)f_x = f_e, \quad (6.14)$$

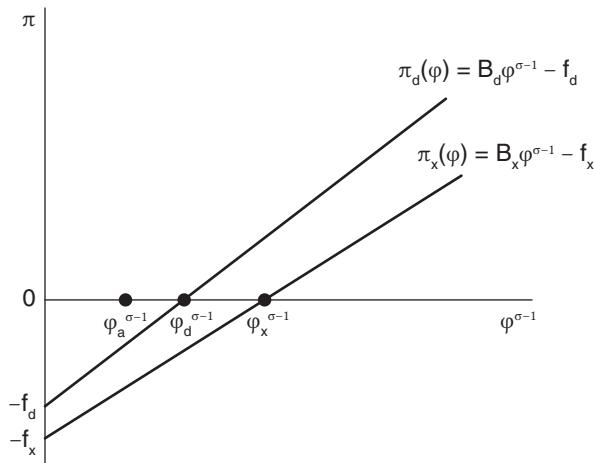


Figure 6.2

for the function $J(\cdot)$ defined in (6.7). Since this function is monotonically decreasing and we know that $\varphi_x > \varphi_d$, it is immediately apparent from comparing (6.7) and (6.14) that we must have $\varphi_a < \varphi_d < \varphi_x$. Therefore, the cutoff productivity for firms selling to the domestic market *rises* with trade, and the cutoff productivity for firms exporting is higher yet. These cutoffs are illustrated in figure 6.2. The finding that that $\varphi_d > \varphi_a$ proves that $P_d < P_a$, as we discussed above when comparing (6.5) and (6.5'). The welfare of the representative consumer is measured by real wages w/P , and normalizing $w=1$ as we have done, it follows that welfare goes up due to trade. The remaining question is the *sources* of the gains from trade, which we turn to now.

THE MELITZ-CHANAY MODEL

Chaney (2008) adds to the Melitz model the assumption that the distribution of productivities across firms follows a Pareto distribution:

$$G(\varphi) = 1 - \varphi^{-\theta} \quad \text{for } \varphi \geq 1, \text{ and } \theta > \sigma - 1 > 0. \quad (6.15)$$

Using this distribution allows us to obtain a closed-form solution to many of the expressions in the Melitz model. For example, consider the function J defined in (6.7). Using the Pareto distribution, we can evaluate this integral as⁴

$$J(\varphi_a) = \left(\frac{\sigma-1}{\theta-\sigma+1} \right) [1 - G(\varphi_a)].$$

Since $f_e = J(\varphi_a)f_d$, we can substitute this expression in (6.10) to find that

$$M_a = \frac{L}{\sigma \left(\frac{f_e}{[1 - G(\varphi_a)]} + f_d \right)} = \left(\frac{\theta-\sigma+1}{\sigma\theta} \right) \frac{L}{f_d}.$$

Thus, the mass of products available to consumers in the autarky economy is proportional to the size of the country, much as we found for the CES version of the Krugman model in chapter 5. For the Melitz-Chaney model, the mass of *entering* firms is $M_e = M_a/[1 - G(\varphi_a)]$, and so

$$M_e = \left(\frac{\theta-\sigma+1}{\sigma\theta} \right) \frac{L}{f_d [1 - G(\varphi_a)]} = \left(\frac{\sigma-1}{\sigma\theta} \right) \frac{L}{f_e}, \quad (6.16)$$

where we again use $f_e = J(\varphi_a)f_d$. So the mass of entering firms under autarky is also proportional to the country size.

Once we introduce trade, it is again possible to solve for the mass of entering firms and surviving firms using the full-employment condition. It turns out that the solution for the mass of entering firms M_e shown in (6.16) continues to hold: this mass is pinned down by the full-employment condition and does not change due to trade.⁵ The assumption of a fixed mass of entrants was made by Chaney (2008) for simplicity,

⁴See exercise 6.1.

⁵See problem 6.4 and the solution in Feenstra (2010c, appendix).

and is therefore an innocuous assumption. The mass of surviving domestic firms when trade is opened is $M_d = M_e [1 - G(\varphi_a)]$, and because the cutoff productivity rises it follows that the mass of domestic surviving firms falls due to trade. This model, therefore, addresses the criticism of the Krugman (1980) model with CES preferences that we studied in the previous chapter, where the number of domestic varieties did not change when trade was opened. Now, in the Melitz-Chaney model, there is a fall in the mass of domestic varieties (which is the same as the mass of surviving domestic firms). To evaluate welfare, this fall in domestic varieties needs to be compared with the increase in import variety, and also compared with the welfare gains from rising *average* productivity as the least-productive domestic firms exit. We will explore the various sources of the gains from trade next.

The Gains from Trade

The CES price index facing consumers is written in general terms in (6.2). Because firms earn zero expected profits by the FE condition, there are no profits redistributed to consumers. So the income of each consumer equals the wage w , and their real income equal w/P . Normalizing the wage at unity, any reduction in price index P therefore corresponds to a rise in utility.

By making the change in variables from the generic index ω to the productivity φ , the CES price index becomes

$$P_d = \left[M_d \int_{\varphi_d}^{\infty} p(\varphi)^{1-\sigma} \frac{g(\varphi)}{[1-G(\varphi_d)]} d\varphi + M_x \int_{\varphi_x}^{\infty} p_x(\varphi)^{1-\sigma} \frac{g(\varphi)}{[1-G(\varphi_x)]} d\varphi \right]^{1/(1-\sigma)}, \quad (6.17)$$

where the first integral reflects the prices of the mass M_d of domestic firms with productivity $\varphi \geq \varphi_d$, and the second integral reflects the export prices of the mass M_x of *foreign* firms with productivity $\varphi \geq \varphi_x$. Recall that we are assuming trade between two identical countries, so these foreign export prices are written in the same manner as home import prices, and the mass of exporting firms is $M_x = M_e [1 - G(\varphi_x)]$. We have already argued that the price index falls as compared to autarky, so that there are overall gains from trade. Now we want to identify the sources of these gains using the various terms in this price index.

There are three potential effects of trade. First, the presence of the second integral in the price index will raise the overall term in square brackets as compared to its autarky value, and because it is raised to a negative power, it therefore lowers the price index and raises welfare. This is the positive effect of *import variety* on welfare. On the other hand, the mass of surviving domestic firms equals $M_d = M_e [1 - G(\varphi_d)]$, and we know that the productivity cutoff for domestic firms φ_d exceeds its autarky φ_a , which means that some firms have exited. With fixed entry M_e this leads to a fall in the mass of products M_d , which lowers the overall term in square brackets and therefore raises the price index. This is the negative effect of *reduced domestic variety* on welfare. The reason this offsetting effect occurs is that import competition lowers demand for domestic firms and forces the least productive of these to exit.⁶ But that exit corresponds to a *positive selection effect of trade*, which lowers the average domestic price through

⁶Melitz (2003, 1715–16) favors an alternative explanation for why the least productive firms exist. Namely, the rise in labor demand from firms that now export bids up the real wage, forcing the least productive firms to exit.

raising φ_d in the first integral and therefore raises welfare. We will argue that with the Pareto distribution, this third effect is the *only* explanation for the gains from trade.

In order to solve for the average domestic price, we follow Melitz (2003) in specifying $\bar{\varphi}$ as *average* productivity:

$$\bar{\varphi}_d \equiv \left[\int_{\varphi_d}^{\infty} \varphi^{(\sigma-1)} \frac{g(\varphi)}{[1-G(\varphi_d)]} d\varphi \right]^{1/(\sigma-1)}. \quad (6.18)$$

We note that a property of the Pareto distribution is that an integral like (6.18) is always a constant multiple of the lower bound of integration,

$$\bar{\varphi}_d = \left(\frac{\theta}{\theta - \sigma + 1} \right)^{-1/(\sigma-1)} \varphi_d, \quad (6.19)$$

as obtained by evaluating the integral in (6.18). The average prices of domestic goods appearing in (6.17) are then

$$\left[M_d \int_{\varphi_d}^{\infty} p(\varphi)^{1-\sigma} \frac{g(\varphi)}{[1-G(\varphi_d)]} d\varphi \right]^{\frac{1}{1-\sigma}} = \left(\frac{\sigma}{\sigma-1} \right) \left(\frac{1}{\bar{\varphi}_d} \right) M_d^{\frac{1}{1-\sigma}}, \quad (6.20)$$

which uses the prices $p(\varphi) = [\sigma/(\sigma-1)](1/\varphi)$ together with the definition of average productivity in (6.18).

When comparing autarky with free trade, we need to take into account the changing price of domestic goods and their changing variety, as in (6.20), along with the fact that all imported goods are new. The ratio of CES price indexes P_d/P_a is what we measured using an “exact” price index in the theorem due to Feenstra (1994) in chapter 5. We now apply that theorem using (6.20) to measure the effective price of domestic goods, continuing to normalize $w^* = 1$. We treat all imported products as new, and let λ_d denote the share of expenditure on domestic goods in the trade equilibrium. Then the chapter 5 theorem due to Feenstra (1994) leads to the following price index comparing trade with autarky:

$$\frac{P_d}{P_a} = \left(\frac{M_d^{\frac{1}{1-\sigma}}/\bar{\varphi}_d}{M_a^{\frac{1}{1-\sigma}}/\bar{\varphi}_a} \right) \lambda_d^{\frac{1}{\sigma-1}}. \quad (6.21)$$

The first term appearing on the right of (6.21) is just the change in the average price of domestic goods, using (6.20). The aggregate domestic good is available in both periods, so this first term reflects the Sato-Vartia index P_{sv} over the “common” good in the chapter 5 theorem. The second term on the right is the spending on domestic goods relative to total spending in the trade equilibrium, or one minus the share of spending on new imported varieties. This term reflects that potential gains due to new import varieties. As we have already argued, those gains may be offset due to a reduced number of domestic varieties. Those offsetting forces can be seen more easily by rewriting (6.21) slightly as

$$\frac{P_d}{P_a} = \left(\frac{1/\bar{\varphi}_d}{1/\bar{\varphi}_a} \right) \left(\frac{\lambda_d}{M_d/M_a} \right)^{\frac{1}{\sigma-1}}. \quad (6.21')$$

This formulation neatly captures the three potential sources of gains from trade in the Melitz model: the selection effect, which is a rise in average productivity and is shown by the first term on the right; the gains due to import variety, which is shown by the numerator of the second term; and the losses due to reduced domestic variety, $M_d < M_a$, which is shown by the denominator of the second term on the right.

We now show that $M_d/M_a = \lambda_d$ in (6.21'), so the welfare loss due to the reduction in the number of domestic varieties just cancels with the welfare gain due to new import varieties. This result is obtained from dividing the cutoff productivities for domestic firms from (6.5) and (6.5'):

$$\left(\frac{\varphi_d}{\varphi_a} \right) = \left(\frac{1/P_d}{1/P_a} \right). \quad (6.22)$$

As we have already argued, the change in the cutoff productivities closely matches the fall in the prices indexes and rise in welfare. Once we add the Pareto assumption, the ratio of the cutoff productivities equals the ratio of the *average* productivities $\bar{\varphi}_d/\bar{\varphi}_a$, as shown in (6.19). Then comparing (6.21) with (6.22), we immediately see that that $M_d/M_a = \lambda_d$, as we intended to show. It follows that

$$\lambda_d = \left(\frac{M_d}{M_a} \right) = \left(\frac{M_e}{M_a} \right) \left(\frac{1 - G(\varphi_d)}{1 - G(\varphi_a)} \right) = \left(\frac{\varphi_d}{\varphi_a} \right)^{-\theta} \Rightarrow \left(\frac{\varphi_d}{\varphi_a} \right) = \lambda_d^{-1/\theta}.$$

The second equality above follows from using the formulas for the mass of surviving firms, and then the third equality follows because the mass of entering firms is equal in the autarky and trade equilibria, and also using the Pareto distribution. We find, therefore, that we obtain a particularly simple formula for the ratio of cutoff productivity by using λ_d —the share of expenditure going to domestic products—raised to negative, inverse of the Pareto parameter. But since the cutoff productivities determine the CES price indexes from (6.22), we have obtained a very simple formula for the gains from trade that is also found by ACR (2012).

THEOREM (ARKOLAKIS, COSTINOT, AND RODRÍGUEZ-CLARE 2012)

Normalizing the wage at unity and comparing autarky with free trade, the rise in welfare for the representative consumer in the Melitz-Chaney model equals

$$\left(\frac{1/P_d}{1/P_a} \right) = \lambda_d^{-1/\theta}, \quad (6.23)$$

where λ_d is the share of expenditure going to domestic products in the trade equilibrium.

As noted above, the fact that $M_d/M_a = \lambda_d$ in (6.21') means that the welfare loss due to the reduction in the number of domestic varieties just cancels with the welfare gain due to new import varieties. So while the change in the *mass* of consumed varieties can either increase or decrease from M_a to $M_d + M_x$,⁷ this change in variety has *zero*

⁷See problem 6.3, where it is shown that $M_a > (<) M_d + M_x$ as $f_x > (<) f_d$. In a model with quadratic rather than CES preferences, Baldwin and Forslid (2010) argue that the total number of product varieties falls with trade liberalization, but that is because they assume that $f_x > f_d$.

welfare impact. Rather, we have shown that the only source of gains from trade in the Melitz-Chaney model becomes the selection effect, whereby exiting domestic firms lead to a rise in average productivity.

Notice that these effects from the Melitz model are very different from the Krugman (1980) model with CES preferences, that we discussed in chapter 5, where neither the scale nor the number of domestic firms changes at all when trade is opened. The difference in results between the Krugman (1980) CES model and the Melitz-Chaney model shows the benefit of allowing for heterogeneous firms, and is why most monopolistic competition research in international trade today will allow for heterogeneous firms.

These results from the Melitz-Chaney model obviously challenge the empirical findings of Broda and Weinstein (2006), discussed in the previous chapter, who held constant the number of domestic varieties as in the Krugman (1980) model. Now that domestic varieties are being reduced in the Melitz-Chaney model, how do the gains from trade compare with what we found from the Krugman model with CES preferences? To answer this question, we use the gains from trade shown in (5.18) in the chapter 5, but evaluated with a single sector (so $\alpha_n = 1$), and change the general industry subscript n to the domestic sector d . The gains from trade in the Krugman (1980) model are then $\lambda_d^{-1/(\sigma-1)}$. How do these gains compare with $\lambda_d^{-1/0}$, as shown in (6.23)? Throughout our analysis, we have relied on the assumption that the Pareto parameter exceeds the elasticity of substitution minus one, $\theta > \sigma - 1 > 0$. Without this assumption, many of the integrals that we have written above would be infinite. Because the share of goods purchased from domestic sources is less than unity, $\lambda_d < 1$, it follows directly from this assumption that $\lambda_d^{-1/0} < \lambda_d^{-1/(\sigma-1)}$. So does this mean that the gains from trade in the Melitz-Chaney model are less than those in the Krugman (1980) model with CES preferences?

The answer is a resounding *no*. In fact, the gains from trade in the Melitz-Chaney model, with heterogeneous firms, will generally exceed those in the Krugman (1980) model, with homogeneous firms, as shown by Melitz and Redding (2014a). The reason that our logic above does not hold is that we were keeping the share of goods purchased from domestic sources, λ_d , fixed when we compare the two models. Melitz and Redding instead compare a heterogeneous firm model to a homogeneous firm model that starts with the same average productivity in autarky. They show that the heterogeneous firm model allows for greater expansion of output of the high-productivity firms when trade is opened, corresponding to higher exports and also imports (with balanced trade), and therefore a *lower value* of λ_d , than in the homogeneous firm model. This lower value of λ_d is enough to counteract the parameter restriction $\theta > \sigma - 1 > 0$, so that the gains from trade in the heterogeneous firm model, $\lambda_d^{-1/0}$, exceed $\lambda_d^{-1/(\sigma-1)}$ in the homogeneous firm model, when both are evaluated at the equilibrium value of λ_d .

This result from Melitz and Redding shows that, in fact, the heterogeneous firm model allows for greater gains from trade due to increased specialization: the most-productive firms become major exporters while the least-productive firms exit the market, with imports correspondingly higher than with homogeneous firms.⁸ It reassures us that the gains from trade from the Krugman (1980) model, computed by Broda and Weinstein (2006) and Ossa (2012) and reported in table 5.1, would be even higher for a heterogeneous firm model. But there is still one nagging question that we might ask. Instead of comparing two models, as Melitz and Redding do, what if we

⁸Balistreri, Hillberry, and Rutherford (2010) analogously argue that with the addition of a second sector, the gains from trade in the Melitz (2003) model are quite different from those in the Krugman model.

simply take the domestic share λ_d as given from the data? In that case, are we forced to the conclusion that $\lambda_d^{-1/\theta} < \lambda_d^{-1/(\sigma-1)}$, so that the gains are lower in the heterogeneous firm model, because of the parameter restriction that $\theta > \sigma - 1$?

Once again, the answer is *no*, for a subtle reason. Simonovska and Waugh (2014a) present a gravity-based method to estimate the parameters of the gains from trade formula: either the elasticity of substitution σ , or the Pareto parameter θ , or a Fréchet parameter (discussed later in this chapter). Estimating each parameter from a gravity equation corresponds to a different model: the homogeneous firm model for the gravity equation with σ appearing on trade costs (as we discussed in chapter 5); the heterogeneous firm model for the gravity equation with θ appearing on trade costs (as we shall discuss in the next section); or the model due to Eaton and Kortum (2002) for a gravity equation with the Fréchet parameter (as we also discuss in the next section). Simonovska and Waugh (2014b) show that when estimating each of these models on the same data, there is a systematic relationship between parameters, with the estimate of the Pareto parameter being *less than* the elasticity of substitution minus one, $\theta < \sigma - 1$. It does not matter that this systematic pattern in the estimates contradicts the theoretical assumption that $\theta > \sigma - 1$, because the parameters estimates are obtained by estimating *different models* on the same data, whereas the restriction $\theta > \sigma - 1$ applies only to the Melitz-Chaney model. Given their empirical finding that $\theta < \sigma - 1$, and with the domestic share $\lambda_d < 1$ given from the data, we therefore conclude once again that $\lambda_d^{-1/\theta} > \lambda_d^{-1/(\sigma-1)}$, so the gains from trade in the heterogeneous firm model exceed those in the homogeneous firm model.

THE GRAVITY EQUATION

Our discussion above indicates how important the gravity equation is in estimating the exponent in the gains from trade formula, and this is the general result from ACR (2012): over a wide class of models, it is the coefficient of trade costs on the gravity equation that governs the exponent in the gains from trade formula. We now provide a quite general derivation of the gravity equation, drawing upon Head and Mayer (2014). They define the “special” gravity equation as an equation of the type

$$X^{ij} = \frac{Y^i}{R^{i*}} \frac{X^j}{R^j} T^{ij}, \quad (6.24)$$

where X^{ij} is the value of export from country i to country j (including X^{ii} which are country i 's sales to itself); $Y^i = \sum_j X^{ij}$ is the total value of production in country i ; $X^j = \sum_i X^{ij}$ is total expenditure in country j ; T^{ij} is a measure of trade barriers between country i and j (reflecting iceberg transport costs τ^{ij} along with its coefficient and any other factors); and the “multilateral resistance” terms are defined as

$$\underbrace{R^j}_{\text{Multilateral Resistance for } j} = \sum_k \frac{T^{kj} Y^k}{R^{k*}}, \quad \text{and} \quad \underbrace{R^{i*}}_{\text{Multilateral Resistance for } i} = \sum_k \frac{T^{ik} X^k}{R^k}. \quad (6.25)$$

The multilateral resistance term for *importing* country j is the weighted sum over the trade barriers T^{kj} that country j has with each source country k , where the weights reflect the production (and therefore potential exports) of country k relative to its own

multilateral resistance. The multilateral resistance term for *exporting* country i is the weighted sum of the trade barriers charged by each destination market k , where the weights reflect the expenditure (and therefore potential imports) of country k relative to its own multilateral resistance. In brief, the importer multilateral resistance R^j is a measure of the trade barriers that country j has when *buying* from its source countries k (hence the superscripts $\cdot j$, indicating that we are summing over source countries); whereas the exporter multilateral resistance $R^{i\cdot}$ reflects the trade barriers faced by country i for *selling* its destination markets k (hence the superscripts $i \cdot$, indicating that we are summing over destination countries).

The above equations are just definitions of a special form of the gravity equation, but take on more substantive meaning when we ask whether these definitions will in fact hold for trade between countries. Head and Mayer (2014) find that the conditions to ensure these special gravity equations are actually quite weak, as shown by the following result.

THEOREM (HEAD AND MAYER 2014)

The special form of the gravity equation in (6.24)–(6.25) will hold whenever there are positive “supply terms” S^i , depending only on data for country i , such that the value of exports from country i to j satisfies

$$\frac{X^{ij}}{X^j} = \frac{S^i T^{ij}}{R^j}, \text{ with } R^j = \sum_k S^k T^{kj}. \quad (6.26)$$

Proof:

Comparing the first equation in (6.26) with (6.24), we can define the remoteness term $R^{i\cdot}$ as $R^{i\cdot} \equiv Y^i/S^i$ so that (6.24) holds. Then substituting in $S^k = Y^k/R^{k\cdot}$, it follows that the second equation in (6.26) becomes $R^{i\cdot} = \sum_k (Y^k/R^{k\cdot}) T^{ik}$, so that the first equation in (6.25) holds. The second equation in (6.25) also holds by using the market clearing condition $Y^i = \sum_k X^{ik}$, which is rewritten by substituting in from (6.26), $X^{ik} = (S^i T^{ik}/R^{i\cdot}) X^k = (Y^i T^{ik}/R^{i\cdot} R^{k\cdot}) X^k$ using $S^i = Y^i/R^{i\cdot}$, so that $Y^i = \sum_k (Y^i T^{ik}/R^{i\cdot} R^{k\cdot}) X^k \Leftrightarrow R^{i\cdot} = \sum_k (T^{ik}/R^{k\cdot}) X^k$. QED

To interpret (6.26), the “supply terms” S^i reflect the supply capacity of country i , which is then sent to the importing countries j according to the trade barriers relative to overall multilateral resistance in j , T^{ij}/R^j . In addition, multilateral resistance in j is a weighted sum of its trade barriers on each exporter k , weighted by the supply capacity in k . Whenever these conditions hold, the special form of the gravity equation in (6.24)–(6.25) also holds. Head and Mayer (2014) further show how these supply terms can be identified in various models that we describe next, which then give rise to a gravity equation.⁹

⁹Head and Mayer (2014) obtain a gravity equation from additional models that we do not discuss, including a model with heterogeneous consumers that leads to logit demand.

Krugman (1980) Model

We start by showing how the Krugman (1980) CES model gives rise to the Anderson-Van Wincoop (2003) gravity equation that we already discussed in the previous chapter. From equations (5.25) and (5.26) and using $p^{ij} = \tau^{ij} p^i$ and balanced trade with $X^j = Y^j$, we have

$$\frac{X^{ij}}{X^j} = N^i \left(\frac{\tau^{ij} p^i}{P^j} \right)^{1-\sigma} \quad \text{and} \quad P^j = \left(\sum_{i=1}^C N^i (p^i \tau^{ij})^{1-\sigma} \right)^{1/(1-\sigma)}. \quad (6.27)$$

It is immediate that the conditions in (6.26) will hold by defining $T^{ij} \equiv (\tau^{ij})^{1-\sigma}$, $R^j \equiv (P^j)^{1-\sigma}$ and $S^i \equiv N^i (p^i)^{1-\sigma}$. It follows that the special form of gravity in (6.24) holds with balanced trade $X^j = Y^j$. Anderson and van Wincoop further assume symmetry in the trade costs $\tau^{ij} = \tau^{ji}$, from which it follows that the importer and exporter multilateral resistance terms are equal and tightly related to the CES price index, $R^j = R^i = (P^i)^{1-\sigma}$. We therefore obtain the gravity equation

$$X^{ij} = Y^i Y^j \left(\frac{\tau^{ij}}{P^i P^j} \right)^{1-\sigma}, \quad (6.27')$$

which is very close to the gravity equation (5.35) as used by Anderson and van Wincoop (2003): the only difference is that (5.35) also divides by world output Y^w , which is obtained by using a different normalization on the CES prices indexes in (5.35).¹⁰

Melitz-Chaney Model

Chaney (2008) has shown how the Melitz model can be used to obtain a gravity equation that is similar to what we obtained in homogeneous firm case, but with a different interpretation of the coefficients on trade barriers and other variables. Because of these different coefficients, Chaney refers to the gravity equation so obtained as “distorted” gravity.

Consumption in any country in the Melitz-Chaney model is as shown in (6.1), so that exports integrate that magnitude over the profitable firms using the price index shown in (6.17). Note that in the price index the mass of profitable domestic firms is $M_d = M_e [1 - G(\varphi_d)]$ and the mass of profitable exporters is $M_x = M_e [1 - G(\varphi_x)]$. So we can rewrite the CES price index (6.17) by integrating over the unconditional distribution $g(\varphi)$, and letting the mass of entrants M_e appear in front of the integrals. We will generalize our earlier exposition to allow for multiple countries, so that M_e^i are the entrants in country i and φ^{ij} is the zero-cutoff-profit (ZCP) value of productivity for selling to country j . We also allow the wages w^i to differ across countries. Then the value of exports from country i to j relative to total consumption $X^j = w^j L^j$ in country j is

¹⁰See note 25 of chapter 5. To understand this normalization, let us work from $R^i \equiv Y^i / S^i$ and $S^i \equiv N^i (p^i)^{1-\sigma}$ in the above proof of the Head-Mayer theorem, which we can substitute into the CES price index in (6.27) to obtain $(P^j)^{1-\sigma} = \sum_{i=1}^C N^i (p^i \tau^{ij})^{1-\sigma} = \sum_{i=1}^C (Y^i / R^i) (\tau^{ij})^{1-\sigma} = \sum_{i=1}^C Y^i (\tau^{ij} / P^i)^{1-\sigma}$, where the last equality uses $R^j = R^i = (P^i)^{1-\sigma}$. This recursive equation to solve for the price indexes differs from (5.32) by the constant Y^w , which appears in (5.32) via the share $s^i = Y^i / Y^w$. Notice that if the supply term S^i satisfies (6.26), then αS^i will equally satisfy this condition for any $\alpha > 0$. Anderson and van Wincoop are using the supply terms $S^i \equiv N^i (p^i)^{1-\sigma} / Y^w$ to obtain (5.32), instead of $S^i \equiv N^i (p^i)^{1-\sigma}$. The choice of the constant α affects the recursive equation to obtain the CES prices indexes and therefore implies a normalization on their solution.

$$\frac{X^{ij}}{X^j} = \frac{M_e^i \int_{\varphi^j}^{\infty} p^{ij}(\varphi)^{1-\sigma} g(\varphi) d\varphi}{\sum_{k=1}^C M_e^k \int_{\varphi^j}^{\infty} p^{kj}(\varphi)^{1-\sigma} g(\varphi) d\varphi} = \frac{M_e^i \int_{\varphi^j}^{\infty} (\tau^{ij} w^i / \varphi)^{1-\sigma} g(\varphi) d\varphi}{\sum_{k=1}^C M_e^k \int_{\varphi^j}^{\infty} (\tau^{kj} w^k / \varphi)^{1-\sigma} g(\varphi) d\varphi}, \quad (6.28)$$

where the second equality follows by using the prices $p^{ij}(\varphi) = [\sigma/(\sigma-1)](\tau^{ij} w^i / \varphi)$. Notice that the iceberg trade costs τ^{ij} can be moved outside the integrals in the above expression.

Comparing (6.28) with (6.26), at first glance it seems that we can fulfill the hypothesis of the Head-Mayer theorem by defining $T^{ij} \equiv (\tau^{ij})^{1-\sigma}$ and $S^i \equiv M_e^i \int_{\varphi^j}^{\infty} (w^i / \varphi)^{1-\sigma} d\varphi$. But these definitions do not actually work, because the supply term S^i is supposed to depend on data for country i only, but $S^i \equiv M_e^i \int_{\varphi^j}^{\infty} (w^i / \varphi)^{1-\sigma} g(\varphi) d\varphi$ also depends on country j through the lower-limit of integration, φ^j . So we need to investigate the ZCP value of productivity more carefully and define the supply terms independent of country j in order to obtain a gravity equation.

The ZCP condition for productivity is in (6.12), which is written more generally for country i exporting to j (and without normalizing the wage at unity) as

$$(\varphi^{ij})^{\sigma-1} = \frac{w^i f^{ij} (w^i \tau^{ij})^{\sigma-1} (\sigma-1)^{1-\sigma}}{w^i L^j (P^j)^{\sigma-1} \sigma^{-\sigma}}. \quad (6.29)$$

We now make use of the Pareto distribution in (6.15) to evaluate the integral

$$\begin{aligned} \int_{\varphi^j}^{\infty} (\tau^{ij} w^i / \varphi)^{1-\sigma} g(\varphi) d\varphi &= (\tau^{ij} w^i)^{1-\sigma} \theta \int_{\varphi^j}^{\infty} \varphi^{\sigma-\theta-2} d\varphi \\ &= \frac{(\tau^{ij} w^i)^{1-\sigma} \theta}{\sigma - \theta - 1} \varphi^{\sigma-\theta-1} \Big|_{\varphi^j}^{\infty} \\ &= \frac{(\tau^{ij} w^i)^{1-\sigma} \theta}{\theta - \sigma + 1} (\varphi^{ij})^{\sigma-\theta-1}, \end{aligned}$$

where the last line uses $\theta > (\sigma-1)$. Combining this result with (6.28) and (6.29) it follows that

$$\frac{X^{ij}}{X^j} = \frac{M_e^i (w^i \tau^{ij})^{-\theta} (w^i f^{ij})^{1-\frac{\theta}{\sigma-1}}}{\sum_{k=1}^C M_e^k (w^k \tau^{kj})^{-\theta} (w^k f^{kj})^{1-\frac{\theta}{\sigma-1}}}. \quad (6.30)$$

This equation is in a form that readily translates into the Head-Mayer theorem, by defining $T^{ij} \equiv (\tau^{ij})^{-\theta} (f^{ij})^{1-\frac{\theta}{\sigma-1}}$ and $S^i \equiv M_e^i (w^i)^{1-(\frac{\theta}{\sigma-1})^\theta}$. Thus, the supply term reflects the mass of entering firms in country i adjusted for the wage, while the trade cost term reflects iceberg transport costs with the coefficient $-\theta$ and fixed costs with the coefficient $1 - [\theta/(\sigma-1)] < 0$. As pointed out by Chaney (2008), this coefficient on the iceberg transport costs is quite different from what is found for the Krugman (1980) model, where iceberg costs were raised to the exponent $1 - \sigma < 0$. In the Melitz-Chaney model, we find that it is instead the Pareto parameter that determines the impact of trade costs. Furthermore the gains from trade in these two models are $\lambda_d^{-1/(\sigma-1)}$ and $\lambda_d^{-1/\theta}$, as discussed above, and so the exponent from trade costs in the gravity equation is what also appears as the exponent of the domestic share λ_d in the gains from trade formula, as emphasized by ACR (2012).¹¹

¹¹Because we are now denoting countries by the superscripts i and j , the share of expenditure purchased from domestic firms, λ_d , is more appropriately written as λ_d^{ij} , i.e., the share of expenditure in country i purchased from its own firms or goods.

Eaton and Kortum (2002) Model

We turn now to a quite different justification for the gravity equation, arising from the model of Eaton and Kortum (2002). These authors extend the model of Dornbusch, Fischer, and Samuelson (1977), discussed in chapter 3, to allow for random draws of productivity in each country for each of the goods in the continuum.¹² Specifically, we will let the indexes i and j denote countries, while $z \in [0, 1]$ denotes a product. A country's productivity in producing good z in country i is $\varphi^i(z)$. There is only one input in this model, labor with the wage w^i , and with constant returns to scale the cost of producing good z in country i is $w^i/\varphi^i(z)$.

Because there are constant returns to scale in production, this model differs from Melitz-Chaney by using perfect competition rather than monopolistic competition. But as in the Melitz-Chaney model we still introduce iceberg trade costs of τ^{ij} . The cost of good z produced in country i and delivered to country j is then

$$p^{ij}(z) = \frac{\tau^{ij} w^i}{\varphi^i(z)}. \quad (6.31)$$

Under perfect competition, consumers will choose a good z from a country j that offers the lowest price. So equilibrium price of good z imported into country j will be

$$p^j(z) = \min_i \{p^{ij}(z); i=1, \dots, C\}, \quad (6.32)$$

where C is a number of countries. Given this equilibrium price in country j , we assume that the consumer has CES preferences over the goods $z \in [0, 1]$, with elasticity σ .

The key feature in the Eaton-Kortum model is to choose a probability distribution for $\varphi^i(z)$ in each country so that it is possible to solve for the equilibrium prices from (6.31) and (6.32) in a relatively simple manner. To this end, Eaton and Kortum use the Fréchet distribution for the productivities, which is written as

$$F^i(\varphi) = \exp(-S^i \varphi^{-\theta}), \varphi > 0 \text{ with } F(0)=0, \quad (6.33)$$

where $S^i > 0$ is a country specific technology parameter that defines how productive a country by influencing the mean value for productivity, and $\theta > 1$ is a Fréchet parameter that is common for all countries and determines the variance of productivity. This distribution applies to any good $z \in [0, 1]$, so we now drop that index. The probability that country i exporting to j has a price at least as low as p is $\Pr[p^{ij} \leq p] = 1 - F^i(w^i \tau^{ij}/p)$, or

$$G^{ij}(p) = 1 - \exp(-S^i (w^i \tau^{ij})^{-\theta} p^\theta). \quad (6.34)$$

Now applying expression (6.32), the probability that the price of a particular good in country j is at least as low as p equals one minus the probability that all exporting countries have prices greater than p . Because the productivity distributions are

¹²See problem 6.6 for the connection to Dornbusch, Fischer, and Samuelson (1977). See also Alvarez and Lucas (2007) for a detailed exposition of the Eaton-Kortum model, including properties of the Fréchet distribution.

independent across countries, the minimum price of a particular good in country j is therefore distributed as

$$G^j(p) = 1 - \prod_{i=1}^C [1 - G^{ij}(p)] = 1 - \exp(-R^j p^\theta), \quad (6.35)$$

where the second equality follows by substituting from (6.34) with $R^j \equiv \sum_{i=1}^C (w^i \tau^{ij})^{-\theta}$. We see that this minimum price has (one minus) a Fréchet distribution, with parameters R^j and θ . More generally, the minimum or maximum of a set of draws from a Fréchet distribution inherits that type of distribution, which is why Eaton and Kortum use this convenient family.

There is one more step to obtain a gravity equation, and that is to solve for the probability that country i is in fact the lowest-cost supplier to country j . In other words, letting p^{ij} denote the price charged by country i and p^j denote the equilibrium price in country j , we need to solve for $\Pr[p^{ij} \leq \min\{p^{kj}; k \neq i\}]$. It turns out that this probability is solved as¹³

$$\Pr[p^{ij} \leq \min\{p^{kj}; k \neq i\}] = \frac{S^i (w^i \tau^{ij})^{-\theta}}{R^j}. \quad (6.36)$$

For any particular good z , we do not know which country i will be the lowest-cost supplier. But because we are working with a continuum of goods $z \in [0, 1]$, by the law of large numbers we know that the probability shown in (6.36) will equal the fraction of goods in the continuum for which country i is the lowest-cost supplier, and therefore the exporter to j . Furthermore, conditional on being the lowest-cost supplier, the expected exports using CES demand are identical across goods and suppliers. So integrating over all goods, the value of exports X^{ij} relative to total purchases X^j in country j will also equal that shown in (6.36):

$$\frac{X^{ij}}{X^j} = \frac{S^i (w^i \tau^{ij})^{-\theta}}{R^j}. \quad (6.36')$$

It turns out that $R^j \equiv \sum_{i=1}^C S^i (w^i \tau^{ij})^{-\theta}$ is proportional to the CES price index,¹⁴ so that (6.36') is a familiar CES demand equation. It is immediate that this equation fits the conditions of the Head-Mayer theorem, with the supply term S^i as shown and with $T^{ij} \equiv (w^i \tau^{ij})^{-\theta}$. We therefore obtain a gravity equation from the Eaton-Kortum model that is extremely similar to that from the Melitz-Chaney model, but now re-interpreting θ as the Fréchet parameter rather than the Pareto parameter. Furthermore, because (6.36') corresponds to CES demand, we can apply the gains from trade formula from chapter 5, which is $\lambda_d^{-1/(\sigma-1)}$, where σ is the elasticity of substitution and λ_d is the share of purchases from domestic sources.¹⁵ But the elasticity of substitution appears as $(1-\sigma)$ in the export share equation (see (6.28), for example), whereas that parameter appears as $-\theta$ in (6.36'), so making this substitution the gains from trade formula is instead written as $\lambda_d^{-1/\theta}$ for the Eaton-Kortum model. This result again confirms formula (6.23) in the ACR theorem, whereby the exponent on trade costs in the

¹³See problem 6.5.

¹⁴See Eaton and Kortum (2002, 1749, note 18).

¹⁵See note 11 of this chapter.

gravity equation becomes the exponent on the domestic share in the gains from trade formula.

Zeros in Trade

A limitation of the gravity equations derived above is that all countries will export to (and import from) every other country. The reason for this outcome in the Krugman (1980) model, leading to the gravity equation of Anderson and van Wincoop (2003), is that the CES demand equation has positive demand even for goods of arbitrarily high price, so even countries with very high wages will be able to export a small amount. The Melitz-Chaney model has fixed costs of exporting, so that goods whose prices are too high will not be able to cover these costs and will not be exported. Instead, the feature leading to trade between all countries is that the productivities $\varphi > 0$ for the Pareto distribution (6.15) are allowed to be arbitrarily high, so there will always be some firm in every country that is productive enough to overcome the fixed and the iceberg costs of exporting. Likewise, for the Eaton-Kortum model the productivities $\varphi > 0$ in the Fréchet distribution (6.33) are allowed to be arbitrarily high, so every country will have some goods z for which they are the lowest-cost exporter, overcoming their iceberg trade costs no matter how distant they are from other countries.

These theoretical results are unrealistic, because in disaggregate trade data we observe zero trade between many country pairs, and even in aggregated trade data it is not unusual to find that some countries do not trade at all with some others. This fact raises two questions: how can we achieve zero in trade in the theoretical models; and what is the appropriate estimation technique when there are zeros in trade in the data?

On the first question, the approach of Helpman, Melitz, and Rubinstein (2008) is to consider a truncated Pareto distribution with an upper bound to firm productivity φ . Depending on the demand from the importing country, the highest-productivity exporter might not earn sufficient profits to cover the fixed costs of exporting, in which case there will be no trade in that good. On the second question, the appropriate estimation technique in this case is a Heckman two-step estimator, where the first equation is a Probit selection equation for whether or not country i exports to country j , while the second equation is a gravity equation conditional on exporting and includes the Heckman “correction” variable reflecting correlation of the errors across equations.

One challenge is that for strong identification, the Heckman estimator generally requires some variable(s) to appear in the Probit equation that do not appear in the gravity equation. For this purpose, Helpman, Melitz, and Rubinstein (2008) consider two such variables: the cost of opening a firm (e.g., the number of days required), available for a subset of countries; and the probability that two randomly chosen people from different countries would have the same religious denomination. They argue that the cost of opening a firm should affect fixed costs only, but find similar results with the religion measure, which is available for the entire sample. They find that by using the two-step estimator, the effect of distance and other border variables are reduced in the gravity equation as compared to the case where the Heckman correction is not made. In other words, the effect of distance and border variables on the value of exports is overestimated when we do not control for selection into exporting.

A simpler approach than the Heckman two-step estimator is to use the single gravity equation, while recognizing that in the data there will be zero trade between some country pairs. This means, for example, that the OLS log-linear gravity equation as used by Anderson and van Wincoop (2003) in the previous chapter, possibly

including fixed effects, will not be appropriate because the natural log of a zero trade flow does not exist (i.e., it is $-\infty$). In practice, the zero trade flows are simply omitted.¹⁶ In order to include the zeros we need an alternative estimator. The recommendation of Santos Silva and Tenreyro (2006) is to use the Poisson estimator, which is commonly used for “count” data where the dependent variable takes on the nonnegative integers 0, 1, 2, 3, Santos Silva and Tenreyro argue that this estimator is valid for the gravity equation including zeros, where the level of exports will be in the range $[0, +\infty)$. While the estimator can clearly handle zeros, the justification that they give is motivated by the issue of heteroskedasticity rather than only the issue of zeros in the data.

To see their argument, we will assume that the gravity equation takes the form

$$E(X^{ij}|Z^{ij}) = \exp(Z^{ij}\beta), \quad (6.37)$$

where X^{ij} is the value of exports from country i to country j , Z^{ij} is the row vector of explanatory variables in the gravity equation, and β is the column vector of coefficients. We are treating X^{ij} as stochastic with an unspecified distribution (e.g., log-normal, Poisson, etc.) but with the conditional expectation as shown above. Then without loss of generality, we can write that the gravity equation with an additive error $\varepsilon^{ij} \equiv X^{ij} - E(X^{ij}|Z^{ij})$ as

$$X^{ij} = \exp(Z^{ij}\beta) + \varepsilon^{ij}, \quad (6.37')$$

where ε^{ij} satisfies $E(\varepsilon^{ij}|Z^{ij})=0$. In order for this form of the gravity equation to be valid, however, we must have $\varepsilon^{ij} > -\exp(Z^{ij}\beta)$ to prevent exports from becoming negative. In other words, it *must* be the case that the value of the error term and therefore its variance, too, depends on $\exp(Z^{ij}\beta)$, so that there is correlation between ε^{ij} and $\exp(Z^{ij}\beta)$ and also heteroskedasticity in the error. This means that the additive error does not satisfy the desired properties of an error term when the gravity equation is specified in levels, and we now show how to correct for these problems.

To correct for the heteroskedasticity, suppose that we had an estimate \hat{V}^{ij} of the variance $V(X^{ij}|Z^{ij})=V(\varepsilon^{ij}|Z^{ij})$. Then we would choose $\hat{\beta}$ to solve

$$\min_{\beta} \sum_{i,j} [X^{ij} - \exp(Z^{ij}\beta)]^2 / \hat{V}^{ij}.$$

The solution to this problem is obtained from the first-order condition

$$0 = \sum_{i,j} [X^{ij} - \exp(Z^{ij}\beta)] \exp(Z^{ij}\beta) Z^{ij} / \hat{V}^{ij}. \quad (6.38)$$

The solution depends on what variance estimate \hat{V}^{ij} we use, and we consider several cases:

Case 1: Suppose that $\varepsilon^{ij} = \exp(Z^{ij}\beta)\eta^{ij}$ for some random variable η^{ij} with $E(\eta^{ij}|Z^{ij})=0$ and with constant variance σ_{η}^2 . Then the gravity equation (6.37') can be rewritten as

¹⁶Another approach is to add unity to all the trade values X^{ij} , so that $\ln(1)=0$ and $\ln(1+X^{ij})>0$ for $X^{ij}>0$. Alternatively, we could add a positive constant α to all trade values, with α estimated. These approaches are *ad hoc*, however, and the evidence in Santos Silva and Tenreyro (2006) is that they do not pass the RESET specification test that we discuss below.

$$X^{ij} = \exp(Z^{ij}\beta)(1 + \eta^{ij}).$$

Notice that we have transformed the additive error in (6.37') into a multiplicative error. Taking the natural log we obtain the log-linear regression equation

$$\ln X^{ij} = Z^{ij}\beta + \ln(1 + \eta^{ij}). \quad (6.39)$$

This case, therefore, corresponds to the log-linear regression we used in chapter 5. The error term in (6.39) is now well behaved, but we still cannot take the natural log of $X^{ij}=0$ if these values occur for exports. So instead of the log-linear regression, we compute $\hat{V}^{ij} = [\exp(Z^{ij}\hat{\beta})]^2 \sigma_\eta^2$ and use this variance in (6.38) so that the first-order condition becomes

$$0 = \sum_{i,j} [X^{ij} - \exp(Z^{ij}\hat{\beta})] Z^{ij} / \exp(Z^{ij}\hat{\beta}). \quad (6.40)$$

It turns out that (6.40) is identical to the first-order conditions for a *gamma* pseudo-maximum likelihood (PML) estimator, so that command in STATA can be used to obtain $\hat{\beta}$.¹⁷ This derivation makes it clear that the gamma estimator will be similar to the log-linear estimation, except that the gamma estimator also allows us to include zeros for export values.

Case 2: Perhaps we do not believe that the error terms move fully in proportion to the gravity variables, as was assumed above with $\varepsilon^{ij} = \exp(Z^{ij}\beta)\eta^{ij}$. A simple alternative model is to assume that $\varepsilon^{ij} = [\exp(Z^{ij}\beta)]^{0.5} \eta^{ij}$, with the random variable η^{ij} still satisfying $E(\eta^{ij}|Z^{ij})=0$ and constant variance σ_η^2 . Then the predicted variance of the error is $\hat{V}^{ij} = [\exp(Z^{ij}\hat{\beta})]\sigma_\eta^2$. Using this variance in (6.38) the first-order condition becomes

$$0 = \sum_{i,j} [X^{ij} - \exp(Z^{ij}\hat{\beta})] Z^{ij}. \quad (6.41)$$

This equation is identical to the first-order conditions for a *Poisson* PML estimator. The fact that we end up with a Poisson estimator in this case is not surprising, because the conditional variance for the Poisson model is equal to the conditional mean, and that is consistent with our assumption that $V(X^{ij}|Z^{ij}) = [\exp(Z^{ij}\beta)]\sigma_\eta^2 \propto [\exp(Z^{ij}\beta)] = E(X^{ij}|Z^{ij})$.¹⁸ Like the gamma estimator, we can again include zeros for export values, but unlike that case, we should not expect the Poisson estimator to necessarily be similar to log-linear OLS.¹⁹

Case 3: A third approach to the issue of zeros in the trade data is taken by Eaton, Kortum, and Sotelo (2012). Rather than appeal to the law of large numbers in matching the probability of a country exporting to its trade share, as in (6.36) and (6.36'),

¹⁷ See empirical exercises 6.1 and 6.2.

¹⁸ As noted by Head and Mayer (2014, 174), the Poisson PML estimator is efficient whenever there is *proportionality* between the conditional variance and conditional mean, and remains consistent even without this proportionality. In the case of “over-dispersion” (with the variance exceeding the mean) it might be tempting to apply the negative binomial estimator to the gravity equation. But Head and Mayer caution against this approach, since then the coefficients obtained depend in a nontrivial way on the *units* in which exports are measured.

¹⁹ In fact, Santos Silva and Tenreyro (2006) argue that when the Poisson model holds, a log-linear regression will result in biased coefficients. To see this, substitute $\varepsilon^{ij} = [\exp(Z^{ij}\beta)]^{0.5} \eta^{ij}$ into (6.37) and take natural logs to obtain $\ln X^{ij} = Z^{ij}\beta + \ln \{1 + \eta^{ij}[\exp(Z^{ij}\beta)]^{-0.5}\}$. The error term appearing in curly brackets is not independent of the regressors Z^{ij} , so the coefficient estimates will be biased.

they assume instead that the number of products is finite, so the gravity equation (6.36') takes on a slightly different form:

$$E\left(\frac{X^{ij}}{X^j}\right) = \frac{S^i(w^i \tau^{ij})^{-\theta}}{R^j}. \quad (6.42)$$

Think of the right-hand side of (6.42) as a gravity model $\exp(Z^{ij}\beta)$, as we used above. Then in place of (6.37) we are writing the gravity equation as $E[(X^{ij}/X^j)|Z^{ij}] = \exp(Z^{ij}\beta)$. We see that the difference between (6.37) and (6.42) is that the left-hand side variable is a *share* rather than a value of exports. Eaton, Kortum, and Sotelo argue that the estimator then becomes multinomial PML. Furthermore, Head and Mayer (2014, 179) refer to unpublished notes by Sotelo and state that multinomial PML can be implemented by applying the Poisson PML command in STATA to the share (X^{ij}/X^j) rather than to the level of exports X^{ij} .

Estimating Gravity Once Again

To illustrate the various estimators that allow for zero values in trade, in table 6.1 we report the estimates of the gravity equation for Canadian provinces and U.S. states that we introduced in the previous chapter, table 5.2. The gravity equation (6.24) is estimated as

$$E(X^{ij}|Z^{ij}) = (d^{ij})^\alpha e^{\beta_1 \delta_1^i} e^{\beta_2 \delta_2^j}, \quad (6.43)$$

where the explanatory variables Z^{ij} are specified as in (5.39): d^{ij} is the distance from country i to j ; δ_1^i is an indicator variable that is unity if region i is the exporter, and zero otherwise; and δ_2^j is an indicator variable that is unity if region j is the importer,

Table 6.1: Gravity Equation for Canada-U.S. Trade, 1993

Estimator: ^a	Dependent Variable: Value of Exports for Province/State Pair						
	OLS $\ln(X^{ij})$	Poisson $X^{ij} > 0$	Poisson X^{ij}	Gamma $X^{ij} > 0$	Gamma X^{ij}	Multinomial $X^{ij} > 0$	Multinomial X^{ij}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance	-1.25 (0.04)	-1.01 (0.04)	-1.00 (0.04)	-1.33 (0.05)	-1.27 (0.04)	-1.01 (0.04)	-1.12 (0.04)
Border indicator	-1.55 (0.06)	-1.24 (0.09)	-1.24 (0.09)	-1.21 (0.07)	-1.24 (0.07)	-1.01 (0.12)	-1.08 (0.12)
Border effect ^b	4.7	3.5	3.5	3.6	3.5	2.75	2.94
RESET test ^c	0.000	0.025	0.025	0.000	0.000	0.130	0.095
Observations	1,511	1,511	1,560	1,511	1,560	1,511	1,560

Source: Author's calculations (see empirical exercise 6.1).

Notes: Standard errors are in parentheses. All regressions include fixed effects for source and destination provinces or states.

a. For the multinomial estimator, the dependent variable is the value of exports divided by GDP in the destination province or state.

b. Computed as the exponent of the (absolute value of the) coefficient on the border indicator.

c. A linear predictor of the dependent variable is constructed from the explanatory variables with estimated coefficients. Then the squared predicted values are included into the regression. A Wald test is performed with the null hypothesis that the coefficient on the squared predicted values is equal to zero, and the *p*-value for this test is reported.

and zero otherwise. Because we have included these source and destination country fixed-effects, there is no need to also include the GDP of the source and destination countries when estimating the gravity equation for a single year, as we shall do for the Canada and U.S. province/state data in 1993.

In table 6.1, we start in column (1) by reporting the OLS estimates using $\ln(X^{ij})$ as the dependent variable and omitting trade flows of zero. These are the same estimates that appear in column (5) of table 5.2. We then report the estimates of the same equation using the Poisson, gamma and multinomial PML estimators, each of these with and without zeros. As noted above, the multinomial PML estimator uses (X^{ij}/X^j) as a dependent variable like in (6.42), with the same explanatory variables as in (6.43).

We can see from table 6.1 that the estimation method makes a substantial difference to the results obtained, and in particular, to the estimated border effect. The impact of the Canada-U.S. border in the OLS estimate is 4.7, meaning that there is 4.7 times as much trade *within* Canada or the United States, on average, than across the border. That estimate is lower in all other specifications. The impact of distance is similar for OLS and gamma PML, which, as we argued above, should be the case. But in Poisson PML we find that distance has less impact on international trade. The results with and without zero trade flows are quite similar in each PML estimator, but that result follows from an unusual feature of the dataset: out of 1560 observations, there are only forty-nine zero trade flows. Actually, the Canada-U.S. dataset is highly incomplete in that it includes all ten Canadian provinces but only thirty U.S. states. Presumably, many of the “missing” states have zero trade with some other states and provinces.

Following Santos Silva and Tenreyro (2006), we can test for heteroskedasticity in each specification by using the RESET test. Under this test, a linear predictor of the dependent variable is constructed from the explanatory variables with estimated coefficients. Then the squared predicted values are included into the regression. A Wald test is performed with a null hypothesis that the coefficient on the squared predicted values is equal to zero, and the p -value for this test if reported. We see from table 6.1 that the RESET test allows us to reject the hypothesis of no heteroskedasticity for OLS and the gamma PML (with or without zeros); but it fails to reject this hypothesis for the Poisson PML specification at the 1 percent level, though it does reject it at the 5 percent level; and it also fails to reject this hypothesis for the multinomial PML specification at about the 10 percent level. Based on this test, the multinomial PML estimator with the null hypothesis of no heteroskedasticity appears to fit the data best. It gives a lower value of the border effect than the other estimators, and also has a distance coefficient that is at least as low as the other estimators (and lower when zeros are included).

Because the Canada-U.S. dataset has very few zeros in it, we have repeated this exercise for the data provided by Santos Silva and Tenreyro (2006) on bilateral trade among 136 countries in 1990. In this dataset, 52 percent of the sample has positive exports and the remaining 48 percent has zero exports. The results are shown in table 6.2. We see that the indicator variable for a common border is significant only for the Poisson PML specification and for the gamma specification including zeros. There is substantial variation in the estimated coefficient of the distance variable, which is closest to zero for Poisson PML and farthest from zero for gamma PML. The addition of zero trade flows, which nearly doubles the number of observations, does not change the results that much for the Poisson and multinomial estimators, but it has a surprisingly large impact on the gamma estimates. Despite this sensitivity, the OLS estimates for most coefficients tend to lie in between the two sets of gamma estimates, and so the OLS estimates are closer to the gamma estimates than to other estimators, as expected.

Table 6.2: Gravity Equation for 136 Countries, 1990

Estimator: ^a	Dependent Variable: Value of Exports for Country Pair						
	OLS $\ln(X^{ij})$	Poisson $X^{ij} > 0$	Poisson X^{ij}	Gamma $X^{ij} > 0$	Gamma X^{ij}	Multinomial $X^{ij} > 0$	Multinomial X^{ij}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance	-1.35 (0.03)	-0.75 (0.04)	-0.77 (0.04)	-1.93 (0.08)	-1.17 (0.04)	-1.11 (0.06)	-1.06 (0.06)
Common border	0.17 (0.13)	0.37 (0.09)	0.35 (0.09)	-0.46 (0.30)	0.37 (0.14)	0.01 (0.14)	0.03 (0.14)
Common language	0.41 (0.07)	0.38 (0.09)	0.42 (0.09)	0.68 (0.14)	0.42 (0.07)	0.77 (0.14)	0.78 (0.14)
Colonial tie	0.67 (0.07)	0.08 (0.13)	0.04 (0.13)	0.81 (0.15)	0.51 (0.08)	0.56 (0.14)	0.51 (0.13)
FTA indicator	0.31 (0.10)	0.38 (0.08)	0.37 (0.08)	1.47 (0.33)	0.58 (0.12)	1.03 (0.17)	1.00 (0.16)
RESET test ^b	0.000	0.567	0.113	0.001	0.000	0.007	0.018
Observations	9,613	9,613	18,360	9,613	18,360	9,613	18,360

Source: Columns (1) – (3) are from Santos Silva and Tenreyro (2006, table 5, p. 652), and the remaining columns are from author's calculations (see empirical exercise 6.2).

Notes: Standard errors are in parentheses. All regressions include fixed effects for source and destination countries.

a. For the multinomial estimator, the dependent variable is the value of exports divided by GDP in the destination province or state.

b. A linear predictor of the dependent variable is constructed from the explanatory variables with estimated coefficients. Then the squared predicted values are included into the regression. A Wald test is performed with the null hypothesis that the coefficient on the squared predicted values is equal to zero, and the *p*-value for this test is reported.

Once again, we use the RESET test for heteroskedasticity. Now we find that that the Poisson PML estimates fail to reject the null of no heteroskedasticity at any level of significance, whereas the multinomial PML estimates fail to reject this hypotheses at the 1 percent or 5 percent level, depending on whether zeros are included or not. So in contrast to the estimates in table 6.1 for trade within and between Canada and the United States, in table 6.2 the Poisson estimator is preferred to the multinomial estimator. The results shown in the two tables are consistent, however, in rejecting the null of no heteroskedasticity in the OLS and gamma specifications; and as argued by Santos Silva and Tenreyro (2006), such heteroskedasticity can lead to biased coefficient estimates.²⁰ So our recommendation from these exercises is that researchers use the Poisson or multinomial PML estimators, which differ only in the way that the dependent variable is constructed (the level of exports for Poisson, and the share of exports in the destination country for multinomial).²¹ In addition, researchers should check for the presence of heteroskedasticity using the RESET or other tests before being satisfied with their gravity estimates.

²⁰See the previous note.

²¹Fally (2014) provides a further reason for using the Poisson PML estimator including fixed effects for countries. He shows that the sum of fitted outward flows (including domestic trade) will then automatically equal the actual output of countries. It follows that the fixed effects can be used to accurately reconstruct the multilateral resistance terms, without needing the type of calculation done by Anderson and van Wincoop (2003) as discussed in chapter 5.

Estimating Trade Costs and the Elasticity of Trade

While we have made good progress toward estimating the gravity equation, along the way we have lost sight of the trade costs T^{ij} . These costs are important not only in their own right but also because they enable us to uncover the *exponent* of the iceberg transport costs τ^{ij} . In the model of Krugman (1980) leading to the gravity equation (6.27') of Anderson and van Wincoop (2003), log trade costs have the coefficient $1 - \sigma$, related to the elasticity of substitution. In the Melitz-Chaney model, the log trade costs in (6.30) have the coefficient $-\theta$ where θ is the Pareto parameter, and likewise in the Eaton-Kortum model with the gravity equation (6.36'), except that θ is then the Fréchet parameter. In general, we will call the (negative of) the coefficient of log trade costs the *elasticity of trade* with respect to iceberg trade costs. The question, then, is how to estimate these trade costs or the elasticity of trade.

Head and Mayer (2014, 153) provide a concise summary of methods they refer to as “ratio-type estimates.” These methods simplify the gravity equation in (6.24), $X^{ij} = (Y^i/R^i)(X^j/R^j) T^{ij}$, by expressing it in ratio form. For example, if the trade costs T^{ij} (consisting of τ^{ij} and its exponent) are symmetric so that $T^{ij} = T^{ji}$, and also that $T^{ii} = T^{jj} = 1$, then the trade costs can be recovered as

$$T^{ij} = \left(\frac{X^{ij} X^{ji}}{X^{ii} X^{jj}} \right)^{1/2}. \quad (6.44)$$

One difficulty with this method, due to Head and Ries (2001), is that the amount of trade with your own country—which is what X^{ii} and X^{jj} represent—is hard to measure in practice. An alternative, due to Romalis (2007) and Hallak (2006), is to form the ratio

$$\frac{X^{ij}/X^{ik}}{X^{lj}/X^{lk}} = \frac{T^{ij}/T^{ik}}{T^{lj}/T^{lk}}. \quad (6.45)$$

This is a more complicated estimate of trade costs, but is still independent of other variables in the gravity equation.

Neither of the above methods uncovers, however, the elasticity of trade with respect to iceberg trade costs. To achieve that end we need to use additional data. Caliendo and Parro (2015) use *ad valorem* (or percentage) tariffs between countries, denoted here by t^{ij} , and suppose that the trade costs are related to tariffs and distance by $T^{ij} = [(1 + t^{ij})(d^{ij})^\eta]^\eta$, where η is the elasticity of trade with respect to trade costs. Since distance is symmetric, $d^{ij} = d^{ji}$, then using the trade cost equation and (6.24), we can show that

$$\frac{X^{ij} X^{jh} X^{hi}}{X^{ih} X^{hj} X^{ji}} = \left[\frac{(1 + t^{ij})(1 + t^{jh})(1 + t^{hi})}{(1 + t^{ih})(1 + t^{hj})(1 + t^{ji})} \right]^\eta, \quad (6.46)$$

so that the elasticity of trade η can be estimated as the exponent on the right-hand side. Caliendo and Parro use these estimates to calculate the impact of the North American Free Trade Agreement on the countries involved, as we shall discuss in the next chapter. Bas, Mayer, and Thoenig (2014) apply a similar technique to (6.45) to obtain the elasticity of trade.

An alternative to using tariffs to estimate the elasticity of trade is to use price data. In the appendix to chapter 5 we discussed the technique due to Feenstra (1994), which

uses the share of the prices (or more precisely, unit-values) of imports by destination country to estimate the elasticity of substitution across these countries, from which the elasticity of trade $1 - \sigma$ is readily obtained. Earlier in this chapter we also described the work of Simonovska and Waugh (2014a,b), which also relies on prices of traded goods in different countries to estimate the elasticity of trade. Their technique applies to any of the models leading to a gravity equation that we have discussed and shows how the estimates obtained depend on the assumed model structure.

CONCLUSIONS

In this chapter we have covered a large amount of new research dealing with the Melitz (2003) model as well as the Eaton and Kortum (2002) model, and drawing upon the survey chapters by Melitz and Redding (2014a) and Head and Mayer (2014). The reader is referred to those chapters for many additional results. Despite the progress we have made, however, there is a good deal of recent research that we have not been able to include.

Specifically, we have omitted discussion of the results obtained with functional forms other than the CES. As stressed at the beginning of chapter 5, the original model of Krugman (1979), using a general additively separable utility function, allowed for a reduction in the markups charged by firms due to international trade. That channel does not operate in the Melitz (2003) model because of CES preferences, and of course, does not operate in the Eaton and Kortum (2002) model because it assumes perfect competition. To discard with the CES assumption, there are three major approaches. First, there are papers that follow Krugman (1979) in assuming an additively separable utility function.²² Second, the work of Melitz and Ottaviano (2008) uses a quadratic utility function, with linear demand curves, to restore the impact of trade on the markups charged by firms. Both these approaches assume nonhomothetic utility functions. A third approach is the translog expenditure function, used by Feenstra (2003) and Feenstra and Weinstein (2010), which corresponds to a utility function that is homothetic but does not require constant price elasticities of demand. Bergin and Feenstra (2000, 2001, 2009) and Rodríguez-López (2011) use the translog function to analyze questions in international macroeconomics. Mrázová and Neary (2013) encompass all these cases and provide a general framework to analyze how the demand structure affects the comparative statics results.

Variable markups are also obtained by De Blas and Russ (2015) in the context of the model by Bernard, Eaton, Jensen, and Kortum (2003). That model generalizes Eaton and Kortum (2002) to allow for imperfect competition between firms. Specifically, Bernard, Eaton, Jensen, and Kortum (2003) assume Bertrand competition, which leads to markups that equal the difference between the productivity of the most efficient and second-most efficient firms. With entry by a finite number of potential rivals, De Blas and Russ (2015) show that these markups are not fixed by the productivity distribution of firms, but depend on the number of rivals. When opening to trade alters the number of potential rivals, then markups will also change.

Allowing for preferences beyond the CES also influences the gains from trade. Arkolakis, Costinot and Rodriguez-Clare (ACR 2012) have shown that *with* CES

²²See note 7 in chapter 5, referring to the work of Behrens and Murata (2007, 2012), Dhingra and Morrow (2013), Kichko et al. (2013), Saure (2009), Simonovska (2015), and Zhelobodko et al. (2011, 2012), for example.

preferences, the monopolistic competition model gives rise to a remarkably simple formula for the gains from opening trade: those gains are equal to one minus the import share of the economy, raised to an exponent that depends on the specific details of the model. In the Krugman (1980) model with homogeneous firms, that exponent is the inverse of one minus the elasticity of substitution. In the Melitz-Chaney model with heterogeneous firms and a Pareto distribution for productivities, the exponent equals minus the inverse of the Pareto parameter. As we have argued in this chapter, this result is obtained because import competition drives out a mass of domestic varieties that just cancels out in welfare terms with increased import variety, so that the only remaining source of gains from trade is from the positive selection of firms (governed by the Pareto parameter). In the Eaton-Kortum model, the same formula holds where now the exponent is minus the inverse of the Fréchet parameter. Furthermore, Atkeson and Burstein (2010) show that essentially the same gains from trade hold in a model that also allows for product innovation.

Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012) have recently argued that a similar result holds in a broader class of models that allows for a family of preferences beyond the CES, including the additively separable and the quadratic preferences. They find that a similarly simple formula for the gains from trade—equal to one minus the import share raised to an exponent—continues to hold. That result hinges very strongly, however, on the assumption of a Pareto distribution for the productivities of firms that is *unbounded above*, like in (6.15). Feenstra (2014) shows that this assumption eliminates any gains from trade due to increased variety or due to a changing distribution of markups, even with quite general preferences.²³ All that is left are the gains due to the positive selection of firms. In contrast, once the Pareto distribution is bounded or truncated above, then the gains from trade due to increased product variety and reduced markups of firms are restored, though interestingly, the formula in ACR (2012) continues to hold as an upper bound to the total gains from trade.

We conclude from this discussion that along with examining preferences other than the CES, it is equally important to examine distributions for firm productivity other than the unbounded Pareto. Melitz and Redding (2013) examine the truncated Pareto distribution, with CES preferences, and show important differences as compared to the unbounded case. Head, Mayer, and Thoenig (2014) and Bas, Mayer, and Thoenig (2014) argue that the log-normal distribution offers a better approximation to actual firm sizes than the Pareto, as is applied by Yang (2014). Edmond, Midrigan, and Xu (2012) allow for a “double” Pareto as a way to better approximate the actual distribution, and Holmes, Hsu, and Lee (2014) also investigate distributions other than the Pareto. These authors show that the pro-competitive gains from trade tend to be greater outside the unbounded Pareto case. Finally, Feenstra and Weinstein (2010) estimate the pro-competitive gains for the United States over 1992–2005 without making any assumption regarding the distribution of firm productivities. They find that

²³Even though the markups of every active firm will change when trade costs are reduced, it is still the case that the *distribution* of markups is unchanged when the Pareto distribution is unbounded above. The intuition provided by Feenstra (2014) is as follows. Suppose that we measure markups by the *ratio* of price to marginal cost, minus unity. The most productive firm has infinite productivity and zero cost, but a nonzero price, so its markup is infinite, while in the absence of fixed costs the least productive surviving firm has a markup of zero. This range of $[0, +\infty)$ for the markups applies to both domestic and foreign firms and is not affected by a change in trade costs.

the gains from reduced markups are on the same order of magnitude as the gains from increased product variety.

Beyond investigating distributions other than the unbounded Pareto, there are many other aspects of the production side that should be generalized. First, it is important to allow for nonsymmetric countries, as is done in Bernard, Redding, and Schott (2007), ACR (2012), and Melitz and Redding (2014a). Second, once symmetry has been removed, the nature of the fixed costs of exporting needs to be examined more carefully. In the original Melitz (2003) model, it is unclear whether the fixed costs apply to a single export market or to all export markets: the assumed symmetry prevents this distinction from being made. In practice the distinction is important, and we might expect that once the fixed costs to one destination have been paid, it is easier for firms to export to other destinations with similar characteristics. That hypothesis has been confirmed empirically by Morales, Sheu, and Zahler (2014). For the motion picture industry, Hanson and Xiang (2011) find evidence to support the idea that there is a single, global fixed cost that must be paid in order for firms to export worldwide.

Third, the production side of the heterogeneous firms model can be enriched by adding stages of production. Melitz and Redding (2014b) introduce a *sequence* of production activities that rely on traded intermediate inputs. As the number of sequential stages grows, the gains from trade also grow, and the simple formula obtained by ACR (2012) no longer applies. Chaney and Ossa (2012) have likewise introduced sequential stages of production into the Krugman (1979) model with homogeneous firms. These are only a few examples of the many ways in which the monopolistic competition and trade model can be generalized.

PROBLEMS

The Pareto distribution of firm productivities is shown by (6.15). In the following three questions we will use this distribution to explicitly solve for the equilibrium in the Melitz-Chaney model.

- 6.1 (a) By differentiation, show that the function $J(\varphi_a) \equiv \int_{\varphi_a}^{\infty} [(\varphi/\varphi_a)^{\sigma-1} - 1] g(\varphi) d\varphi$ defined in (6.7) is monotonically decreasing in φ_a .
 (b) Using the Pareto distribution in (6.15), show that $J(\varphi_a) = [(\sigma-1)/(\theta-\sigma+1)] [1 - G(\varphi_a)]$.
- 6.2 (a) Using the Pareto distribution in (6.15) and the result of problem 6.1(b), solve for the value of productivity φ_a satisfying the autarky equilibrium condition (6.7).
 (b) Using the Pareto distribution in (6.15) and the result of problem 6.1(b) once again, together with (6.13), solve for the value of productivities φ_d and φ_x satisfying the trade equilibrium condition (6.14). Show that $\varphi_d > \varphi_a$.
- 6.3 (a) Using the result of problem 6.2(a), and also using $M_a = M_e [1 - G(\varphi_a)]$ while treating M_e as given, solve for M_a . How does your answer compare with (6.10)?
 (b) Using the result of problem 6.2(b), and also using $M_d = M_e [1 - G(\varphi_d)]$ and $M_x = M_e [1 - G(\varphi_x)]$ while treating M_e as given, solve for M_d and M_x .
 (c) Compare the value of M_a from part (a) with $M_d + M_x$ in part (b). Are there simple conditions under which one of these values exceeds the other?

6.4 The full employment condition under trade is

$$L = M_e f_e + M_d \int_{\varphi_d}^{\infty} \left[\frac{y(\varphi)}{\varphi} + f_d \right] \frac{g(\varphi)}{[1 - G(\varphi_d)]} d\varphi + M_x \int_{\varphi_x}^{\infty} \left[\frac{y_x(\varphi)}{\varphi} + f_x \right] \frac{g(\varphi)}{[1 - G(\varphi_x)]} d\varphi,$$

which can be compared to (6.9) in autarky. We can rewrite this condition by using prices net of transport costs $p(\varphi) = (1/\varphi)[\sigma/(\sigma-1)]$ to obtain

$$\begin{aligned} L &= (M_e f_e + M_f + M_x f_x) + \left(\frac{\sigma-1}{\sigma} \right) \left[M_d \int_{\varphi_d}^{\infty} \frac{p(\varphi) y(\varphi) g(\varphi)}{[1 - G(\varphi_d)]} d\varphi + M_x \int_{\varphi_x}^{\infty} \frac{p(\varphi) y_x(\varphi) g(\varphi)}{[1 - G(\varphi_x)]} d\varphi \right] \\ &= (M_e f_e + M_f + M_x f_x) + \left(\frac{\sigma-1}{\sigma} \right) L, \end{aligned}$$

where the second line uses the definition of $GDP = L$, with wages normalized at unity and zero expected profits. It follows immediately that $L = \sigma(M_e f_e + M_d f_d + M_x f_x)$. Then the above full employment condition is simplified as

$$\left(\frac{\sigma-1}{\sigma} \right) L = M_d \int_{\varphi_d}^{\infty} \left[\frac{y(\varphi)}{\varphi} \right] \frac{g(\varphi)}{[1 - G(\varphi_d)]} d\varphi + M_x \int_{\varphi_x}^{\infty} \left[\frac{y_x(\varphi)}{\varphi} \right] \frac{g(\varphi)}{[1 - G(\varphi_x)]} d\varphi.$$

- (a) Using the Pareto distribution and CES demand, evaluate the above two integrals.
- (b) Therefore, show that $L = [\sigma\theta/(\theta-\sigma+1)](M_d f_d + M_x f_x)$.
- (c) Using $L = \sigma(M_e f_e + M_d f_d + M_x f_x)$ solve for M_e and compare your answer to (6.16).

The next two questions deal with the Eaton-Kortum model.

- 6.5 To derive equation (6.36), we follow Eaton and Kortum (2002, note 16) and use the result from probability theory stating that $\Pr[X \leq Y] = \int [1 - F_Y(x)] f_x(x) dx = \int [1 - F_Y(x)] dF_x(x)$. Therefore

$$\Pr[p^{ij} \leq \min\{p^{kj}; k \neq i\}] = \int_0^\infty \prod_{k \neq i} [1 - G^{kj}(p)] dG^{ij}(p).$$

Complete this calculation to establish (6.36).

- 6.6 Dornbusch, Fischer, and Samuelson (1977) do not introduce a functional form for $A(z)$ and assume only that $A(z)$ is decreasing in z , while Eaton and Kortum (2002) specify the technology explicitly. In this problem we will use the model of Eaton and Kortum (EK) to derive the explicit solution for $A(z)$ and for relative wages in Dornbusch, Fischer, and Samuelson (DFS).²⁴ We consider the case of two countries without trade costs.
- (a) Recall that $A(z) = a^*(z)/a(z)$ in DFS, where $a(z)$ and $a^*(z)$ are the amount of labor needed for one unit of output in the home and foreign countries, respectively. Now suppose we model these productivities as $a = x^{1/\theta}$, where x has exponential distribution, namely, $F(x) = 1 - \exp(-Tx)$. Defining productivity as $z = 1/a$, then show that the distribution of productivity will be $G(z) = 1 - \exp(-Tz^{-\theta})$, which is the Fréchet distribution.

²⁴This problem draws on Eaton and Kortum (2002, note 15), and I thank Andrés Rodríguez-Clare for providing the solution shown here.

- (b) Suppose that the distribution for home productivity $a = x^{1/0}$ is the exponential $F(x) = 1 - \exp(-Tx)$, while the distribution for foreign productivity $a^* = (x^*)^{1/0}$ is the exponential $F^*(x) = 1 - \exp(-T^*x)$. According to part (a), these assumptions will correspond to the EK model. To make the connection to the DFS model, notice that $A = a^*/a = (x^*/x)^{1/0}$ becomes a random variable. From (3.28) in chapter 3, we know that the home county will have lower costs and will export whenever $A = (x^*/x)^{1/0} > w/w^*$. Therefore, the probability that $x^*/x > (w/w^*)^0$ equals the mass of home exports.

To solve for the mass of home exports, we use the following properties of the exponential distributions: if $x \sim \exp(\lambda)$, then $kx \sim \exp(\lambda/k)$ and if $x \sim \exp(\lambda)$ and $y \sim \exp(\gamma)$ are independent, then $\Pr[x \leq y] = \lambda/(\lambda + \gamma)$. Then solve for the probability that $x^*/x > (w/w^*)^0$.

Hint: We know that in the equilibrium of the DFS model, $A(z') = w/w^*$. So for notational convenience, replace w/w^* with A , and use the two properties of the exponential distribution to show that

$$\Pr[x^*/x > A^0] = 1 - \Pr[x^* \leq xA^0] = 1 - \frac{T^*}{T^* + A^{-0}T}.$$

- (c) Set the expression obtained in part (b) equal to z' , which is the range of goods exported in the DFS model. Invert this expression to obtain an explicit solution for $A(z')$.
 (d) Combine this expression for $A(z)$ with equation (3.30) to obtain the following solution for relative wages: $w/w^* = (\frac{T^*L}{T^*E})^{1/(1+0)}$.

EMPIRICAL EXERCISES

- 6.1 In this exercise you are asked to reproduce the empirical results presented on table 6.1. This exercise extends the empirical exercise 5.1 in chapter 5. There are four datasets in the folder: “trade_93” reports exports in 1993 between Canadian and American provinces and states; “gdp_ci” and “gdp_ce” contain data on GDPs of all the regions from “trade_93”; and “dist” contains information on the distance between any two regions in the data. To complete the exercise these files should be stored in the directory c:\Empirical_Exercise\Chapter_6\.

When the files are located there, run the STATA program “data_trans2.do”, which will convert these datasets to STATA files with the same name. Run the program “gravity_3.do” to replicate the gravity equation in columns (1)–(7) of table 6.1. Why does the coefficient at the border have a negative sign?

- 6.2 In this exercise you are asked to reproduce the empirical results presented in table 6.2. Download the data from the “log of gravity” webpage, located at <http://privatewww.essex.ac.uk/~jmcss/LGW.html> (accessed October 1, 2014). Run the program “gravity_4.do” to replicate the gravity equation in columns (1)–(7) of table 6.2.

Gains from Trade and Regional Agreements

In earlier chapters we suggested that trade brings gains to a country, but at the same time, there are both winners and losers. The Stolper-Samuelson theorem made that especially clear. In the Heckscher-Ohlin (HO) model the *abundant factor* gains from trade (through the rise in the relative export price, increasing the real return to that factor, used intensively in exports), while the *scarce factor* loses from trade (through the fall in the relative import price, lowering the real return to that factor). Can we be sure that the gains always exceed the losses? That is the topic of this chapter.

We begin with a discussion of “lump sum transfers,” whereby the government has the ability to tax the gainers and transfer income to the losers, without changing their behavior in the process. That is, the lump sum transfers are assumed to be nondistorting. We show that it is indeed possible to achieve Pareto gains from trade, that is, a situation where everyone gains. This result is unrealistic, however, because any attempt by the government to achieve such a redistribution of income would lead people to change their behavior (so as to receive more transfers), and these effects are not taken into account.

As a second approach, we investigate a set of commodity and factor taxes/subsidies that has been proposed by Dixit and Norman (1980). By construction, this policy requires much less information than the lump sum transfers, and any changes in individuals’ behavior due to the tax or subsidies are fully taken into account. We show that this policy can *also* lead to a situation where everyone gains. There are still some assumptions needed to achieve this result, as we will discuss, but the policy is generally thought to be much more feasible than lump sum transfers. Indeed, we provide an example from Akerlof et al. (1991), where factor subsidies were proposed as a means to obtain Pareto gains from trade after the unification of Germany.

The use of lump sum subsidies to achieve Pareto gains from trade can be readily extended to compare any two trading situations, and identify conditions under which the country in question is better off or not, as in Grinols and Wong 1991 and Ju and Krishna 2000a, 2000b. Furthermore, the analysis can be extended to cover tariff reform by multiple countries. This is the goal of the World Trade Organization (WTO): to reduce barriers to trade among its more than 140 member countries. Among the newest members are the People’s Republic of China, accepted into the WTO on November 10, 2001, with Taiwan joining the next day. The WTO was known before 1994 as the General Agreement on Tariffs and Trade (GATT), and

was set up after World War II along with other international institution such as the International Monetary Fund and the World Bank. The GATT made great progress in reducing trade barriers worldwide through “rounds” of discussion, which included the Uruguay Round lasting from 1986 to 1994, and the agreement in Doha, Qatar, on November 14, 2001, to launch a new round of negotiations.

Some of the founding principles of GATT, known as articles, are displayed in table 7.1. The most fundamental of these is the “most favored nation” (MFN) principle (see Article I), which states that all countries belonging to the GATT should be treated equally; that is, if the United States reduces its tariffs on goods coming from Europe, it must do the same for those goods coming from any other member country. Such tariff reductions are called *multilateral*, and all the GATT rounds have been aimed at multilateral reductions in trade barriers.

There are instances, however, where a group of countries will decide to completely eliminate all tariffs between them, *without* eliminating tariffs on goods imported from the rest of the world. This is called a *regional agreement* or *preferential agreement* among the countries involved. In the case where this group of countries also *unifies* its tariffs against the rest of the world, with zero tariffs internally, this is called a *customs union*. The European Economic Community (EEC) is one example of a customs union, and another is MERCOSUR, established in 1995, which now includes Argentina, Brazil, Paraguay, Uruguay, and Venezuela. In contrast, if the group of countries maintains *its own* tariffs against the rest of the world, with zero tariffs internally, then this is called a *free trade area*. Examples of free-trade areas are the Canada-U.S. Free Trade Agreement (CUSFTA) of 1989, which later led to the North American Free Trade Agreement (NAFTA) of 1994, which consists of Canada, the United States, and Mexico. Both customs unions and free trade areas violate the MFN principle, since the countries within the union or area are treated differently from those outside. But such agreements are permitted under Article XXIV of the GATT, provided that “the duties [with outside parties] shall not on the whole be higher or more restrictive than the general incidence of the duties . . . prior to the formation.”

Because customs unions and free trade areas involve only partial elimination of tariffs, they can lead to unexpected results: as pointed out by Viner (1950), the countries left out of the union and even the countries *inside* the union can become worse off. We discuss this possibility at the end of the chapter, and identify conditions under which this outcome can be avoided, as in Kemp and Wan (1976). This also provides us with an important area for empirical research: measuring the gains and losses from preferential agreements, of which there have been an increasing number in recent years. We discuss several empirical studies: Grinols (1984) on Britain’s membership in the EEC; Clausing (2001) and Romalis (2007) on trade creation and diversion due to CUSFTA and NAFTA; and Chang and Winters (2002) on the terms of trade effects of MERCOSUR. We conclude the chapter by extending the results to include imperfect competition and increasing returns to scale, and also by briefly discussing the welfare effects of regional agreements under monopolistic competition.

LUMP SUM TRANSFERS

Our discussion of lump sum transfers follows Dixit and Norman (1980, 76–78). We want to distinguish the various consumers in the economy, so we index them by $h=1,\dots,H$. We will suppose that individual h has the increasing and quasi-concave

TABLE 7.1: Key Provisions of the GATT Agreement Preamble

Recognizing that their relations in the field of trade and economic endeavour should be conducted with a view to raising standards of living, ensuring full employment and a large and steadily growing volume of real income and effective demand, developing the full use of the resources of the world and expanding the production and exchange of goods.

Being desirous of contributing to these objectives by entering into reciprocal and mutually advantageous arrangements directed to the substantial reduction of tariffs and other barriers to trade and to the elimination of discriminatory treatment in international commerce.

Have through their Representatives agreed as follows:

Article I
General Most-Favoured-Nation Treatment

- With respect to customs duties . . . and with respect to all rules and formalities in connection with importation and exportation . . . any advantage, favour, privilege or immunity granted by any contracting party to any product originating in or destined for any other country shall be accorded immediately and unconditionally to the like product originating in or destined for the territories of all other contracting parties. . . .

Article III
National Treatment on Internal Taxation and Regulation

- The contracting parties recognize that internal taxes and other internal charges, and laws, regulations and requirements affecting the internal sale . . . should not be applied to imported or domestic products so as to afford protection to domestic production. . . .

Article VI
Anti-dumping and Countervailing Duties

- The contracting parties recognize that dumping, by which products of one country are introduced into the commerce of another country at less than the normal value of the products, is to be condemned if it causes or threatens material injury to an established industry. . . . [A] product is to be considered . . . less than its normal value, if the price of the product exported from one country to another
 - is less than the comparable price . . . for the like product when destined for consumption in the exporting country, or,
 - in the absence of such domestic price, is less than either
 - the highest comparable price for the like product for export to any third country in the ordinary course of trade, or
 - the cost of production of the product in the country of origin plus a reasonable addition for selling cost and profit. . . .

Article X
Publication and Administration of Trade Regulations

- Laws, regulations, judicial decisions and administrative rulings of general application, made effective by any contracting party, pertaining to the classification or the valuation of products for customs purposes, or to rates of duty, taxes or other charges, or to requirements, restrictions or prohibitions on imports or exports or on the transfer of payments therefore . . . shall be published promptly in such a manner as to enable governments and traders to become acquainted with them. . . . The provisions of this paragraph shall not require any contracting party to disclose confidential information which would impede law enforcement or otherwise be contrary to the public interest or would prejudice the legitimate commercial interests of particular enterprises, public or private.

(continued)

TABLE 7.1: (*Continued*)

Article XI
General Elimination of Quantitative Restrictions

1. No prohibitions or restrictions other than duties, taxes or other charges, whether made effective through quotas, import or export licenses or other measures, shall be instituted or maintained by any contracting party on the importation of any product of the territory of any other contracting party or on the exportation or sale for export of any product destined for the territory of any other contracting party. . . .

Article XVI
Subsidies

1. If any contracting party grants or maintains any subsidy, including any form of income or price support, which operates directly or indirectly to increase exports of any product from, or to reduce imports of any product into, its territory, it shall notify the contracting parties in writing of the extent and nature of the subsidization, of the estimated effect of the subsidization on the quantity of the affected product or products imported into or exported from its territory and of the circumstances making subsidization necessary. In any case in which it is determined that serious prejudice to the interests of any other contracting party is caused or threatened by any such subsidization, the contracting party granting the subsidy shall, upon request, discuss with the other contracting party . . . the possibility of limiting the subsidization.

Article XVII
Restrictions to Safeguard the Balance of Payments

1. Notwithstanding the provisions of paragraph 1 of Article XI, any contracting party, in order to safeguard its external financial position and its balance of payments, may restrict the quantity or value of merchandise permitted to be imported.

Article XIX
Emergency Action on Imports of Particular Products

1. (a) If, as a result of unforeseen developments and of the effect of the obligations incurred by a contracting party under this Agreement, including tariff concessions, any product is being imported into the territory of that contracting party in such increased quantities and under such conditions as to cause or threaten serious injury to domestic producers in that territory of like or directly competitive products, the contracting party shall be free, in respect of such product, and to the extent and for such time as may be necessary to prevent or remedy such injury, to suspend the obligation in whole or in part or to withdraw or modify the concession....

Article XXIV
Territorial Application-Frontier Traffic-Customs Unions and Free-Trade Areas

4. The contracting parties recognize the desirability of increasing freedom of trade by the development, through voluntary agreements, of closer integration between the economies of the countries party to such agreements. They also recognize that the purpose of a customs union or of a free-trade area should be to facilitate trade between the constituent territories and not to raise barriers to the trade of other contracting parties with such territories.
5. Accordingly, the provisions of this Agreement shall not prevent [the formation of customs unions and free-trade areas, provided that:]
 - a. . . . the duties [with outside parties] shall not on the whole be higher or more restrictive than the general incidence of the duties . . . prior to the formation. . . .

Source: Yoffie and Gomes-Casseres (1994, 420–22).

utility function $u^h(c^h, v^h)$, where c^h is the N -dimensional vector of consumption, and v^h is the M -dimensional vector of factor supplies. Under autarky, each consumer maximizes utility subject to the budget constraint

$$p^a' c^h \leq w^a' v^h, \quad (7.1)$$

where p^a is the autarky vector of commodity prices, and w^a is the autarky vector of factor prices. The resulting utility for each consumer is $u^h(c^h, v^h)$.

Summing over all consumption and factor supplies, the total output in the economy is therefore $\sum_{h=1}^H c^{ha} = y^a$, and the total factor inputs are $\sum_{h=1}^H v^{ha} = v^a$. Of course, the point (y^a, v^a) must be feasible for producers, and also optimal at the prices (p^a, w^a) . We will assume that the technology for the economy is subject to constant returns to scale. This means that the payments to factors are exactly equal to the revenue collected, so that profits in the economy are

$$p^a' y^a - w^a' v^a = 0.$$

We now consider the free-trade equilibrium, which includes a system of lump sum taxes or transfers from the government to each individual. We will let R^h denote the transfer to each individual h , so that $R^h < 0$ is a lump sum tax. Then the total revenue collected by the government is $-\sum_{h=1}^H R^h$, and we will need to establish that this is nonnegative in order for this system of lump sum transfers to be feasible.

Denote the equilibrium commodity and factor prices under free trade by (p, w) . With each individual receiving the transfer R^h , the budget constraint is

$$p' c^h \leq w' v^h + R^h. \quad (7.2)$$

We want to choose a pattern of transfers that allows each individual to be as well off as they were in autarky. To this end, consider the following system:

$$R^h = (p - p^a)' c^{ha} - (w - w^a)' v^{ha}. \quad (7.3)$$

To interpret this, if a price for a commodity rises going from autarky to free trade, the government will subsidize each individual by the price rise times the individual's autarky consumption. Conversely, if the earnings of a factor rise going from autarky to free trade, the government will tax each individual by the wage increase times the individual's autarky factor supply. The goal of this transfer system is to ensure that all individuals can still afford their autarky choices (c^{ha}, v^{ha}) . To see that this is indeed the case, substitute these autarky choices into the budget constraint (7.2), and make use of (7.3), to obtain

$$p' c^{ha} \leq w' v^{ha} + (p - p^a)' c^{ha} - (w - w^a)' v^{ha} \Leftrightarrow p' c^{ha} \leq w' v^{ha}.$$

Since the second equation above is just the individual's autarky budget constraint in (7.1), we know that it is satisfied. This proves that the autarky choices (c^{ha}, v^{ha}) are feasible for the budget constraint in (7.2) and transfers in (7.3).

This result can also be seen from figure 7.1. We illustrate there the indifference curves for an individual choosing between a single consumption good c^h and a single factor supply v^h . Under autarky the budget constraint is $p^a' c^h \leq w^a' v^h$, or $c^h \leq (w^a/p^a) v^h$, as shown by the line through the origin with slope (w^a/p^a) , and the individual chooses point A.

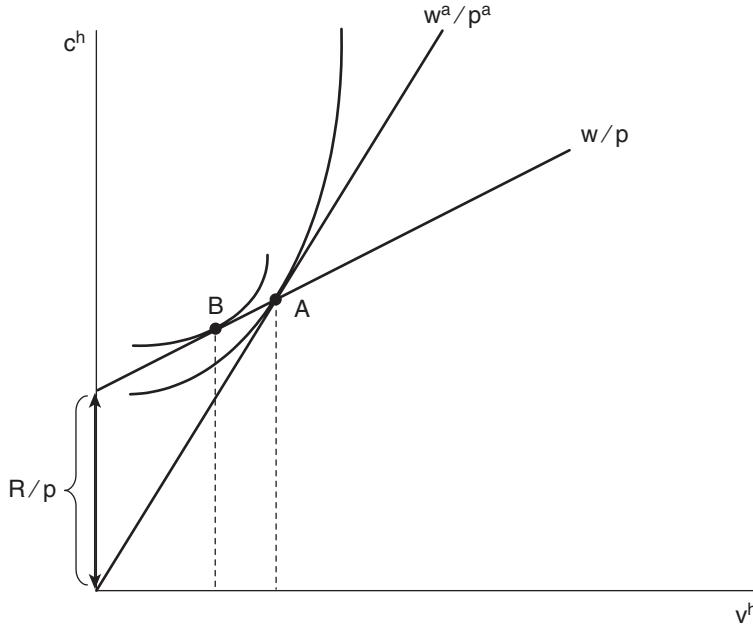


Figure 7.1

Suppose that the free-trade prices lead to a decline in the real wage, $(w/p) < (w^a/p^a)$. Then in order to ensure that the consumer can still afford point A, the government would transfer the amount R/p in terms of the consumption good, leading to the budget constraint starting at that point and rising with slope (w/p) . On this budget constraint the individual can still afford point A, but would prefer to consume at point B. This illustrates the consumer gains available using the lump sum transfers described in (7.3).

Let us now sum up the transfers, and reverse its sign, to obtain total government revenue:

$$\begin{aligned}
 -\sum_{h=1}^H R^h &= (p^a - p)' \sum_{h=1}^H c^{ha} - (w^a - w)' \sum_{h=1}^H v^{ha} \\
 &= (p^a - p)' y^a - (w^a - w)' v^a \\
 &= (p^a y^a - w^a v^a) - (p' y^a - w' v^a) \\
 &= -(p' y^a - w' v^a) \\
 &\geq -(p' y^a - w' v^a) = 0,
 \end{aligned}$$

where the second line follows by using $\sum_{h=1}^H c^{ha} = y^a$ and $\sum_{h=1}^H v^{ha} = v^a$, the third line follows by rearranging terms, and the fourth line follows because autarky profits are zero as in (7.2). The expression $(p' y^a - w' v^a)$ in the fourth line shows the value of profits under free trade if producers chose the feasible but nonoptimal point (y^a, v^a) , which will be *less than* profits at the optimal point (y^a, v^a) , and these profits are zero from constant returns to scale. This establishes the final line.

Therefore, we see that $-\sum_{h=1}^H R^h \geq 0$, and so this system does not cost the government anything: the lump-sum taxes collected from those gaining from trade are more

than enough to cover the subsidies to those harmed by trade. No one is harmed by the move to free trade, and there will generally be gains of two sorts: the *consumption gains* shown in figure 7.1 from consuming at a point different from autarky; and the *efficiency gains* that arise when the value of output ($p'y - w'\nu$) is strictly greater than ($p'y^a - w'\nu^a$), so that the inequality on the last line of (7.4) is strict. To understand this source of gains, consider a special case where the factor supplies chosen by consumers do not change, so that the total factor inputs $\bar{\nu}$ are fixed. Then we have that $w'\nu^a = w'\bar{\nu} = p'y$, so that the revenue collected by the government in (7.4) is

$$-\sum_{h=1}^H R^h = -(p'y^a - w'\nu^a) = p'y - p'y^a. \quad (7.4')$$

In figure 7.2, for example, suppose that the autarky equilibrium is at point A and the free-trade production point at B. Then the revenue collected by the government is exactly the difference between the value of production at these points (with free-trade prices p), or the difference between the two budget lines that are drawn. Figure 7.2 thus illustrates the *production or efficiency gains* due to opening trade, and with this revenue given back to consumers via the lump sum transfers, they enjoy a rise in welfare for that reason in addition to the *consumption gains* illustrated in figure 7.1.

How large are these gains from trade? There are a few historical examples where countries have moved from autarky to free trade, or vice versa, rapidly enough that we can use time-series data to estimate the gains from trade. One such episode in the United States occurred during December 1807 to March 1809, when the U.S. Congress imposed a nearly complete embargo on international trade at the request of President Thomas Jefferson. Irwin (2005) calculates that the welfare cost to the United States was some 8 percent of GDP, which compares to pre-embargo exports of about 13 percent of GDP. Another case was Japan's rapid opening to the world economy in 1859 after some two hundred years of self-imposed autarky. Bernhofen and Brown (2004b) calculate an upper-bound to the gains from trade for Japan of 4 to 5 percent of GDP.

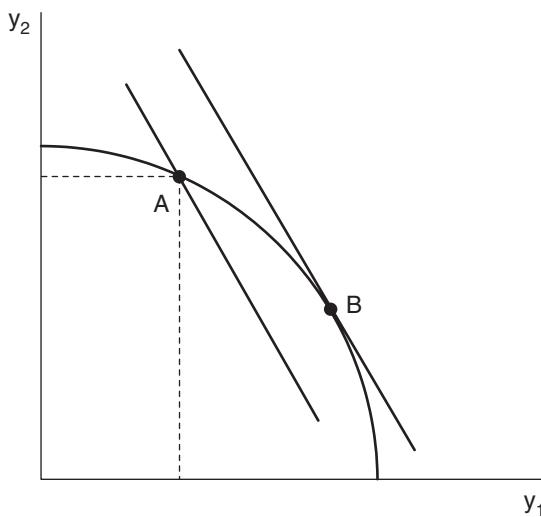


Figure 7.2

As appealing as these results on the gains from trade are, the idea of using lump sum transfers is not realistic because this policy requires too much information to implement. This can be seen from (7.3), where the transfers require knowledge of *each individual's autarky consumption and factor supplies*. Suppose that the government announced a plan to collect this information using a survey and then to implement the transfers in (7.3). Given knowledge of the formula used in (7.3), people would then have an incentive to change their consumption and factor supplies so as to have more beneficial values of (c^{ha}, v^{ha}) recorded in the survey, and therefore get more transfers, thereby leading to a government budget deficit. This is an example of an *incentive compatibility* problem: the attempt to implement a policy would lead individuals to not truthfully report the information needed.

There are two possible solutions to such a problem. The first, which we describe next, is to use policy instruments that do not require individual information. A second solution is to *build into the policy itself* an incentive for people to tell the truth. The transfers in (7.3) certainly will not do this: everyone would claim that they had consumed more of goods whose prices increased from autarky to trade, and had supplied more of factors whose wages fell. But perhaps there is some other system of transfers that would lead to truthful reporting of individual information, and is therefore incentive compatible while at the same time leading to Pareto gains. This topic is explored by Feenstra, McMillan, and Lewis (1990), Feenstra and Lewis (1991a) and Facchini and Willmann (1999), but is beyond the scope of the present chapter.

COMMODITY TAXES AND SUBSIDIES

Dixit and Norman (1980, 79–80) and Dixit (1986) have proposed a system of commodity taxes and subsidies that is designed to achieve Pareto gains from trade, without requiring individual information (or at least, requiring much less information). Specifically, their idea is to allow good and factor prices *for producers* to move to their free-trade levels, but hold good and factor prices *for consumers* fixed at their autarky levels. Let us denote the autarky prices by (p^a, w^a) and the equilibrium prices with free trade (and the commodity tax/subsidies in place) by (p, w) . We let firms face these free trade prices, and set the vector of consumer taxes on goods at $(p^a - p)$ and the vector of consumer subsidies on factors at $(w^a - w)$. It follows that consumers face the price vector (p^a, w^a) , which is the same as in autarky. Therefore, they would make exactly the same decisions as in autarky, and achieve their autarky utility, $u^h(c^{ha}, v^{ha})$.

Next, we check the government budget. The revenue collected from this system of commodity taxes and subsidies equals

$$(p^a - p)' \sum_{h=1}^H c^{ha} - (w^a - w)' \sum_{h=1}^H v^{ha} = (p^a - p)' y^a - (w^a - w)' v^a. \quad (7.5)$$

Note that this is *identical* to the expression in (7.4), and is therefore nonnegative for exactly the same reasons. As we discussed in (7.4') and figure 7.2, this expression is strictly positive whenever there are efficiency gains from having producers face free-trade rather than autarky prices. If this is the case, then the government redistributes this revenue back to consumers as a “poll subsidy” (i.e., an equal amount to each person). In that way, all consumers receive *more than* their autarky utility level, and Pareto gains from trade are achieved!

This system of commodity taxes and subsidies just requires information on the goods and factor prices (p^a, w^a) and (p, w) , which is obviously much less than required to compute the lump sum transfers in (7.3). On the other hand, it should not be thought that the factor price information is always available. If we can think of the quantity of a factor supplied as measuring *effort* rather than just hours worked, then the factor prices need to be measured *per unit of effort*. If effort is not observed, then neither are the true prices or factor supplies. To state this problem differently, a worker who was *guaranteed* his autarky wage even when trade is opened would no longer have an incentive to work hard. Thus, there is again an incentive compatibility problem related to the collection of factor prices whenever individual's effort cannot be observed. A similar problem occurs for workers paid annual salaries rather than wages, so that we observe their income wv^h rather than wages w . In this case, we would have to infer their factor supply (or effort level) v^h before knowing their wage w .¹

A second potential difficulty with this system of commodity taxes and subsidies occurs when individuals have *moving costs*. To model this, suppose that the vector of factor supplies v^h appearing in the utility function is indexed by the *location where the individual works*. That is, if the economy has M primary factors and L locations, the vector v^h is of dimension LM rather than M . An individual who is indifferent toward working in one location versus the other would have indifference curves that are linear between these two factor supplies: when offered a wage slightly higher in one location, the individual would shift entirely there. But if instead the utility function $u^h(c^h, v^h)$ is *strictly* quasi-concave, then this individual would *not* be willing to shift between locations when wages are unchanged. This formulation is a simple way to capture the idea that individuals have preferences over where to work and psychic costs to relocating. Note that it is still consistent with the Arrow-Debreu general equilibrium model, which allows goods or factors to be indexed by location.

What are the implications of this for the system of commodity tax and subsidies proposed by Dixit and Norman (1980)? Suppose that we use the factor subsidies to bring all individuals back to their autarky factor prices. Since v^h is indexed by location, and $u^h(c^h, v^h)$ is *strictly* quasi-concave, this means that individuals have a unique choice of where to work, and are unwilling to shift between locations at the autarky wages. Assume that the various goods in the economy are produced at different locations. Then opening the economy up to trade, but holding wages fixed at w^a means that it is *impossible* for the economy to shift from point A to point B in figure 7.2: if workers are unwilling to move between industry/locations at constant wages, the economy will be stuck at point A despite the change in producer prices. It is appropriate to think of the preferences over locations, combined with wages fixed at their autarky level, as introducing a "kink" into the production possibility frontier at point A , as illustrated by the dashed lines in figure 7.2. Because production no longer changes in response to the free trade prices, there are no efficiency gains, and government revenue in (7.5) is

¹This difficulty occurs especially for specific factors in an industry, where at most we would observe their total income but not their wage. This is an important example because of our earlier assumption that the technology of the economy has constant returns to scale. Suppose instead that the technology had decreasing returns, in which case the revenue earned exceeds the sum of factor payments, so (7.2) does not hold. Then define the difference between revenue earned and factor payments as the income of a hypothetical fixed factor of production, so that including payments to this fixed factor, we have constant returns and (7.2) holds again. Then we would observe only income of this fixed factor, and *not* the separate wage and factor supply.

zero. In this case, the scheme of commodity tax/subsidies proposed by Dixit and Norman fails to achieve any gains.

Feenstra and Lewis (1994) have investigated this case, and argue that *moving subsidies* can restore Pareto gains through subsidizing workers to move between industries/locations. An example of such subsidies in the United States is Trade Adjustment Assistance (TAA). As described by Richardson (1982), this program was established in the United States under the Trade Act of 1974 to compensate workers in industries that experienced unemployment due to import competition. It was used heavily by autoworkers during the U.S. recession of the early 1980s, for example. However, it has often been the case that workers using this program have *not shifted* into new jobs, but instead used TAA as an income supplement during their period of unemployment (see the evidence in Aho and Bayard 1984). This was especially true for the automobile workers laid off in Detroit in 1979 and 1980: many of them simply collected the TAA benefits, and when the automobile industry recovered, they returned to their former jobs. Findings such as this eventually led Congress to realize that the program was not encouraging relocation of workers, and its funding was cut drastically. A much smaller TAA program was later implemented along with NAFTA in 1994, and this program required job retraining (see Irwin 2002, 106–10).

There is some theoretical validity to the idea that a TAA program is worthwhile only if it leads to the relocation of workers between industries. In terms of figure 7.2, the economy achieves production gains of going from point *A* to point *B* only if workers are willing to move between industries/locations; simply giving income supplements to workers but having production remain at point *A* cannot be a self-financing policy. Furthermore, it is essential that workers move in the “right” direction, that is, toward those industries whose international prices have increased, or from point *A* to point *B* in figure 7.2. Feenstra and Lewis (1994) show that a program of moving subsidies can be designed to achieve this goal, so that in conjunction with the commodity taxes proposed by Dixit and Norman, the program achieves Pareto gains and is self-financing.²

As a final example of how the commodity tax/subsidies proposed by Dixit and Norman relate to actual policies, consider the case of the unification of East and West Germany. As described by Akerlof et al. (1991, 1): “At midnight on June 30, 1990, German economic, monetary, and social union occurred: the mark of the German Democratic Republic [East Germany] was replaced by the deutsche mark; trade barriers were lifted; legal, tax, and social insurance systems were harmonized; and all existing barriers to capital and labor movement were removed.” The replacement of the mark with the deutsche mark at par, and the pressure from labor unions to achieve wage parity between the east and west, meant that companies in former East Germany were faced with wages that were far above their market levels. Indeed, Akerlof, et al. (1991) estimate that only 8 percent of former East German conglomerates would have been profitable at those high wages. In the absence of government intervention, it could be expected that wages in the industries located in eastern Germany would fall dramatically, or severe bankruptcy and unemployment would result.

²A richer comparison of trade adjustment assistance with wage and employment subsidies is provided by Davidson and Matusz (2006), in a model where workers search for new jobs and therefore have temporary spells of unemployment. They argue that *temporary* wage subsidies (in industries to which workers should be moving) and *temporary* employment subsidies (in industries that workers have difficulty leaving) are preferable to adjustment assistance and can be used to achieve Pareto gains.

The government attempted to increase wages in the east to levels closer to those in the west in order to avoid migration out of the east. Thus, not only did government policies explicitly allow workers in former East Germany to maintain their standard of living following unification, these policies attempted to achieve an increase in living standards in order to avoid migration. At the same time, it was expected that workers in former West Germany would not be harmed by unification. This historical episode is therefore a perfect example of an attempt to achieve Pareto gains from trade in practice.

How could such gains be achieved in the face of nonmarket clearing wages in East Germany, combined with unemployment? The answer proposed by Akerlof et al. (1991) was to use “flexible employment bonuses” or *wage subsidies* in the east:

So far . . . the package of policies that has been enacted fails to deal realistically with the questions of how to preserve existing jobs, how to speed new job creation, and how to make existing companies viable enough to be privatized. The major problem is that wages in the East are too high for most former state-owned enterprises to cover their costs. High wages also deter new investment. This creates an obvious need for governmental measures to close the gap between the high private cost of labor, caused by high Eastern wages, and the low marginal product of labor, caused by outmoded capital and technology. We propose a program of self-eliminating flexible employment bonuses . . . to eliminate this gap. Our analysis shows that such a program would give many workers a chance to keep their jobs and would also raise the level of new job creation through faster private investment. According to our estimates, even deep wage subsidies (for example, an employment bonus equal to 75 percent of current wages) would have very low budgetary costs. They might even *reduce* the budget deficits—largely for the same reason that infrastructure investment is not costly: the government is already committed to a high level of income support even if workers are unemployed.

Thus, by subsidizing workers in East Germany the government could save enough income from unemployment benefits that the policy could be self-financing. This is very similar to the commodity tax/subsidies proposed by Dixit and Norman (1980), and shows how wage subsidies have a potentially important role to play in an actual attempt to achieve Pareto gains from trade. Unfortunately, the wage subsidy proposal of Akerlof et al. (1991) was not followed in Germany and migration from east to west continues to be a policy issue.³

PARTIAL REFORM OF TARIFFS

We now consider a country that undertakes a partial reform of its tariff structure, moving from one situation of restricted trade (possible autarky) to another. This is a very realistic example, since countries undertake partial reform of their tariff frequently (though it is often done *in conjunction* with tariff reforms abroad). The question is this: are individuals better off in a country or not? To answer the question, we make use of the lump sum transfers as a means to compensate individuals for this move, so we will need to check whether the government budget constraint allows for this compensation or not. While the use of lump sum transfers is not realistic, the reader should think of this as simply a method to evaluate the desirability of the tariff

³Interestingly, Phelps (1997) also proposes a system of wage subsidies to offset the drop in wages for low-skilled workers in the United States, whose wages have fallen during the 1980s and 1990s, as discussed in chapter 4.

reform, that is, if the tariff reform combined with lump sum transfers *could* make everyone better off, then we accept the tariff reform itself as a worthwhile policy, even when the transfers are not made. This criterion for evaluating a reform is called the “compensation principle” (see Chipman 1987). Our analysis will follow that in Grinols and Wong (1991) and Ju and Krishna (2000a), though the latter authors also show how the results can be extended to non-lump sum transfers, using commodity tax and subsidies as in Dixit and Norman (1980) and Dixit (1986).⁴

To set up our notation, let the domestic price vector for consumers and producers be denoted by $p = p^* + t$, where p^* is the vector of world prices, and t is the vector of trade taxes and subsidies. Notice that $t_i > 0$ for an imported good indicates a tariff (raising the domestic price above the world price), whereas $t_i < 0$ for an imported good indicates a subsidy. Conversely, $t_i > 0$ for an exported good indicates a subsidy (since firms receive $p_i^* + t_i$ from exporting, so the domestic price would also have to rise by the same amount), while $t_i > 0$ for an exported good indicates a tax. It follows that the total revenue collected by the government from the trade tax/subsidies is $t'(c - y)$, where $c = \sum_{h=1}^H c^h$ is total consumption and y is the output of the economy at the prices p . For convenience, let us denote the *import* vector as $m = c - y$, so that revenue collected is simply $t'm$.

We compare two different trade tax/subsidies t^0 and t^1 , with the equilibrium domestic price vectors $p^0 = p^{*0} + t^0$ and $p^1 = p^{*1} + t^1$. We stress that the world price p^* changes in these two situations, and in general, the world prices are endogenously determined by the tariffs chosen at home. More generally, our setup allows for a change in tariffs abroad simultaneously with those at home, as part of a bilateral or multi-lateral tariff reform.

Our argument will follow the earlier discussion of lump sum compensation quite closely. In the initial situation, we suppose that consumers maximize utility subject to their budget constraints:

$$p^{0'} c^{h0} \leq w^{0'} v^{h0} + (t^{0'} m^0)/H, \quad (7.6)$$

where p^0 is the autarky vector of commodity prices, w^0 is the autarky vector of factor prices, and we include the redistributed tariff revenue $t^{0'} m^0 / H$, where H is the number of consumers. The resulting utility for each consumer is $u^h(c^{h0}, v^{h0})$.

Summing over all consumption and factor supplies, the total consumption in the economy is therefore $\sum_{h=1}^H c^{h0} = c^0$, and the total factor inputs are $\sum_{h=1}^H v^{h0} = v^0$. Firms will be maximizing profits at the equilibrium prices (p^0, w^0) , resulting in the outputs y^0 and factor demands v^0 . Again, we assume constant returns to scale, so that the resulting profits are zero, $p^{0'} y^0 - w^{0'} v^0 = 0$. Imports in this equilibrium are $m^0 = c^0 - y^0$, so the revenue collected is $t^{0'} m^0$, which has been transferred back to consumers in (7.6).

After the tariff reform, domestic prices change to $p^1 = p^{*1} + t^1$, and the revenue collected from the trade tax/subsidies becomes $t^1 m^1$. We denote the transfer to consumers by R^h , which may differ from just the revenue $t^1 m^1$. The budget constraint for consumers is

$$p^{1'} c^h \leq w^{1'} v^h + R^h. \quad (7.7)$$

⁴See problems 7.2 and 7.3, provided by Jiandong Ju. An alternative approach to the piecemeal reform of tariffs, relying on small reforms and using calculus, is provided in Diewert, Turunen-Red, and Woodland (1989); Ju and Krishna (2000b); and Turunen-Red and Woodland (1991, 2000).

Our goal is to identify conditions under which the government can compensate individuals for the change in prices, and at the same time, balance the government budget. Using a similar system of lump sum transfers that we considered earlier, this can be achieved by the following theorem.

THEOREM (GRINOLS AND WONG 1991; JU AND KRISHNA 2000A)

When the prices change from $p^0 = p^{*0} + t^0$ to $p^1 = p^{*1} + t^1$, suppose that the government transfers the following amount to each individual:

$$R^h = (p^1 - p^0)' c^{h0} - (w^1 - w^0)' v^{h0} + (t^0' m^0)/H. \quad (7.8)$$

Provided that $(p^{*0} - p^{*1})' m^0 + t^1'(m^1 - m^0) \geq 0$, then no individual is worse off and the government budget is balanced.

Proof:

Substituting the transfers in (7.8) into individual's budget constraint (7.7), we can confirm that the initial choices (c^{h0}, v^{h0}) are still feasible:

$$\begin{aligned} p^1' c^{h0} &\leq w^1' v^{h0} + (p^1 - p^0)' c^{h0} - (w^1 - w^0)' v^{h0} + (t^0' m^0)/H \\ &\Leftrightarrow p^0' c^{h0} \leq w^0' v^{h0} + (t^0' m^0)/H. \end{aligned} \quad (7.9)$$

Since the second equation shown is just the initial budget constraint in (7.6), the initial choices (c^{h0}, v^{h0}) are feasible and utility is at least $u^h(c^{h0}, v^{h0})$.

To check the government budget, we subtract the transfers from the tariff revenue, obtaining

$$\begin{aligned} t^1' m^1 - \sum_{h=1}^H R^h &= t^1' m^1 - (p^1 - p^0)' \sum_{h=1}^H c^{h0} + (w^1 - w^0)' \sum_{h=1}^H v^{h0} - t^0' m^0 \\ &= (p^0 - p^1)' c^0 - (w^0 - w^1)' v^0 + t^1' m^1 - t^0' m^0 \\ &= (p^0 - p^1)' y^0 - (w^0 - w^1)' v^0 + (p^0 - p^1)' m^0 + t^1' m^1 - t^0' m^0 \\ &= (p^0' y^0 - w^0' v^0) - (p^1' y^0 - w^1' v^0) + (p^{*0} - p^{*1})' m^0 + t^1' (m^1 - m^0) \\ &\geq (p^{*0} - p^{*1})' m^0 + t^1' (m^1 - m^0). \end{aligned} \quad (7.10)$$

The second, third, and fourth equalities follow from simple algebra, and the inequality on the final line follows because initial profits $(p^0' y^0 - w^0' v^0)$ are zero, while $(p^1' y^0 - w^1' v^0)$ is less than $(p^1' y^1 - w^1' v^1)$, which is also zero. Provided that the final line of (7.10) is nonnegative, then so is the government budget. QED

This theorem provides us with two expressions that summarize the welfare impact of tariff reforms. The first of these is $(p^{*0} - p^{*1})' m^0$, which indicates the impact of the reform on the *terms of trade*, or international prices: if the price of imports goes up or the price of exports goes down, this expression will be negative and gains cannot be assured. The second term is $t^1' (m^1 - m^0)$, which is interpreted as the *change in imports*

evaluated at the *final tariff vector*. The tariff vector reflects the difference between domestic and international prices: a positive value indicates that marginal costs at home exceed international prices, so it would be more efficient for the country to import the good; and if the value is negative, the country should be exporting the good. Therefore the term $t^*(m^1 - m^0)$ gives a measure of the efficiency gain (if positive) due to attracting imports toward protected sectors. Thus, policies to raise welfare must avoid adverse terms of trade or efficiency effects to ensure Pareto gains.

Note that the term $t^*(m^1 - m^0)$ can also be interpreted as the *change in tariff revenue* evaluating at the *final tariff vector*. In some cases, this will equal the overall change in tariff revenue. For example, suppose that a small country starts with prohibitive tariffs on some goods, and then moves to free trade in those goods. Since there is zero tariff revenue before and after the change, the term $t^*(m^1 - m^0)$ shows what happens to revenue collected on the *other* goods, whose tariffs are unchanged. If tariff revenue falls, then we cannot be sure that the country gained from the partial movement toward free trade. Working with a one-consumer small economy, Ohyama (1972) was the first to identify the sufficient condition that tariff revenue (evaluated at the final tariff vector) must rise to ensure gains.

Two special cases of the above theorem can be noted:

COROLLARY

Applying the transfers to individuals as in (7.8):

- (a) Starting in autarky and moving to trade with tariffs and subsidies makes no one worse off provided that $t^*m^1 \geq 0$;
- (b) For small country with constant world prices, $p^{*0} = p^{*1}$, a change in tariffs and subsidies makes no one worse off provided that $t^*(m^1 - m^0) \geq 0$.

In the first case, we set $m^0 = 0$ in the above theorem, and see that the criterion to achieve Pareto gains is just $t^*m^1 \geq 0$, that is, the tariffs and subsidies must raise non-negative revenue. Ohyama (1972) describes this condition as “self-financing” tariffs. The result in corollary (a) is interpreted as saying that “some trade is better than no trade.” If we suppose that the second situation has free trade, $t^1 = 0$, then the condition in corollary (a) is clearly satisfied and this result is interpreted as “free trade is better than autarky” (Samuelson 1962; Kemp 1962). Corollary (b) considers a “small country” facing fixed world prices and shows that a tariff reform will raise welfare provided that the efficiency effects are nonnegative. Again, if we suppose that the second situation has free trade, $t^1 = 0$, then the condition in corollary (b) is clearly satisfied, and this result can be interpreted as “free trade is better than restricted trade for a small country.” We will provide another demonstration of this result in the next chapter.

REGIONAL TRADE AGREEMENTS

We now extend our results to the case where a group of countries decides to pursue free trade internally, while maintaining tariffs against the rest of the world. Under a customs union, the countries involved choose a *common* external tariff with the rest

of the world, whereas under a free trade area the countries maintain different tariffs on imports from the rest of the world. Let us consider the customs union case first, since it is easier.

The analysis of customs unions dates back to Viner (1950), who introduced the terms *trade creation* versus *trade diversion*. Trade creation refers to a situation where two countries within the customs union begin to trade with each other, whereas formerly they produced the good in question for themselves. In terms of corollary (a) above, the countries go from autarky (in this good) to trading with zero tariffs, and they both gain. Trade diversion, on the other hand, occurs when two countries begin to trade within the union, but one of these countries had formerly imported the good from *outside* the union. We may presume that the importing country formerly had the same tariffs on all countries, but purchased from *outside* the union because that price was lowest. After joining the union, the country *switches* its purchases from the lowest-price to a higher-price country, so in terms of corollary (b) there is a negative efficiency effect, and the country could lose from joining the union.

To verify these results more formally, we need to expand our notation to incorporate the customs union. In our analysis above we implicitly assumed that each country i had a single tariff vector t^i applied against *all other* countries. This is no longer the case: countries in the union will face zero tariffs among themselves, and still maintain the same external tariff with respect to outside countries. An easy way to capture this is to *index goods by their region of origin and their region of destination*. That is, let the first N goods denote those produced in the union and sent to the union; the next N goods are produced outside the union and sent outside the union; the next N are produced in the union and sent outside the union; and the last N are produced outside the union and sent into the union. In total we therefore have $4N$ goods, but only the last $2N$ of these involve trade between the union and outside countries. This treatment of goods still is consistent with the Arrow-Debreu model, where goods can be indexed by location.⁵

With this notation, the above theorem and corollary continue to hold. For a union country i , the first N tariffs are zero, the second N tariffs are irrelevant (since these goods are traded between nonunion countries), and the last $2N$ tariffs are *equalized* for all countries of the union. We will denote this post-union tariff vector by t^i . Notice that with zero tariffs on countries within the union, the term $t^i(m^{ii} - m^{i0})$ measures the *change in tariff revenue for country i's imports from outside the union*. Trade diversion means that country i switches from the lowest-price supplier from outside the union (where tariff revenue is collected) to another supplier within the union (with no tariff revenue), so that $t^i(m^{ii} - m^{i0}) < 0$ and tariff revenue has fallen. Therefore, country i can be worse off. This result illustrates what is called a “second best” problem: by eliminating its tariffs with all other countries, a small country would necessarily be better off, but if it goes only part way toward this goal (eliminating tariffs with the union countries only), there is no guarantee of gains.

This outcome makes it seem like no general result on the desirability of customs unions is possible. Remarkably, Kemp and Wan (1976) were still able to obtain such a result. Their goal was to ensure that the customs union benefits its members, and also does no harm to the rest of the world. To achieve this, Kemp and Wan used the following rule: *the customs union should keep the world price p^* fixed, or equivalently,*

⁵Note that if consumers are indifferent toward a good coming from two locations, the indifference curves between these locations are linear.

keep the purchases from the rest of the world x^ fixed.* This rule can be contrasted with the criterion used in the WTO on customs unions. In Article XXIV in table 7.1, we see that customs unions are “not prevented” provided that “the duties [with outside parties] shall not on the whole be higher or more restrictive than the general incidence of the duties . . . prior to the formation.” Thus, the criterion used in WTO is that the *tariffs with respect to countries outside the union* should not be higher than before, whereas the rule used by Kemp and Wan is that *trade with respect to outside countries* is not affected by the formation of the customs union. The Kemp-Wan rule is the correct one to use for avoiding losses, not the existing WTO criterion.

To develop the result of Kemp and Wan, let us denote the countries in the customs union by $I=1, \dots, C$. Prior to the union they have tariffs t^{i0} . We let p^{*0} denote the world prices in this equilibrium, and m^{i0} denote the import vector of country i . Let the net export vector of the rest of the world be denoted by $x^*(p^{*0})$. Note that this vector is zero for the first $2N$ goods (which refer to trade strictly inside or outside the union), and potentially nonzero otherwise. Summing imports over all countries in the union, the equilibrium condition is

$$\sum_{i=1}^C m^{i0} = x^*(p^{*0}). \quad (7.11)$$

After joining the customs union the member countries have a unified tariff t^1 , the elements of which are zero for purchases from within-union countries. Let m^{i1} denote the vector of imports of country i in this situation. Kemp and Wan (1976) prove that there exists a unified tariff t^1 , which ensures that

$$\sum_{i=1}^C m^{i1} = x^*(p^{*0}). \quad (7.12)$$

The fact that the prices p^{*0} and exports from the rest of the world $x^*(p^{*0})$ are *unchanged* due to the customs union guarantees that the *rest of the world is not affected by it*. The question now is whether the members of the customs union also gain, and whether such a gain requires some transfers between the countries. This is answered by the following theorem.

THEOREM (KEMP AND WAN 1976; GRINOLS 1981)

Let the countries $I=1, \dots, C$ form a customs union, with the common external tariff such that prices and purchases from the rest of the world are unchanged. Suppose that each country within the customs union receives the transfers $-t^1(m^{i1} - m^{i0})$ from the rest of the union. Then there is a pattern of lump sum transfers within each country such that no individual is worse off, and the government budget in each country is nonnegative.

Proof:

Consider the transfers described in (7.8), but applied within each country. It follows that all consumers in every country of the union can still afford their previous consumption bundles. Furthermore, noting that prices p^* from the rest of the world are fixed, then from (7.10) the government budget in each country is no less than

$-t^{i1}(m^{i1} - m^{i0})$. Therefore, each country can afford to give up this amount (or receive it if negative), and have a nonnegative budget. It remains to be shown that these transfers within the union are self-financing, that is, $\sum_{i=1}^C t^{i1}(m^{i1} - m^{i0}) = 0$. But this follows immediately from the fact that total union imports are equal in (7.11) and (7.12): $\sum_{i=1}^C m^{i1} = \sum_{i=1}^C m^{i0}$. QED

It is apparent that the way losses are avoided for the union is by keeping total imports fixed: in that way, the total tariff revenue (evaluated at post-union external tariffs) is also fixed. But keeping total imports fixed also means that world prices are fixed, which eliminates any terms of trade impacts for the union and the rest of the world. Thus, keeping total union imports fixed immediately satisfies the criterion identified in the earlier theorem: when summed over the countries of the union, the term $(p^{*0} - p^{*1})'m^{i0} + t^{i1}(m^{i1} - m^{i0})$ is zero, and so the union as a whole is no worse off. An individual country may experience a loss in its own tariff revenue of $t^{i1}(m^{i1} - m^{i0})$, but this could be compensated by transfers from other countries within the union.

It has been more than twenty-five years since Kemp and Wan demonstrated this result, and in the meantime, regional agreements under GATT Article XXIV have proliferated. Many of these have not been customs unions, but instead, are free trade areas (FTAs). Is it possible to extend the Kemp-Wan theorem to cover these FTAs? As explained above, FTAs differ from customs unions in that the countries involved have their own tariff vector with respect to the rest of the world. At the same time, tariffs are zero within the union, so goods can be sent between these countries freely. This means that *producer prices* must be equalized within the union, but by virtue of their differing tariffs on outside countries, *consumer prices* can differ within the union. It follows that the method of proof used by Kemp and Wan (1976) cannot be used for free-trade areas. Krishna and Panagariya (2002) have developed a new proof, however, which shows that Kemp-Wan theorem can indeed be extended to cover FTAs. Their approach can be described quite easily using our notation above.

Let $x^*(p^{*0})$ denote the initial net exports from the rest of the world to the countries planning to form an FTA. Recall that this vector is zero for the first $2N$ goods (which refer to trade strictly inside or outside the union), and potentially nonzero otherwise. The approach of Krishna and Panagariya is to ensure that *every country in the FTA* engages in exactly the same trade with the rest of the world as it did initially. Letting m^{i0} and m^{i1} denote the imports of country $i = 1, \dots, C$, which is a member of the FTA, we require that

$$m_j^{i0} = m_j^{i1}, \text{ for goods } j = 2N + 1, \dots, 4N. \quad (7.13)$$

Thus, we require that every FTA member import and export the same quantity with the rest of the world as it did before joining the free-trade area. While Krishna and Panagariya do not provide a general existence result, that was provided by Grinols and Silva (2007), so we can be confident that there are country-specific tariffs ensuring that (7.13) holds and we can denote these by $t^{i1}, i = 1, \dots, C$. Notice that (7.13) implies that (7.12) still holds, that is, *total* net exports from the FTA to the rest of the world equals $x^*(p^{*0})$, the value initially. So we conclude that the equilibrium world prices are still given by p^{*0} .

Since trade with the outside world is not affected by the FTA, and neither are prices, we have obtained the result of the following theorem.

THEOREM (KRISHNA AND PANAGARIYA 2002; GRINOLS AND SILVA 2007)

Let the countries $i=1,\dots,C$ form a free-trade area, with external tariffs t^{il} for each country such that prices and purchases from the rest of the world are unchanged. Then there is a pattern of lump sum transfers within each country such that no individual is worse off, and the government budget in each country is nonnegative.

Proof:

Again we use the transfers described in (7.8) applied within each country. It follows that all consumers in every country of the FTA can still afford their previous consumption bundle, and from (7.11), the government budget in each country is no less than $t^{il}(m^{il} - m^{i0})$. The first N elements of t^{il} are zero due to free trade within the FTA, and the next N elements of m^{i0} and m^{il} are zero because those goods are produced and delivered outside the FTA. Then (7.13) ensures that $t^{il}(m^{il} - m^{i0}) = 0$, so the budget is nonnegative in every country $i=1,\dots,C$. QED

One feature of FTAs and the above theorem should be emphasized. Because tariffs with respect to the rest of the world differ among the member countries, outside countries would have an incentive to export their goods into the *lowest-tariff* country of the FTA, and then transship it to other member countries. This type of activity needs to be ruled out to obtain the above theorem. That is, transshipment through the lowest-tariff country would *violate* condition (7.13), which specifies that each country must import the same quantity from outside countries as before the FTA. More generally, transshipment through the lowest-tariff country would lead to a drop in tariff revenue for other countries of the FTA, violating the condition to obtain Pareto gains. To ensure welfare gains, therefore, we need to rule out transshipment of goods not produced in the FTA.

In practice, restrictions on such transshipment within FTAs are known as *rules of origin* (ROO). Within NAFTA, for example, each and every good must be certified as “North American made” before it can be shipped freely between Mexico, the United States, and Canada. This certification is typically made based on the percentage of value added coming from North American production, or the use of some key input produced in North America. This prevents a foreign country from exporting into Canada, for example, and then sending the good to the United States or Mexico, unless it had enough further processing to certify it as “North American made.” Such rules are enormously complex, since they require a specification of the production process for each and every good. Due to their complexity and the fact that the ROO are an artificial restriction on the movement of goods, they can have unexpected consequences (see Krishna and Krueger 1995; Krishna 2003). But the theorem of Krishna and Panagariya sheds new light on these practices, since the welfare-improving FTA depends on their use.⁶

⁶Let us be more careful in this statement. Starting at any FTA, the theorem of Kemp and Wan ensures that there is a customs union that is Pareto superior. But that customs union is not achieved by simply dropping the ROO from the FTA; rather, it would require dropping the ROO, changing the external tariffs for all countries to a common level that ensures the same total imports from the rest of the world, and also implementing income transfers across countries within the union. If these actions are politically infeasible, then the ROO are needed in order for the FTA to ensure Pareto gains, as in the theorem of Krishna and Panagariya.

EVIDENCE ON REGIONAL AGREEMENTS

The above theorems allow us to conclude that welfare-improving customs unions and FTAs exist, but from this result we should not infer that regional trade agreements in practice are necessarily a good thing. The possibility of trade diversion identified by Viner (1950), and any changes in the terms of trade need to be evaluated empirically before judging actual agreements. More generally, we should be concerned with whether regional trade agreements *help* or *hinder* the movement toward global free trade through multilateral negotiations under the WTO. This issue of political economy will be discussed in chapter 10.

Britain and the EEC

Grinols (1984) was among the first to measure the gains or losses from membership in a customs union. He examined Britain's joining the European Economic Community in 1973. This meant that Britain had to adopt the common external tariff of the EEC. Not only would this tariff affect domestic price p^i in Britain, but also it could be expected to have an impact on the foreign prices p^{*i} facing Britain, that is, on its terms of trade. We are recognizing here that the prices charged by foreign exporting firms are often responsive to a change in tariffs by just one importing country. This occurs when foreign firms segment their export markets, charging different prices in each. This behavior is related to imperfect competition, as will be discussed in the next chapter.

Grinols measured the prices for British exports and imports during the years after its membership in the EEC. Multiplying these prices by initial quantities gives $(p^{*i0} - p^{*i1})'m^{i0}$, or the value of changes in the terms of trade. In every year from 1973 to 1980, it turned out that the British terms of trade had fallen from 1972, and the amount of this decline averaged about 2.3 percent of British GDP. (This amount includes the lost tariff revenue on imports, which automatically become property of the customs union, but the average drop in the terms of trade without including tariff revenue was still 1.7 percent of GDP.)

In order to compensate for this loss, Britain would have needed to obtain some transfer from the EEC, that is, the amount $t'(m^{i1} - m^{i0})$ in our discussion above. There was a wide range of transfers between countries as part of EEC membership, including the regional development funds (allocated to projects in lower-income regions), special funds in agriculture, coal, and steel, value-added taxes, and so on. The net transfers were slightly in Britain's favor, but not enough to offset the terms of trade loss, so that summing $(p^{*i0} - p^{*i1})'m^{i0} + t'(m^{i1} - m^{i0})$ still gives an average loss for Britain of 2.26 percent of GDP annually.

We conclude that the sufficient condition to ensure Britain's gain from joining the EEC, that is, $(p^{*i0} - p^{*i1})'m^{i0} + t'(m^{i1} - m^{i0}) > 0$, did not hold. This does not necessarily imply that Britain was worse off, however, since it still enjoys free trade within the EEC, and so has had efficiency gains on both the consumption and production side for that reason. In figure 7.1, for example, these gains are illustrated by the movement from point A to point B. Grinols attempts to evaluate these gains for Britain and finds that they are quite small. The conclusion is that its membership in the EEC in 1973 led to a welfare loss that averaged about 2 percent of GDP annually through 1980.

One criticism that can be made of Grinols's methodology is that he simply accepts the trade prices as data, and presumes that any change from 1972 represents the effect of Britain's membership in the EEC. Thus, the terms of trade decline is *attributed*

to Britain's changing its external tariff to the common EEC level. This does not allow for any exogenous changes in the terms of trade due to factors affecting supply and demand on world markets. To control for such factors, we need to specify prices as functions of the tariffs, as well as other relevant factors. Using a log-linear regression to estimate prices as a function of tariffs and other factors allows us to *infer* the specific impact of the tariff change. This is the approach taken by Chang and Winters (2002) for MERCOSUR.

MERCOSUR

Specifically, Chang and Winters (2002) consider prices from various countries (United States, Japan, German, etc.) to Brazil, which is the largest market in MERCOSUR. This customs union had the effect of eliminating tariffs between Brazil and Argentina, Paraguay and Uruguay, as well as equalizing their external tariffs. Of interest, then, is the *direct* effect of the change in the external tariff of Brazil on the prices charged by foreign firms; as well as the *indirect* effect of the elimination of the tariffs between Brazil and Argentina on the prices charged by foreign firms to Brazil. Both direct and indirect effect will lead to a terms of trade change for Brazil, as well as for the countries selling in that market. The regression framework used to measure these effects draws on Feenstra (1989), as discussed in detail in the next chapter, and specifies prices as a log-linear function of tariffs and other factors. By pooling data across disaggregate products and years, Chang and Winters can use annual fixed effects to control for any changes in prices that are common across commodities.

For the case of MERCOSUR, Chang and Winters find that many of the prices charged by external countries, including the United States, Japan, Germany, Korea, and Chile, indeed fell as a result of the change in Brazilian tariffs. The *direct* impact of the change in the Brazilian external tariff is smaller than the *indirect* effect of the elimination of its tariff with Argentina. On average, foreign prices charged to Brazil fell by one-third as much as the drop in the tariff with Argentina (and by even more if yearly fixed effects are used). This implies a significant terms of trade gain for Brazil and loss for the exporting countries. For example, Chang and Winters evaluate this loss in 1996 as \$624 million for the United States (which exceeds 10 percent of its export sales to Brazil of \$5.4 billion in 1991), \$59 million for Japan, \$236 million for Germany, \$14 million for Korea, and \$17 million for Chile. These figures illustrate the potential losses that a customs union can impose on the countries left out.⁷

CUSFTA and NAFTA

For the North American Free Trade Agreement (NAFTA) of 1994 and the earlier Canada-U.S. Free Trade Agreement (CUSFTA) of 1989, Clausing (2001) and Romalis (2007) estimate the impact on trade quantities rather than prices, using highly disaggregate data for the United States. Clausing estimates the impact of CUSFTA on U.S. imports from Canada over 1989–94, and on U.S. imports from the rest of the world. These trade quantities are modeled as log-linear function of tariffs and other factors,

⁷Schiff and Chang (2003) extend the work of Chang and Winters (2002) to examine whether the *possibility* of Argentine exports to Brazil (as indicated by its exports to another country in the same product category) served to lower the foreign prices charged in Brazil. They find evidence in support of this “contestability” effect.

and pooling across commodities and years allows annual fixed effects to be used. Romalis examines where the United States sourced its imports of some five thousand commodities, and compares this information with where the European Union sourced the same commodities, using a “differences in differences” technique to identify the impact of both CUSFTA and NAFTA.

Clausing finds that U.S. imports from Canada responded predictably to CUSFTA, with a 10 percent drop in the U.S. tariff leading to a 1 percent expansion of its imports from Canada. This is an example of trade creation between the countries. At the same time, the drop in U.S. tariffs had a negative but statistically *insignificant* impact on its imports from the rest of the world, so there is little evidence to support trade diversion. This conclusion suggests that the Canada-U.S. Free Trade Agreement was relatively benign for other countries, but should be interpreted with caution: to make welfare conclusions, it is better to estimate the effect of tariff changes on *prices*, as done by Chang and Winters (2002), rather than quantities. Romalis also finds a significant impact of CUSFTA on U.S. imports from Canada, along with a large impact of NAFTA on U.S. imports from Mexico. However, he suggests that NAFTA may have led to substantial trade diversion, because the largest tariff reductions were often in commodities where the United States relied on imports from outside of North America.

Direct evidence on the role of trade creation versus trade diversion in CUSFTA and NAFTA is provided by Trefler (2004) and Caliendo and Parro (2015). Drawing on Krishna (2003), Trefler develops a simple condition that must be satisfied in order for trade creation in Canada under CUSFTA to dominate trade diversion, and he finds that this condition indeed holds. Then he checks for evidence that U.S. prices rose in Canada—which would be a terms of trade loss for that country—and finds no evidence to support this. Putting together these two results, he concludes that Canada gained from the free trade agreement with the United States. Trefler also investigates the impact of CUSFTA on wages, employment, and productivity in Canada, as discussed in chapter 5.

Caliendo and Parro (2015) examine the effect of NAFTA on Canada, Mexico, and the United States. Their method differs from that of Trefler, in that rather than examining data from these countries before and after NAFTA, they instead calibrate a general equilibrium model to the pre-NAFTA data and then simulate the impact of the tariff cuts on the terms of trade and on trade creation and diversion. Like Trefler finds for Canada under CUSFTA, Caliendo and Parro conclude that trade creation exceeded trade diversion for all three countries as a result of NAFTA. However, the tariff cuts led to terms of trade losses for both Canada and Mexico (though not for the United States); and in the case of Canada, the terms of trade losses combined with those from trade diversion were enough to overpower the gains from trade creation. In other words, Canada ended up losing from NAFTA while the United States and Mexico both gained. The terms of trade loss for Canada is traced to a falling price of exports in a few key sectors, namely, autos, other transport equipment, and basic metals.

It would certainly be of interest to compare the predictions from the general equilibrium model used by Caliendo and Parro to the change in the actual prices of Canadian exports under NAFTA, to verify the accuracy of the simulated results. The general equilibrium model that they use is based on the Eaton and Kortum (2002) framework, but they extend their model to incorporate input-output linkages between industries, trade in intermediate inputs, and, of course, tariff changes. Dekle, Eaton, and Kortum (2008) were the first to show how that model can be expressed

in rate-of-change form—analogous to the Jones algebra that we studied in chapter 1, but instead expressing the equilibrium conditions in terms of one equilibrium relative to another—which allows the *change* between equilibria to be computed without knowledge of all the model parameters. This parsimonious solution method due to Dekle, Eaton and Kortum (2008) is adapted by Caliendo and Parro (2015) to analyze the change in tariffs under NAFTA and has many other applications.

IMPERFECT COMPETITION AND INCREASING RETURNS TO SCALE

An omitted topic in the chapter thus far is the role of imperfect competition. This raises two distinct issues. First, in industries with monopolistic pricing there is a gap between price and marginal costs. Because prices reflect consumer valuations of a product, this gap indicates that consumers would be better off by *increasing* the output of that industry. If this occurs due to the opening of trade, we might expect that the economy is better off; but if trade reduces the outputs of industries with imperfect competition, this seems to indicate potential losses.

Second, models of imperfect competition often have increasing returns to scale, such as in the monopolistic competition models we studied in chapters 5 and 6. The presence of increasing returns creates a reason to be concerned about losses from trade. Many years ago, Graham (1923) argued that with increasing returns an industry facing strong import competition might be forced to contract, leading to higher average costs, and that this would be a reason for protection. Ethier (1982a) has confirmed this result in a model relying on increasing returns that are *external* to the firm, finding that the *small country* may experience losses from trade due to the contraction of its increasing-returns industry. Interestingly, Markusen (1981) finds that the monopolistic firm in the *large country* might be forced to shrink due to trade, so the large country potentially faces losses. So once again, a sufficient condition for gains from trade under increasing returns appears to be that the increasing-returns industries do not contract. We will derive a condition that is weaker than this using the analysis of Grinols (1991), and then summarize other results that apply under monopolistic competition, due to Helpman and Krugman (1985, chapter 9).

In contrast to the perfectly competitive case used earlier in the chapter, we will now need to be more specific about the structure of industry production and costs. With industry i output denoted by y_i , we will suppose that total minimum costs are $C_i(y_p, w)$. This formulation rules out having several goods produced jointly within an industry, and also rules out having firms in an industry produce with differing costs. But we allow for constant or increasing returns to scale and also allow for zero or positive profits in each industry.⁸ Define $AC_i(y_p, w) \equiv C_i(y_p, w)/y_i$ as average costs, so that under increasing returns to scale $AC_i(y_p, w)$ falls as output rises.

Total profits earned by firms in the autarky equilibrium are $p^{a'}y^a - w^{a'}v^a$, and each consumer is entitled to some fixed fraction λ^h of these. Then the autarky budget constraint of consumer h is rewritten from (7.1) as

$$p^{a'}c^{ha} \leq w^{a'}v^{ha} + \lambda^h(p^{a'}y^a - w^{a'}v^a). \quad (7.14)$$

⁸The case of decreasing returns to scale can be handled by defining an artificial factor of production that transforms the technology into constant returns to scale, as in note 1 of this chapter.

Let us now open the economy to trade at the prices p , and introduce a system of lump sum transfers, so the post-trade budget constraint of each consumer is

$$p'c^h \leq w'v^h + R^h + \lambda^h(p'y - w'v), \quad (7.15)$$

and the lump sum transfers to each individual are

$$R^h = (p - p^a)'c^{ha} - (w - w^a)'v^{ha} + \lambda^h(p^a'y^a - w^a'v^a) - \lambda^h(p'y - w'v). \quad (7.16)$$

It is straightforward to show that these transfers will allow all individuals to obtain at least their autarky consumption bundle, so that they are no worse off. The question, then, is whether the government can afford this policy.

Summing the transfers in (7.16), and reversing the sign, we obtain

$$\begin{aligned} -\sum_{h=1}^H R^h &= (p^a - p)' \sum_{h=1}^H c^{ha} - (w^a - w)' \sum_{h=1}^H v^{ha} - \sum_{h=1}^H \lambda^h [(p^a'y^a - w^a'v^a) - (p'y - w'v)] \\ &= (p^a - p)'y^a - (w^a - w)'v^a - (p^a'y^a - w^a'v^a) + (p'y - w'v) \\ &= (p'y - p^a'y^a) + \left[w'v^a - \sum_{i=1}^N C_i(y_p, w) \right] \\ &= (p'y - p^a'y^a) - \sum_{i=1}^N [C_i(y_p, w) - C_i(y_i^a, w)] + \left[w'v^a - \sum_{i=1}^N C_i(y_i^a, w) \right], \end{aligned} \quad (7.17)$$

where the second equality follows by using $\sum_{h=1}^H c^{ha} = y^a$ and $\sum_{h=1}^H \lambda^h = 1$, the third equality follows by canceling terms and using total economy costs $w'v = \sum_{i=1}^N C_i(y_p, w)$, and the fourth equality follows by adding and subtracting the term $\sum_{i=1}^N C_i(y_i^a, w)$.

If condition (7.17) is positive, then this is sufficient to ensure that Pareto gains can be achieved. The final term in brackets is guaranteed to be nonnegative, because it is *feasible* to produce the autarky output y^a using the autarky factor supplies v^a , but not necessarily optimal with the wages w , so that economy-wide costs are $\sum_{i=1}^N C_i(y^a, w) \leq w'v^a$. Thus, by combining the first two terms, the condition to ensure Pareto gains from trade becomes

$$\sum_{i=1}^N p_i \omega_i (y_i - y_i^a) > 0, \quad (7.18)$$

where,⁹

$$\omega_i \equiv 1 - \left[\frac{C_i(y_p, w) - C_i(y_i^a, w)}{p_i(y_i - y_i^a)} \right], \text{ for } y_i \neq y_i^a. \quad (7.19)$$

The terms ω_i act as weights in the summation of (7.19). To interpret these weights, consider the case of constant returns to scale, so that $C_i(y_p, w) = p_i y_i = y_i AC_i(w)$ and $C_i(y_i^a, w) = y_i^a AC_i(w)$, where average costs AC_i are independent of output. Then (7.19) readily becomes $\omega_i = (p_i - AC_i)/p_i$, which is the markup of prices over average costs. Industries with perfect competition then have $\omega_i = 0$, and more generally these weights

⁹If $y_i = y_i^a$ then (7.19) is not defined, and we simply exclude that industry i from the summation in (7.18).

satisfy $0 \leq \omega_i \leq 1$. We argue in the next result that these bounds hold under increasing returns to scale, too.

THEOREM (GRINOLS 1991)

A sufficient condition to obtain Pareto gains from trade is that a weighted average of the outputs in industries subject to increasing returns to scale or imperfect competition expands, or (7.18) holds, where the weights ω_i defined in (7.19) satisfy $0 \leq \omega_i \leq 1$.

Proof:

From (7.17) and (7.18) the government budget is positive, so everyone can be made better off with a poll subsidy. Then the only thing left to show is that $0 \leq \omega_i \leq 1$. Since costs are non-decreasing in output the numerator and denominator of the bracketed term in (7.19) have the same sign, and it follows that the bracketed term is nonnegative, so that $\omega_i \leq 1$. The numerator of the bracketed term can be written as $C_i(y_p, w) - C_i(y_i^a, w) = AC_i(y_p, w)y_i - AC_i(y_i^a, w)y_i^a$. Consider the case where output is increasing, $y_i > y_i^a$. Since average costs are constant or falling in output, it follows that $AC_i(y_p, w) \leq AC_i(y_i^a, w)$, and so the numerator satisfies $C_i(y_p, w) - C_i(y_i^a, w) \leq AC_i(y_p, w)(y_i - y_i^a)$. Using this in (7.19), it follows that $\omega_i \geq [p_i - AC_i(y_p, w)]/p_i \geq 0$, where the final inequality is obtained since profits are nonnegative, so that $p_i \geq AC_i(y_p, w)$. A similar proof that $\omega_i \geq 0$ applies when $y_i < y_i^a$. QED

COROLLARY:

A sufficient condition for Pareto gains from trade is that the output of every industry subject to increasing returns to scale or imperfect competition does not fall when going from autarky to free trade, and at least one such industry expands.

Versions of this corollary have been obtained by Kemp and Negishi (1970), Markusen (1981), and Helpman (1984b). The theorem of Grinols is more general in its assumptions than is the work by any of these authors, and also shows how output increases in some industries and decreases in others can be summed together using the weights ω_i .

One concern with the above results is that it may be hard to achieve increases in industry output across several countries: if output increases in one country, is it not more likely to fall in another? This was the case analyzed by Markusen (1981) with a single industry subject to imperfect competition. For two countries that are identical except for size, he found that the smaller country would experience an increase in the imperfect competitive industry, but not the larger country. This means that gains from trade are assured only for the smaller country.

Sufficient conditions to have *all* countries gain from trade have been derived using a monopolistic competition model with homogeneous firms (Helpman and Krugman 1985, chapter 9). In this case, each firm is producing a differentiated product and there is free entry and exit from the industry. As a first step in their argument, Helpman and Krugman find that Pareto gains from trade for a single country are assured if

(a) the output of *each firm still producing* does not fall going from autarky to free trade; (b) total *industry output* (equal to the number of firms times the output of each) in that country does not fall going from autarky to free trade; (c) the set of goods available to consumers is *not reduced* due to the opening of trade. Notice that condition (a) allows some firms to exit due to the opening of trade, as discussed in chapter 5. Indeed, their departure gives room for other firms to increase their outputs and take advantage of economies of scale.

Extending the argument to deal with multiple countries opening to free trade, Helpman and Krugman (1985, 187–90) have identified quite weak assumptions to ensure that conditions (a)–(c) hold across all countries.¹⁰ Specifically, they assume that each industry in the economy is organized like the Krugman (1979) model considered at the beginning of chapter 5, with consumption in each industry coming from an additively separable utility function. Assume that (a) the elasticity of demand is non-decreasing in output; (b) there is a constant share of expenditure spent on each differentiated goods industry; (c) the production function for each differentiated good is homothetic; (d) equilibrium in each differentiated goods industry is symmetric, with identical costs, prices, and output. Then Helpman and Krugman demonstrate that the opening of trade will lead to at least as many goods, and at least as much output of each, provided only that the *world output of each differentiated goods industry does not contract*. This becomes a very weak condition to ensure Pareto gains from trade across all countries.

CONCLUSIONS

The idea that countries gain from trade is as old as the idea of comparative advantage itself—Ricardo wrote his model of trade between England and Portugal to demonstrate both claims. But gains “for a country” does not have a well-defined meaning unless we specify what this condition implies for the many different individuals located there. In this chapter we have taken the heterogeneity of individuals seriously, and have identified conditions under which *all agents* can gain from trade, that is, Pareto gains are achieved. The lump sum transfers that we have discussed require too much information to be implemented in practice, but are still valuable because they allow us to show that Pareto gains are *in principle* possible. The commodity tax/subsidies introduced by Dixit and Norman (1980) are an alternative means of achieving Pareto gains. In the presence of mobility costs, however, these may need to be supplemented with some form of trade adjustment assistance.

Our results for Pareto-gains within a country can be extended to cover the comparison of any two trading situations, and also to cover multiple countries, such as those belonging to customs unions and free-trade areas. In addition, we showed how imperfect competition and increasing returns to scale can be incorporated. Missing from our analysis, however, is an important topic related to imperfect competition and the gains from regional trade agreements. Suppose that trade in differentiated products occurs under monopolistic competition, subject to “iceberg” transportation costs as discussed in chapter 5. Furthermore, suppose that the countries forming a regional agreement do not aim to keep welfare in the rest of world unchanged; on the contrary,

¹⁰ Examples can be constructed, however, where one or more of these conditions is violated so that losses from trade are possible. See Lawrence and Spiller (1983) and François and van Ypersele (2002).

these countries adjust the tariffs applied to the rest of the world to their own optimal advantage. In the next chapter, we will call this the “optimal” tariff, and it is applied to shift a country’s terms of trade in its favor. Finally, suppose that the world splits up into a number of such regional areas, each of them equally sized. Then how does the formation of these areas impact world welfare?

This question has been addressed by Krugman (1991a,b), with two quite distinct answers. First, ignoring transportation costs, Krugman (1991a) argues that with the world divided into a number of equally sized regional trading areas, the *worst* number of such areas is *three*. The logic of this proposition is that with just one area, encompassing all countries, we have the highest level of world welfare (i.e., global free trade). Conversely, with many equal-sized areas, each group of countries will have little influence in changing their terms of trade, so tariffs are low and we are again close to free trade. But at some intermediate number of areas, tariffs will be higher and world welfare lower. Stylized calculations suggest that world welfare is *minimized* with three areas, which is a depressing result in view of the possible formation of three trading regions in reality: the Americas, Europe, and Asia.

But this view of the world ignores the fact that actual trading regions are often between geographically close countries (e.g., the EEC and NAFTA), and so the “natural” trading advantages of these countries are simply enhanced by the regional agreement. To see how this tendency affects the argument, suppose now that there are three continents (America, Europe, and Asia) with multiple countries in each. As in Krugman (1991b), take an extreme case where transportation costs are infinite for trade between the continents, but zero within each continent. In the absence of any regional agreements, the countries in each continent will be applying optimal tariffs against each other. But with a regional free trade agreement on each continent, tariffs are eliminated, and world welfare is maximized: the fact that outside countries are excluded from each regional agreement is irrelevant, since transportation costs between the continents are so high. So in this case, three regional trading areas *maximize* world welfare!

These two starkly different results suggest that the pattern and level of transportation costs between countries is crucial to determining the welfare effects of regional agreements. Obviously, we need to go beyond the two extreme cases of Krugman (1991a,b), and this is done by Frankel (1997). Continuing with Krugman’s example of a world divided into continents, each of which is divided into multiple countries, Frankel explores the welfare impact of regional agreements within or across continents. An agreement within a continent (where transportation costs are low) is called “natural,” while an agreement across continents (where transportation costs are high) is called “unnatural.” Frankel confirms that “natural” trading areas are more likely to lead to world welfare gains than “unnatural” areas.¹¹ However, it turns out that having regional agreements *within* continents is not always welfare improving, and the case where these “natural” areas fail to lead to world gains is called “supernatural.” This brief summary is meant to suggest that the impact of regional trade agreements—in theory or in practice—is still a wide open area for research. The papers summarized above use a monopolistic competition model with homogeneous firms, as we studied in chapter 5, but the case of heterogeneous firms discussed in chapter 6 has not yet been analyzed. We shall return to this topic in chapter 10, which deals with the political economy of tariffs.

¹¹In a test of this proposition, Krishna (2003) provides empirical evidence that raises doubts on the link between “natural” trading areas (between close countries) and welfare gains.

PROBLEMS

In earlier chapters, our only demonstration of the gains from trade has been a diagram like figure 7.3, where we show the indifference curves U being lifted off the production possibility frontier (PPF) due to free trade. The question is this: should we think of these indifference curves as applying to a single “representative” consumer, or is there some other interpretation that allows for consumers who differ in their utility functions? The following results are due to Samuelson (1956), and give a precise interpretation to the indifference curves shown.

Suppose that we have H consumers with utility functions $u^h(c^h)$, $h=1, \dots, H$, where we ignore individual supplies of factors for simplicity. We assume that these are increasing and quasi-concave. Further, suppose that we have a social welfare function $W(u^1, \dots, u^H)$, which is non-decreasing and quasi-concave. Samuelson defined the social utility function as

$$U(c) = \max_{c^h \geq 0} W[u^1(c^1), L, u^H(c^H)] \text{ st. } \sum_{h=1}^H c^h \leq c. \quad (7.20)$$

Thus, this social utility function takes the total amount of consumption goods available, c , and allocates it across individuals so as to maximize social welfare. This allocation is a form of “lump sum” transfers where we are physically allocating the goods (though the outcome would be similar if we instead used lump sum transfers of income).

Samuelson argued that we could think of the indifference curves in figure 7.3 as being iso-curves of the function U , so moving to a higher indifference curve due to trade (as illustrated) means that social welfare has indeed increased.

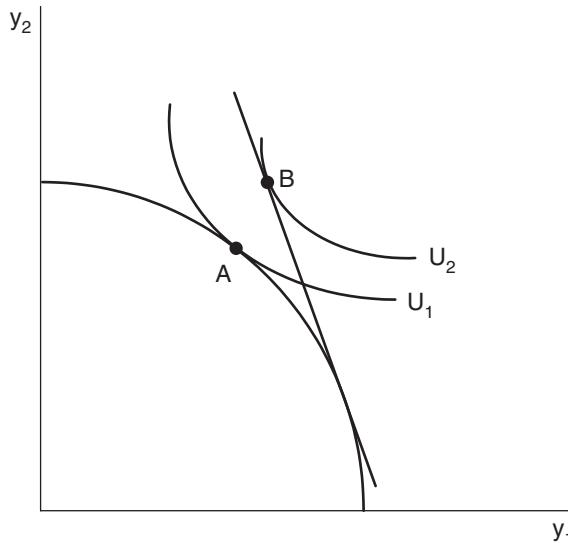


Figure 7.3

To further understand Samuelson's approach, answer the following:

- 7.1 The fact that we are moving from a welfare level of U_1 in figure 7.3 to U_2 due to the opening of trade proves that *some persons' utility must have increased*, but it does not prove that *everyone's* utility has increased. Can you think of a special form for the welfare function W that will ensure that when U goes up, *everyone's* utility must go up as well? (This social welfare function is used by Chipman (1987, 528), for example.)
- The next two questions extend our treatment of partial tariff reform:
- 7.2 In this problem we extend the analysis of partial tariff reform to use the commodity taxes and subsidies suggested by Dixit and Norman (1980), rather than lump sum transfers. Suppose that the government imposes taxes of $(p^0 - p^1)$ on goods and subsidies of $(w^0 - w^1)$ on factors. In addition, we will suppose that the poll subsidy of $R \equiv (t'^m^0)/H - (t'm^1)/H$ is given to each person (which is a tax if negative), in addition to the tariff revenue.
- Show that consumers can still afford their initial consumption bundles (c^{h0}, v^{h0}) .
 - Show that the government budget is nonnegative if $(p^{*0} - p^{*1})'m^0 + t'(m^1 - m^0) \geq 0$.
- 7.3 Using the assumption of trade balance at international prices, show that the condition $(p^{*0} - p^{*1})'m^0 + t'(m^1 - m^0) \geq 0$ can be alternatively expressed as $p'(m^1 - m^0) \geq 0$, and interpret the latter condition. This formulation of the condition to ensure Pareto gains under partial reform is used by Ju and Krishna (2000a).

Import Tariffs and Dumping

There are various reasons why countries use import tariffs and other types of trade policies. Nearly all countries have used these instruments in the early stages of their development to foster the growth of domestic industries, in what is called *import substitution*. Such policies have been heavily criticized for protecting inefficient domestic industries from international competition. Many countries have later switched to an *export promotion* regime, under which industries are expected to meet international competition through exports, albeit with subsidies (hopefully temporary) given to exporters. The more than 140 members of the WTO have all committed to abandon such heavily regulated trade regimes, and move toward substantially freer trade. One question, then, is whether the use of import tariffs and other trade policies at early stages of the development process has any rationale at all, especially when other markets (such as for capital) might not be functioning well. While this is too big a question to deal with adequately in this chapter, we will briefly discuss the rationale for temporary tariffs in what is called “infant industry” protection.

A second question concerns the welfare cost of tariffs and quotas in situations where other markets are working well. Even under the GATT/WTO, countries are permitted to apply tariffs in a number of cases, including: (i) “escape clause” tariffs, under which countries temporarily escape from their promise to keep tariffs low, due to injury in an import-competing industry; (ii) antidumping duties, under which tariffs are applied to offset import prices that are “too low.” For theoretical purposes, we can think of escape clause tariffs as *exogenously imposed* on exporting firms, and this will be our assumption in the first part of the chapter. We will examine the response of the exporters, as well as the response of import-competing firms, to such tariffs. In particular, we are interested in whether there is any “strategic” role for trade policy, whereby it gives an advantage to importing firms that leads to a welfare gain for the importing country. This possibility is associated with the work of Brander and Spencer (1984a,b). The response of firms to the tariff depends on the market structure, and we shall investigate a series of cases, concluding with an empirical investigation of these tariffs as applied to U.S. imports of compact trucks and heavyweight motorcycles (Feenstra 1989).

Following this, we turn to a discussion of dumping. We suggest that this phenomenon can be viewed as a natural attempt of imperfectly competitive firms to enter each other’s markets, as in the “reciprocal dumping” model of Brander (1981) and Brander and Krugman (1983). As such, it is likely to bring gains to consumers through lower prices. These gains will be offset by the use of antidumping duties, especially since

these duties must be treated as *endogenous*: their application will depend on the prices charged by the exporting firms. Exporting firms will have an incentive to raise their prices even if there is only a threat of antidumping duties being imposed, and to raise them even further if the duties are actually imposed. For these reasons, consumer and social losses due to antidumping actions are particularly high. We review empirical work by Prusa (1991, 1992), Staiger and Wolak (1994), Blonigen and Haynes (2002), and Ruhl (2014), who estimate the price and quantity effects of antidumping actions.

TARIFFS UNDER THE GATT/WTO

For member countries of the WTO the use of tariffs or quotas is allowed under some circumstances. For example, Article XIX of the GATT (see table 7.1), called the “escape clause,” allows for the temporary use of protection when domestic industries are experiencing unusual import competition. In order for the articles of GATT to have legal standing in its member countries, they must be reflected in a set of trade laws within each country. In the United States, Article XIX was incorporated into Title II, Section 201 of the Trade Act of 1974:

Title II, Relief From Injury Caused By Import Competition

Section 201. Investigation By International Trade Commission

(a) (1) A petition for eligibility for import relief for the purpose of facilitating orderly adjustment to import competition may be filed with the International Trade Commission . . . by an entity, including a trade association, firm, certified or recognized union, or a group of workers, which is representative of the industry....

(b) (1) [U]pon the filing of a petition under subsection (a) (1), the Commission shall promptly make an investigation to determine whether an article is being imported into the United States in such increased quantities as to be a substantial cause of serious injury, or threat thereof, to the domestic industry producing an article like or directly competitive with the imported article...

(4) For the purposes of this section, the term “substantial cause” means a cause *which is important and not less than any other cause*. [Italics added]

Source: U.S. Public Law 93-618, Jan. 3, 1975

It can be seen that in order to qualify for tariffs protection under Section 201, several criteria need to be met, the most important of which is that imports must be a “substantial cause of serious injury, or threat thereof, to the domestic industry,” where a “substantial cause” must be “not less than any other cause.” This legal language is very important in practice: it means that a tariff will be granted only if rising imports are the *most important* cause of injury to the domestic industry. This criterion is difficult enough to establish that escape clause protection is used infrequently, as is shown in table 8.1. Over the years 1980–2011, there were only thirty-one escape-clause cases filed in the United States, of which sixteen, or about one-half, obtained a *negative* recommendation from the U.S. International Trade Commission (ITC). The twelve affirmative cases (not counting ties and a terminated case) went to the president for a final ruling, and he gave a positive recommendation for import protection in only nine cases, or less than one every two years. We will investigate two of these U.S. cases in this chapter, one of which (heavyweight motorcycles) received import protection under Section 201, and the other of which (compact trucks) was denied under Section 201 but was able to get tariff protection by other means.

TABLE 8.1: Administered Protection Cases in the United States

Safeguard or Escape Clause Cases					
Total 1980–89	Total 1990–99	Total 2000–2011	Negative ITC Ruling	Affirmative ITC Ruling*	Total 1980–2011 Affirmative U.S. President Decision
19	9	3	16	12	9
China-Specific Safeguard Cases					
Total 1980–89	Total 1990–99	Total 2000–2011	Negative ITC Ruling	Affirmative ITC Ruling	Total 1980–2011 Affirmative U.S. President Decision
NA	NA	7	2	5	1
Antidumping Cases					
Total 1980–89	Total 1990–99	Total 2000–2011	Duty Levied	Cases Rejected	Cases Withdrawn
468	428	332	548	456	148

Sources: Hansen and Prusa (1995). Update for 1989 to 1994 from Destler (2005, 149, 165). Update for 1995 to 2011 from Chad P. Bown, “Global Antidumping Database” (2011), available at <http://econ.worldbank.org/ttbd/>.

* In addition to the twelve affirmative ITC safeguard cases, there were two ties and one terminated case.

NA: not applicable since the China specific safeguard began in 2001.

In the next panel of table 8.1 are shown special “China-specific” safeguard cases, under Section 421 of U.S. trade law, which arose when China joined the WTO in 2001 and expired in December 2013. The language used for this safeguard is different from that in Section 201:

SECTION 421

Upon the filing of a petition . . . the United States International Trade Commission . . . shall promptly make an investigation to determine whether products of the People’s Republic of China are being imported into the United States in such increased quantities or under such conditions as to cause or threaten to cause market disruption to the domestic producers of like or directly competitive products.

. . . (1) For purposes of this section, market disruption exists whenever imports of an article like or directly competitive with an article produced by a domestic industry are increasing rapidly, either absolutely or relatively, so as to be a significant cause of material injury, or threat of material injury, to the domestic industry.

(2) For purposes of paragraph (1), the term “significant cause” refers to a cause which contributes significantly to the material injury of the domestic industry, *but need not be equal to or greater than any other cause*. [Italics added]

To summarize, Section 421 actions needed to establish “material” injury and not “serious” injury to a domestic industry, and needed to be a “significant” but not a “substantial” cause, in contrast to Section 201. From 2001 to 2011, seven cases were filed under Section 421, of which two were denied by the ITC and five were approved. Of these five approved tariffs, the president also ruled in favor only once—the tariff imposed on imports of tires from China, approved by President Barack Obama in 2009 and in effect until 2012.

As an alternative to the escape clause, firms can file for import protection under the antidumping provision of the WTO, which is Article VI (see table 7.1). This provision has been incorporated into Title VII of U.S. trade laws, where Section 303 deals with “countervailing duties” and Section 731 deals with “antidumping duties.” Countervailing duties are applied to offset foreign subsidies that lower the price of the foreign export good in the home market. Antidumping deals with cases where it is believed that a foreign firm is selling goods in the home (i.e., importing) market at either a *lower price* than in its own market, or if there is no foreign price to observe, then at less than its *average costs* of production. The cases filed in the United States under the antidumping provisions are also shown in table 8.1, which shows that the number of antidumping cases vastly exceeds safeguard cases. From 1980 to 1989 and from 1990 to 1999, there were more than four hundred antidumping cases filed in the United States. In the next twelve years, 2000 to 2011, there were more than three hundred antidumping cases, bringing the total number of antidumping cases filed from 1980 to 2011 to more than twelve hundred!

To have antidumping duties applied, a case must first go to the U.S. Department of Commerce, which rules on whether imports are selling domestically at “less than fair value”; that is, below the price in their own market or below the average cost of making them. These rulings were positive in 93 percent of cases during this period. The case approved by the Department of Commerce is then brought before the ITC, which must rule on whether imports have caused “material injury” to the domestic industry (defined as “harm that is not inconsequential, immaterial, or unimportant”). This criterion is much easier to meet than the “substantial cause of serious injury” provision for a safeguard tariff, and, as a result, the ITC more frequently rules in favor of antidumping duties. Furthermore, the application of duties does not require the additional approval of the U. S. President. Of the twelve hundred antidumping cases filed from 1980 to 2011, about 450 were rejected and another 550 had duties levied.

The remaining 150 antidumping cases (or 148 to be precise, shown at the bottom of table 8.1) fall into a surprising third category: those that are *withdrawn* prior to a ruling by the ITC. It turns out that the U.S. antidumping law actually permits U.S. firms to withdraw their case and, acting through an intermediary at the Department of Commerce, agree with the foreign firm on the level of prices and market shares (Prusa 1991, 1992)! This type of communication between two American firms would be illegal under U.S. antitrust law, of course, but is exempted from prosecution in antidumping cases under the so-called Noerr-Pennington doctrine. As we would expect, these withdrawn and settled cases result in a significant increase in market prices, with losses for consumers, as will be discussed at the end of the chapter.

SOCIAL WELFARE

In the previous chapter we were careful to allow for heterogeneous consumers, and considered systems of lump sum transfers or commodity taxes and subsidies such

that everyone could be better off (Pareto gains from trade). It will now be convenient to adopt a simpler measure of overall social welfare, without worrying about whether every individual is better off or not. To construct this measure we need to have a method to “add up” the utilities of consumers. This will be achieved by assuming that every individual has a quasi-linear utility function, given by $c_0^h + U^h(c^h)$, where c_0^h is the consumption of a numeraire good, and c^h is the consumption vector of all other goods for consumer $h=1, \dots, H$, with U^h increasing and strictly concave. Each consumer maximizes utility subject to the budget constraint $c_0^h + p'c^h \leq I^h$.

Let $c^h = d^h(p)$ denote the optimal vector of consumption for each individual, with remaining income spent on the numeraire good, $c_0^h = I^h - p'd^h(p)$.¹ Then we can define social welfare as the sum of individual utilities,

$$W(p, I) \equiv \sum_{h=1}^H I^h - p'd^h(p) + U^h[d^h(p)], \quad (8.1)$$

where total income is $I = \sum_{h=1}^H I^h$. Note that the sum over individuals of the last two terms on the right-hand side of (8.1) gives consumer surplus. By the envelope theorem, the derivative of (8.1) is the negative of total consumption, $\partial W / \partial p = \sum_{h=1}^H -d^h(p) \equiv -d(p)$. Thus, we can use the social welfare function just like an indirect utility function for the economy as a whole.

Let us further simplify the analysis by supposing that there is only a *single import good* subject to a tariff. Holding the prices of all other goods fixed, we will therefore treat p as a *scalar*, denoting the price of that good in the importing country. Its world price is denoted by the scalar p^* , and the difference between these prices is the *specific import tariff* t , so $p = p^* + t$. We suppose that the numeraire good is also traded, at a fixed world price of unity. Labor is assumed to be the only factor of production, and each unit of the numeraire good requires one unit of labor. It follows that wages in the economy are also unity, so that total income equals the fixed labor supply of L . These assumptions allow the economy to mimic a partial-equilibrium setting, where wages are fixed and trade is balanced through flows of the numeraire good.

Output of the good in question is denoted by the scalar y , which may be produced by competitive or imperfectly competitive firms. We suppose that the industry costs of producing the good are denoted by $C(y)$, with marginal costs $C'(y)$.² Imports are denoted by the scalar $m = d(p) - y$, where $d'(p) < 0$ from the assumption that U^h is strictly concave for all h . We assume that revenue raised from the tariff, tm , is redistributed back to consumers, who are also each entitled to profits from the import-competing industry, which are $py - C(y)$. It follows that social welfare is written as

$$W[p, L + tm + py - C(y)] \equiv W(t). \quad (8.2)$$

This general expression for social welfare holds under perfect or imperfect competition. In the former case, it is common to refer to profits $[py - C(y)]$ as producer surplus, that is, the return to fixed factors of production in the industry. In the latter case, we denote profits of the domestic industry by $\pi = py - C(y)$, and will need to specify how these depend on actions of foreign firms. We will be considering three cases: (a) perfect competition in the home and foreign industries; (b) foreign monopoly,

¹Note that demand for the non-numeraire goods does not depend on income; see problem 8.1.

²Under perfect competition $C(y)$ denotes industry costs, and under monopoly it denotes the firm's costs. Under oligopoly with a homogeneous product, we should replace $C(y)$ with $NC(y)$, where N is the number of firms and $C(y)$ is the costs of each. Similarly, we replace sales py in (8.2) by Npy .

with no domestic import-competing firms; (c) duopoly, with one home firm and one foreign firm engaged in either Cournot or Bertrand competition. A fourth case of monopolistic competition is discussed by Helpman (1990) and is not covered here. We begin with perfect competition, where we distinguish a *small importing country*, meaning that the world price p^* is fixed even as the tariff varies, and a *large country*, whose tariff affects the foreign price p^* .

In each case, we are interested in how social welfare in (8.2) varies with the tariff. To determine this, let us first derive a general expression for the change in welfare. Treating the price p and output y as depending on the tariff, we can totally differentiate (8.2) to obtain

$$\begin{aligned}\frac{dW}{dt} &= -d(p)\frac{dp}{dt} + m + \left(t\frac{dm}{dp} + y\right)\frac{dp}{dt} + [p - C(y)]\frac{dy}{dt} \\ &= m\left(1 - \frac{dp}{dt}\right) + t\frac{dm}{dp}\frac{dp}{dt} + [p - C(y)]\frac{dy}{dt} \\ &= t\frac{dm}{dp}\frac{dp}{dt} - m\frac{dp^*}{dt} + [p - C(y)]\frac{dy}{dt}.\end{aligned}\quad (8.3)$$

The second line of (8.3) follows by noting that $d(p) - y = m$ and combining terms, while the third line follows because $p = p^* + t$, so that $[1 - (dp/dt)] = -dp^*/dt$.

We will be discussing each of the terms on the last line of (8.3) throughout this chapter, but note here the similarity between these terms and those derived in the previous chapter. The first term on the last line of (8.3) can be interpreted as the efficiency cost of the tariff, much like the term $t'(m^1 - m^0)$ in chapter 7; the second term is the effect of the tariff on the foreign price p^* , or the terms of trade effect, like $(p^{*0} - p^{*1})m^0$ in chapter 7; and the third term reflects the change in industry output times the price-cost margin, like (7.18). This third term reflects the fact that with imperfect competition there is a gap between the price or consumer valuation of a product, and the marginal costs to firms. This distortion due to monopolistic pricing creates an efficiency loss, and any increase in domestic output will therefore offset that loss and serve to raise welfare.

PERFECT COMPETITION, SMALL COUNTRY

Holding p^* fixed, we already argued in the previous chapter that “free trade is better than restricted trade for a small country.” In other words, the optimal tariff is zero. We now provide a more direct demonstration of this result, and also show how the welfare loss due to a tariff can be measured. Under perfect competition, profits are maximized when $p = C(y)$, so the final term in (8.3) is zero. When world prices are fixed and domestic prices given by $p = p^* + t$, then $dp^*/dt = 0$ and $dp/dt = 1$. Using these various relations in (8.3), we readily obtain

$$\frac{dW}{dt} = t\frac{dm}{dp}. \quad (8.4)$$

Evaluating this expression at a zero tariff, we have

$$\frac{dW}{dt} \Big|_{t=0} = 0. \quad (8.4')$$

This expression proves that social welfare has a critical point at a zero tariff, but we still need to determine whether this is a maximum or not. Differentiating (8.4) and evaluating at $t = 0$,

$$\frac{d^2W}{dt^2} \Big|_{t=0} = \frac{dm}{dp} < 0, \quad (8.5)$$

where this sign is obtained because $dm/dp = d'(p) - (1/C')$, with $d'(p) < 0$ from concavity of the utility functions and $C' > 0$ from the second-order condition for profit maximization. This proves that the critical point at $t = 0$ is a local *maximum*, and in fact, it is also a global maximum since with $dm/dp = 0$, then $t = 0$ is the *only* tariff at which (8.4) equals zero. Therefore, the *optimal tariff for a small country is zero*.

It is relatively easy to obtain an expression for the loss in welfare from applying a tariff. To do so, let us take a second-order Taylor series approximation of welfare, around the free trade point,

$$W(t) \approx W(0) + t \frac{dW}{dt} \Big|_{t=0} + \frac{1}{2}t^2 \frac{d^2W}{dt^2} \Big|_{t=0}. \quad (8.6)$$

Evaluating this using (8.4') and (8.5), we see that,

$$W(t) - W(0) \approx \frac{1}{2}t^2 \frac{d^2W}{dt^2} \Big|_{t=0} = \frac{1}{2}t^2 \frac{dm}{dp} = \frac{1}{2}\Delta p \Delta m < 0. \quad (8.7)$$

The final expression in (8.7) is negative since the tariff reduces imports, so the welfare loss equals one-half times the increase in price times the change in imports. This loss is illustrated in figure 8.1. In panel (a) we show the domestic demand curve D and supply curve S , together with the constant world price p^* . Under free trade, domestic demand is at c_0 and supply at y_0 , so imports are $m_0 = c_0 - y_0$. This is shown in panel (b), which graphs the import demand curve $M = D - S$. We can think of the fixed world price p^* as establishing a horizontal export supply curve X , which intersects M at the equilibrium imports m_0 .

With the import tariff of t , the export supply curve shifts up to $X + t$ in panel (b), leading to the equilibrium domestic price of $p = p^* + t$. Thus, the domestic price

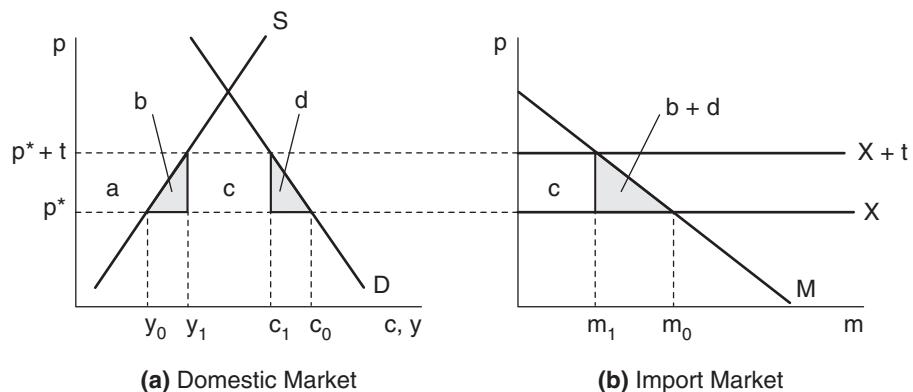


Figure 8.1: Small Country

increases by the full amount of the tariff. In panel (a), this leads to reduced demand of c_1 and increased supply of y_1 . The change in welfare in figure 8.1(a) can be decomposed as $-(a + b + c + d)$ consumer surplus loss + (a) producer surplus gain + (c) tariff revenue $= -(b + d)$, which is always negative. Area $(b + d)$ is the deadweight loss of the tariff, and is also shown as the triangle under the import demand curve in panel (b). The area (d) is interpreted as the consumer surplus loss for those units no longer purchased (i.e., $c_0 - c_1$), while the area (b) is interpreted as the increase in marginal costs (along the supply curve) for the extra units produced (i.e., $y_1 - y_0$).

So we see that the deadweight loss of the tariff can be measured by a triangle under the import demand curve. Since both the height and the base of the triangle depend on the tariff, the area of the triangle itself is of the second order of smalls for a small tariff, that is, the deadweight loss in (8.7) depends on the *square* of the tariff. Notice that we can rewrite (8.7) to measure the deadweight loss as a fraction of import expenditure:

$$\frac{W(t) - W(0)}{pm} \approx \frac{1}{2} \left(\frac{t}{p}\right)^2 \left(\frac{dm}{dp} \frac{p}{m}\right). \quad (8.7')$$

For example, with a 10 percent tariff and an import demand elasticity of 2, the deadweight loss relative to import expenditure will be 1 percent. As is typical for triangle calculations, we see that the deadweight loss is quite small for moderate tariffs. This has led many researchers to conclude that a formula like (8.7') does not capture the inefficiencies present in import substitution regimes. We will be investigating some reasons why the deadweight loss of trade restrictions may well be greater than (8.7'). This can occur if foreign firms anticipate the tariffs, as with antidumping duties (discussed later in this chapter), or if quotas are used instead of tariffs (discussed in the next chapter), or if tariffs have the effect of decreasing the *variety* of import products available. This latter effect has been emphasized by Feenstra (1992) and Romer (1994). Both authors provide simple formulas to compute welfare losses due to a reduction in import varieties, and these are much greater than (8.7'). In addition, Feenstra (1994), Broda and Weinstein (2006), and Feenstra and Weinstein (2010) provide estimates of the welfare *gain* due to new import varieties, which can be substantial.

PERFECT COMPETITION, LARGE COUNTRY

For a large country, we suppose that the world price of imports depends on the tariff chosen, which is written as $p^*(t)$. Generally, we define the *terms of trade* as the price of a country's exports divided by the price of its imports, so that a fall in the import price p^* is an *improvement* in the terms of trade. We will argue that a tariff leads to such an improvement in the terms of trade under perfect competition, as we assume in this section. In the next section we will examine the sign of this derivative under foreign monopoly.

The domestic price of the import good is $p = p^* + t$, so now $dp^*/dt \neq 0$ and $dp/dt \neq 1$. We still have $p = C'(y)$ under perfect competition, so (8.3) becomes

$$\frac{dW}{dt} = t \frac{dm}{dp} \frac{dp}{dt} - m \frac{dp^*}{dt}. \quad (8.8)$$

The first term is interpreted as the marginal deadweight loss from the tariff; and the second term is the terms of trade effect of the tariff, that is, the reduction in the price

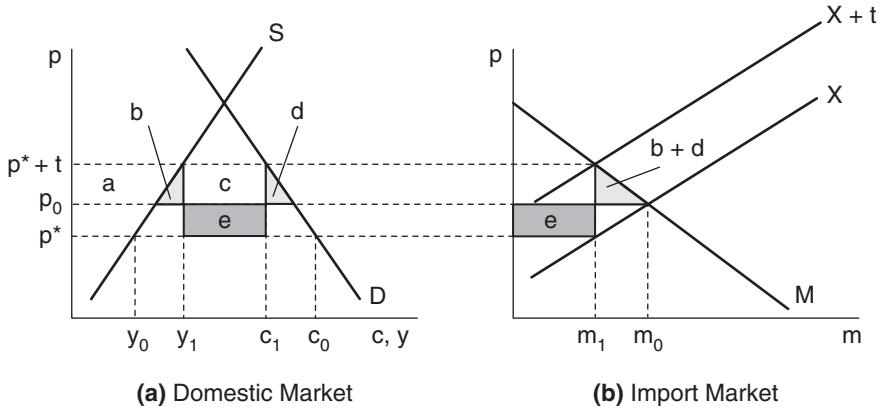


Figure 8.2: Large Country

of p^* times the amount of imports. The sign of dp^*/dt can be determined from figure 8.2. In panel (a), we show the domestic demand curve D and supply curve S , which lead to the import demand curve $M = D - S$ in panel (b). Also shown is the foreign supply curve X , which we assume is upward sloping. The initial equilibrium foreign and domestic price is $p_0^* = p_0$. The tariff shifts up the export supply curve to $X + t$, which intersects import demand at the new domestic price $p = p^* + t$. It is clear from the diagram that the increase in the domestic price p_0 to $p^* + t$ is less than the amount of the tariff t , which implies that the new foreign price p^* is less than its initial value p_0^* . This is a terms of trade gain for the importing country.

Using the result that $dp^*/dt < 0$, we obtain from (8.8),

$$\frac{dW}{dt} \Big|_{t=0} = -m \frac{dp^*}{dt} > 0, \quad (8.9)$$

so that a small tariff will necessarily raise welfare. The change in welfare in figure 8.2 is $-(a + b + c + d)$ consumer surplus loss + (a) producer surplus gain + $(c + e)$ tariff revenue = $e - (b + d)$. The area e equals the drop in the price p^* times the import quantity $m_1 = c_1 - y_1$, which is a precise measure of the terms of trade gain. The dead-weight loss triangle $(b + d)$ is still measured by (8.7), and depends on the square of the tariff, so it is of the second order of smalls. This does not apply to the terms of trade gain e , however, which depends on t rather than t^2 . For these reasons, the net welfare gain $e - (b + d)$ is positive for tariffs that are sufficiently small, while it is negative for large tariffs.

The *optimal tariff* t^* is computed where (8.8) equals zero:

$$\frac{dW}{dt} = 0 \Rightarrow \frac{t^*}{p^*} = \left(\frac{dp^*}{dt} \frac{m}{p^*} \right) \Bigg/ \left(\frac{dm}{dp} \frac{dp}{dt} \right). \quad (8.10)$$

To interpret this expression, we simplify it in two different ways:

Since domestic imports equal foreign exports, then $m = x \Rightarrow \frac{dm}{dp} \frac{dp}{dt} = \frac{dx}{dt}$.

Substituting this into (8.10) we obtain

$$\frac{t^*}{p^*} = \left(\frac{dp^*}{dt} \frac{x}{p^*} \right) / \left(\frac{dx}{dt} \right) = 1 / \left(\frac{dx}{dp^*} \frac{p^*}{x} \right). \quad (8.11)$$

where the second term is obtained because $(\frac{dx}{dt} / \frac{dp^*}{dt}) = \frac{dx}{dp^*}$, which is the slope of the foreign export supply curve. Thus, (8.11) states that the *optimal percentage tariff*, t^*/p^* , equals the *inverse of the elasticity of foreign export supply*.

This formula immediately confirms that for a small country, where the elasticity of foreign export supply is infinite, the optimal tariff is zero. For a large country where the elasticity of foreign export supply is finite and positive, so is the optimal tariff. To implement this formula we need information on the elasticity of foreign supply. It turns out that these supply elasticities are obtained as a by-product of the estimation method in Feenstra (1994), discussed in the appendix to chapter 5. That technique assumed a constant elasticity of substitution (CES) demand system across a product imported from multiple source countries, which have the same demand elasticity (i.e., the elasticity of substitution). Likewise, Feenstra (1994) assumes a constant elasticity supply curve for each product, with the same supply elasticity over the source countries. Broda and Weinstein (2006) apply the Feenstra (1994) method to estimate the demand and supply elasticities for each four-digit Harmonized System (HS) product imported into the United States.

Broda, Limao, and Weinstein (2008) extend this estimation to countries beyond the United States, and specifically focus on the supply elasticities to examine the optimal tariffs.³ They find that the inverse supply elasticity, that is, the formula in (8.11), is between one and three for the typical four-digit HS good. The *actual* tariffs applied by GATT/WTO members will differ greatly from the optimal tariffs, because members of the WTO are constrained to charge “most favored nation” (MFN) tariffs to other WTO members. So Broda, Limao, and Weinstein focus on fifteen countries prior to membership in the WTO. The authors find that these countries set higher tariffs on products in which they had more market power, that is, in which they faced higher inverse supply elasticities, and this effect was significant in thirteen out of the fifteen countries. Their research also confirms a positive link between the inverse supply elasticities and the tariffs set by the United States against non-WTO members, and also the non-tariff barriers used by the United States. So the economic intuition of the simple formula in (8.11) is confirmed by this study.

There is a second interpretation of the optimal tariff formula in (8.10) that depends much less on supply elasticities, but instead on the pricing behavior of imperfectly competitive firms, and which we can also investigate empirically.

An alternative way of writing (8.10) is simply to rearrange terms, obtaining

$$\frac{t^*}{p^*} = \left(\frac{dm}{dp} \frac{p}{m} \right)^{-1} \left(\frac{dp^*}{dt} / \frac{dp}{dt} \right). \quad (8.12)$$

Now the optimal tariff equals the *inverse of the elasticity of import demand*, times the ratio of the change in the relative *foreign* and *domestic* price of imports. Since the

³ Soderber (2014) extends the estimation method to allow for different elasticities of supply for various source countries. He then tests whether the supply elasticities play a role in the “protection for sale” model of tariffs (Grossman and Helpman 1994), discussed in chapter 10.

import demand elasticity is negative, and presuming that $dp/dt > 0$ (which we will confirm below), then the optimal tariff is positive provided that $dp^*/dt < 0$.

Think about the foreign price p^* as being chosen strategically by exporting firms. When these firms are faced with a tariff, how will they adjust the *net of tariff* price p^* that they receive?

Will they absorb part of the tariff, meaning that $dp^*/dt < 0$ and $dp/dt < 1$, in an attempt to moderate the increase in the import price $p = p^* + t$? This may well be a profit-maximizing strategy. Generally, we will refer to the magnitude of the derivative dp/dt as the “pass-through” of the tariff: if domestic prices rise by less than the tariff, so that $dp/dt < 1$, this means that foreign exporters have absorbed part of the tariff, $dp^*/dt < 0$, which is a terms of trade gain. The smaller the pass-through of the tariff, the larger the optimal tariff from (8.12). The nice feature of this formula is that it puts the emphasis on the pricing decisions made by foreign exporters, which is entirely appropriate. We analyze this decision of foreign firms next.

FOREIGN MONOPOLY

We turn now to the case of a single foreign exporter, selling into the home market. The idea that this firm does not have any competitors in the home market is not very realistic: tariffs are designed to protect domestic firms, and if there are none, we would not expect a tariff! So the assumption of a foreign monopolist should be thought of as a simplification. Our analysis of this case follows Brander and Spencer (1984a,b).

Let x denote the sales of the exporting firm into the home market, which equals home consumption, so $x = d(p)$. We will invert this expression and work with the inverse demand curve, $p = p(x)$, with $p' < 0$. The price received by the foreign exporter is $p^* = p(x) - t$, so that foreign profits are

$$\pi^*(x) = x[p(x) - t] - C^*(x), \quad (8.13)$$

where $C^*(x)$ is foreign costs. Maximizing this over the choice of x , the first-order condition is

$$\pi^{''}(x) = p(x) + xp'(x) - [C^{''}(x) + t] = 0. \quad (8.14)$$

The term $p(x) + xp'(x)$ is just marginal revenue, while $C^{''}(x) + t$ is marginal costs inclusive of the tariff, and these are equalized to maximize profits.

Totally differentiating (8.14) we obtain $\pi^{'''}(x)dx - dt = 0$, so that

$$\frac{dx}{dt} = \frac{1}{\pi^{'''}(x)} < 0, \quad (8.15)$$

where the sign is obtained from the second-order condition for profit maximization. It follows that the change in the import price is

$$\frac{dp}{dt} = p'(x)\frac{dx}{dt} = \frac{p'(x)}{\pi^{'''}(x)} > 0. \quad (8.16)$$

Not surprisingly, the tariff inclusive price $p = p^* + t$ rises. We are interested in whether it rises *less than* the amount of the tariff, that is, whether the pass-through of the tariff

is less than complete. Notice that with $p = p^* + t$, then $dp/dt < 1$ if and only if $dp^*/dt < 0$. Thus, “partial pass-through” of the tariff to domestic prices is equivalent to having the foreign firm absorb part of the tariff, which is a terms of trade gain.

Noting that the numerator and denominator of (8.16) are both negative, then $dp/dt < 1$ if and only if

$$p'(x) > \pi''(x) = 2p'(x) + xp''(x) - C^{**}(x). \quad (8.17)$$

The left side of (8.17) is the slope of the inverse demand curve, while the right side is interpreted as the slope of the marginal revenue curve, $[2p'(x) + xp''(x)]$, less the slope of marginal costs. Suppose that marginal costs are constant, $C^{**} = 0$. Then (8.17) will hold if and only if

$$p'(x) + xp''(x) < 0, \quad (8.17')$$

which guarantees that marginal revenue is steeper than demand. This condition will hold for linear or concave demand curves, or for any demand curve that is not “too convex.”

To illustrate these results, in figure 8.3 we show the initial equilibrium with the domestic and foreign price of p_0 . With the increase in the marginal costs due to the tariff t , the import price rises to p_1 . Provided that the marginal revenue is *steeper* than demand, we see this increase in price is less than t , so that the *foreign* price received by the exporters falls, $p_1 - t < p_0$. In that case, $dp^*/dt < 0$ and then the optimal tariff from (8.8) is positive. The change in welfare is shown in figure 8.3 by $-(c+d)$ consumer surplus loss $+(c+e)$ revenue gain $= (e-d)$, which is positive if the tariff is sufficiently small.

If marginal revenue is *flatter* than demand, however, then $dp^*/dt > 0$ and the optimal policy is an *import subsidy*. We regard this as a relatively unusual case, but it is certainly possible. For example, if the elasticity of demand is constant, then since marginal revenue is $mr(x) = p(x)[1 - (1/\eta)]$, we have $-mr'(x) = -p'(x)[1 - (1/\eta)] < -p'(x)$, and it follows that marginal revenue is flatter than demand.

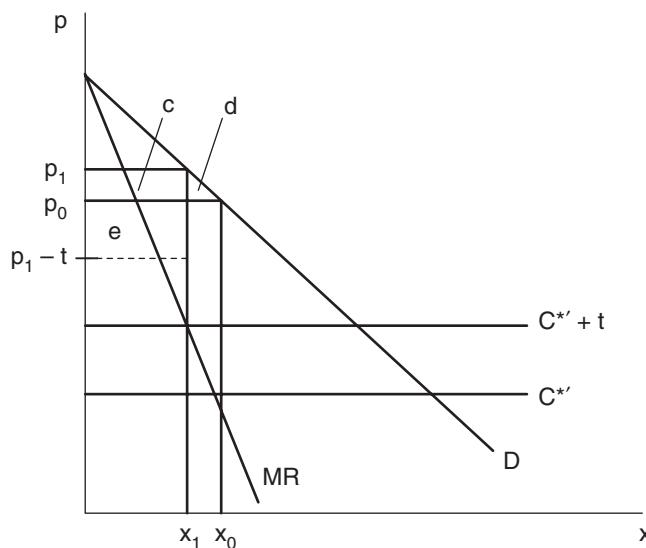


Figure 8.3

We next check whether an *ad valorem* tariff τ leads to similar results. The net of tariff price received by the foreign exporters is now $p^* = p(x)/(1+\tau)$, so that profits are

$$\pi^*(x) = \frac{xp(x)}{(1+\tau)} - C^*(x). \quad (8.18)$$

Maximizing this over the choice of x , the first-order condition is $p + xp' = (1+\tau)C^{*'}(x)$, which can be written as

$$p(x)\left(1 - \frac{1}{\eta}\right) = (1+\tau)C^{*'}(x), \quad (8.19)$$

using the elasticity $\eta(x) = (dx/dp)(p/x) = p/xp'(x)$.

It is straightforward to show that $dx/d\tau < 0$ and $dp/d\tau > 0$, as with the specific tariff. To determine the change in the foreign price p^* , rewrite (8.19) as

$$p^* = \frac{p(x)}{(1+\tau)} = \left(\frac{\eta}{\eta-1}\right)C^{*'}(x). \quad (8.19')$$

Consider the case where the exporter's marginal costs $C^{*'}(x)$ are constant. Then differentiating (8.19'), we obtain

$$\frac{dp^*}{d\tau} = \left[\frac{\eta'}{\eta-1} - \frac{\eta\eta'}{(\eta-1)^2} \right] \frac{dx}{d\tau} C^{*'} = -\frac{\eta'(x)}{(\eta-1)^2} \frac{dx}{d\tau} C^{*'}.$$

Using the fact that $dx/d\tau < 0$, we see that $dp^*/d\tau < 0$ when $\eta'(x) < 0$, meaning that the elasticity of demand *increases* as consumption of the importable falls. In that case the tariff leads to a fall in the price-cost margin set by the foreign exporter, so that p^* falls, which is a terms of trade gains for the importer. Thus, we summarize our results with the following theorem.

THEOREM (BRANDER AND SPENCER 1984A,B)

When the home country imports from a foreign monopolist with constant marginal costs,

- (a) a small specific tariff improves the terms of trade and raises home welfare if marginal revenue is steeper than demand;
- (b) a small *ad valorem* tariff improves the terms of trade and raises home welfare if the elasticity of demand increases as consumption of the importable falls.

Notice that the condition the $\eta'(x) < 0$ in part (b) is identical to the assumption on the elasticity of demand used by Krugman (1979), as discussed at the beginning of chapter 5. This case applies for linear or concave demand curves, and more generally, any curve that is “less convex” than a constant-elasticity demand curve.

COURNOT DUOPOLY

Let us now introduce a domestic firm that is competing with the foreign firm in the domestic market. We will have to specify the strategic variables chosen by the firms

and will assume Cournot competition in quantities in this section, and Bertrand competition in prices in the next. To avoid covering too many cases, we will focus on a *specific* tariff with Cournot competition, and the *ad valorem* tariff with Bertrand competition. To preview our results in this section, we confirm the finding of Helpman and Krugman (1989, sec. 6.1) that a positive tariff is very likely to be the optimal policy under Cournot duopoly.

We let x denote the sales of the foreign exporter in the domestic market, and y denote the sales of the home firm, so total consumption is $z = x + y$. As before, we invert the demand curve $z = d(p)$ to obtain inverse demand $p = p(z)$, with $p' < 0$. Profits of the foreign and home firms are then

$$\pi^* = x[p(z) - t] - C(x), \quad (8.21a)$$

$$\pi = yp(z) - C(y), \quad (8.21b)$$

Maximizing these over the choice of x and y , respectively, the first-order conditions are

$$\pi_x^* = p(z) + xp'(z) - [C'(x) + t] = 0. \quad (8.22a)$$

$$\pi_y = p(z) + yp'(z) - C'(y) = 0. \quad (8.22b)$$

The second-order conditions are $\pi_{xx}^* = 2p' + xp'' - C'' < 0$ and $\pi_{yy} = 2p' + yp'' - C'' < 0$. In addition, we assume that the stability condition $\pi_{xx}^* \pi_{yy} - \pi_{xy}^* \pi_{yx} > 0$ holds.

Using (8.22a), we can solve for foreign exports x as a function of domestic sales y , written as the reaction curve $x = r^*(y, t)$. From (8.22b), we can also solve for domestic output y as a function of foreign sales x , written as a reaction curve $y = r(x)$. These reaction curves are graphed as R^*R^* and RR in figure 8.4, and their intersection determines the Cournot equilibrium denoted by point C . The typical property of these reaction curves is that they are both downward sloping. For stability, the home reaction

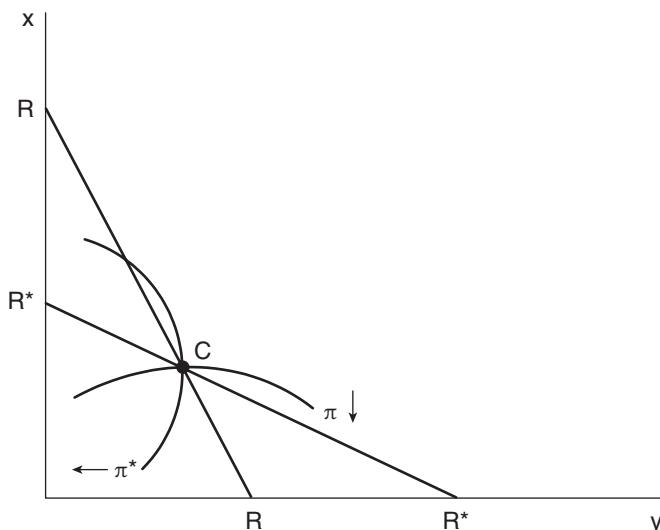


Figure 8.4

curve RR needs to cut the foreign reaction curve R^*R^* from above, as illustrated.⁴ Notice that the iso-profit curves of π have higher profits in the downward direction (i.e., for reduced x), and similarly the iso-profit curve π^* have higher foreign profits in the leftward direction (for reduced y), as illustrated.

With the tariff, the foreign firm will reduce the amount it wishes to export (from (8.22b), $dx/dt = 1/\pi_{xx}^* < 0$), and its reaction curve shifts down to $R'R'$, as shown in figure 8.5. The equilibrium therefore shifts from point C to point D , involving reduced sales of export sales x but increased domestic sales y , together with increased home profits. To determine the effect on the domestic prices we need to calculate the impact on *total sales* $z = x + y$, and then on $p(z)$. To determine this, it is convenient to sum the first-order conditions, obtaining

$$2p(z) + zp'(z) = C'(y) + [C'(x) + t]. \quad (8.23)$$

Suppose that domestic and foreign marginal costs are *both constant*. Then we can totally differentiate (8.23) to obtain

$$\frac{dz}{dt} = \frac{1}{[3p'(z) + zp''(z)]}, \quad (8.24a)$$

and,

$$\frac{dp}{dt} = \frac{p'(z)}{[3p'(z) + zp''(z)]}. \quad (8.24b)$$

From (8.24a) we see that total consumption falls due to the tariff if $3p'(z) + zp''(z) < 0$. In that case the numerator and denominator of (8.24b) are both negative, so that $dp/dt < 1$ if

$$p'(z) > 3p'(z) + zp''(z) \Leftrightarrow 2p'(z) + zp''(z) < 0. \quad (8.25)$$

The condition on the right-hand side of (8.25) states that marginal revenue for the market as a whole, which is $p'(z) + zp'(z)$, *slopes down*. This is a significantly weaker condition than (8.17'), which involved a comparison of the slopes of marginal revenue and demand; now, we only need to check the slope of marginal revenue, and most demand curves will satisfy (8.25). This includes the constant elasticity curve, for which $mr'(z) = -p'(z)[1 - (1/\eta)] < 0$. When (8.25) is satisfied, then $dz/dt < 0$ in (8.24a) and $dp/dt < 1$ in (8.24b), so there is a beneficial terms of trade impact ($dp^*/dt < 0$) for the importing country.

Checking social welfare, we again use the general change in social welfare in (8.3), rewritten here using foreign exports (x) rather than home imports (m):

$$\frac{dW}{dt} = t \frac{dx}{dt} - x \frac{dp^*}{dt} + [p - C'(y)] \frac{dy}{dt}, \quad (8.3')$$

where the final term reflects the change in domestic output from equilibrium point C to point D . The first term on the right-hand side of (8.3') vanishes for small tariffs,

⁴In problem 8.3, you are asked to derive these properties of the reaction curves.

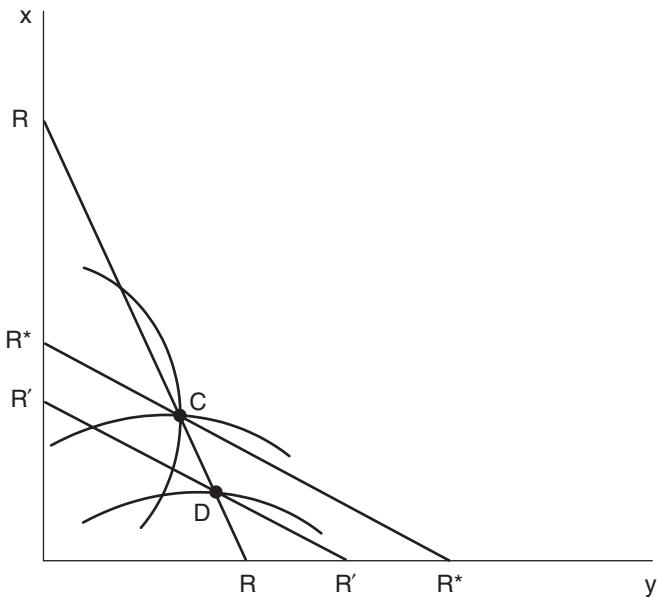


Figure 8.5

and the second term is positive when (8.25) holds, so that a positive tariff improves the terms of trade. For the third term, in the typical case where both reaction curves are downward sloping, with the home curve cutting the foreign curve from above as shown in figure 8.5, we will have $dy/dt > 0$: the tariff will lead to an increase in domestic output, which brings an additional terms of trade gain. In this case, the optimal tariff is unambiguously positive.

However, the typical pattern of the reaction curves shown in figure 8.5 need not hold, and it is possible that the home curve slope upward instead, in which case $dy/dt < 0$, so the tariff *reduces* home output. This will tend to offset the terms of trade gains due to the tariff. With constant marginal costs, it can be shown that this case is avoided, so that $dy/dt > 0$, provided that $p'(z) + yp''(z) < 0$.⁵ This condition states that “perceived” marginal revenue for the foreign firm, $p(z) + yp'(z)$, is steeper than the demand curve $p(z)$, and is somewhat weaker than (8.17'). Combining this condition with our finding that $dp/dt < 1$ when (8.25) holds, so that $dp^*/dt < 0$, we have two distinct reasons why a tariff will raise social welfare: the terms of trade effect; and the efficiency effect, whereby the tariff moves the home firm to a higher level of output at point D and offsets the distortion between price and marginal cost.

The idea that imperfect competition might justify the “strategic” use of trade policy was an area of very active research during the 1980s. In the tariff case we are discussing, it is tempting to refer to the last term in (8.3') as a “profit shifting” motive for the use of tariffs, suggesting that a “strategic” use of tariffs is to shift profits toward the home firm. But this terminology must be interpreted with caution. There is no direct relation between the sign of the last term in (8.3') and the change in profits for the domestic firm. In the case we are considering, profits of the home firm *always* rise from

⁵You are asked to prove this in problem 8.4.

point C to point D in figure 8.5 due to the tariff, but nevertheless, the last term in (8.3') can be positive or negative.⁶

In addition, Horstmann and Markusen (1986) argue the “profit shifting” motive for the tariff disappears if there is *free entry* into the domestic industry with Cournot competition. That is, suppose we allow free entry of home firms. Then the tariff, which would cause an incipient rise in profits, leads to the entry of domestic firms until profits are returned to zero. Domestic welfare is written in this case as $W[p, L + tm + N(py - C(y))]$, where N is the number of domestic firms, but with profits equal to zero this becomes $W(p, L + tm)$. Differentiating this with respect to the tariff t , we obtain the total change in utility:

$$\frac{dW}{dt} = -d(p)\frac{dp}{dt} + m + t\frac{dm}{dp}\frac{dp}{dt}. \quad (8.26)$$

This very simple expression is the alternative to (8.3) when profits are identically zero. For small tariffs, the final term vanishes, so the change in welfare depends on a comparison of the first two terms. It is immediately clear that welfare of the importing country increases for a small tariff if and only if $dp/dt < m/d(p)$. The interpretation of this condition is that the pass-through of the tariff to domestic prices must be *less than* the import share $m/d(p)$. Thus, having a terms of trade gain due to the tariff (meaning that $dp/dt < 1$) is no longer sufficient to obtain a welfare improvement; instead, we must have $dp/dt < m/d(p)$. We will be reviewing estimates of the pass-through later in the chapter, and it is not unusual to find that 50 to 75 percent of a tariff is reflected in the import price. Under Cournot competition and free entry, this tariff would be welfare improving (due to tariff revenue raised) only in an industry where the import share *exceeds* this magnitude.

BERTRAND DUOPOLY

We turn next to the case of Bertrand competition between the home and foreign firm, where they are choosing prices as strategic variables. If the domestic and imported goods are perfect substitutes, as we assumed in the previous section, then Bertrand competition leads to marginal cost pricing. To avoid this case, we will suppose that the domestic and import goods are *imperfect substitutes*, with the price of the import good denoted by p and that of the domestic good by q . Let us momentarily depart from our assumption at the beginning of this chapter that there is an additively separable numeraire good that absorbs all income effects, and instead model demand for the domestic good as $y = d(p, q, I)$, and demand for the imported good as $x = d^*(p, q, I)$, where I is the expenditure on both of these goods. We will treat this expenditure as constant, though this is a strong assumption and will be weakened below. Both these demand functions should be homogeneous of degree zero in prices and expenditure I .

With an ad valorem tariff of τ on imports, their domestic price is $p = p^*(1 + \tau)$. So profits of the foreign and home firms are

$$\pi^* = \frac{pd^*(p, q, I)}{(1 + \tau)} - C^*[d^*(p, q, I)]. \quad (8.27a)$$

⁶See problem 8.4.

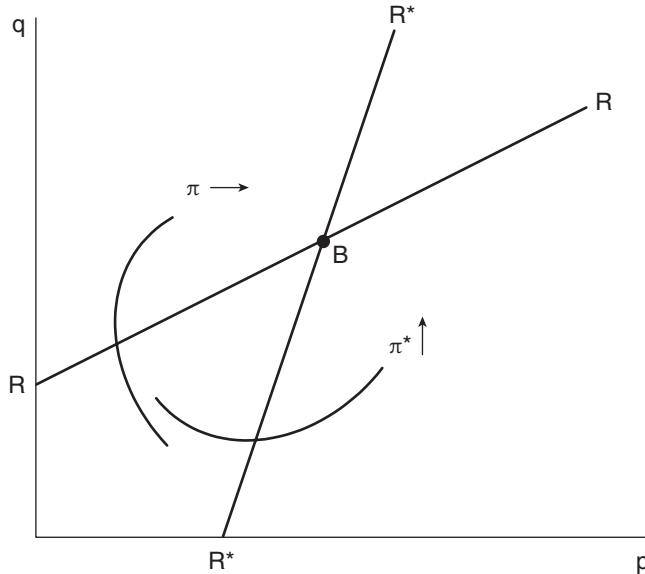


Figure 8.6

$$\pi = qd(p, q, I) - C[d(p, q, I)], \quad (8.27b)$$

Maximizing these over the choice of p and q , respectively, treating total expenditure I as fixed, the first-order conditions $\pi_p^* = \pi_q = 0$ can be simplified as

$$p\left(1 - \frac{1}{\eta^*}\right) = (1 + \tau)C'[d^*(p, q, I)], \quad (8.28a)$$

$$q\left(1 - \frac{1}{\eta}\right) = C'[d(p, q, I)], \quad (8.28b)$$

where $\eta^* \equiv -(\partial d^*(p, q, I)/\partial p)(p/d^*)$ and $\eta \equiv -(\partial d(p, q, I)/\partial q)(q/d)$ are the (positive) elasticities of import and domestic demand, respectively. The second-order conditions are $\pi_{pp}^* < 0$ and $\pi_{qq} < 0$, and for stability we require that $\pi_{qq}\pi_{pp}^* - \pi_{qp}\pi_{pq}^* > 0$.

Given the domestic price q , income I , and the tariff, we can use (8.28a) to solve for the tariff-inclusive import price p , obtaining the reaction curve $p = r^*(q, \tau)$. In addition, given the tariff-inclusive import price p and income I , we can use (8.28b) to solve for the domestic price q , obtaining the reaction curve $q = r(p)$. The intersection of these determines the Bertrand equilibrium, at point B in figure 8.6. The iso-profit curves of π have higher profits in the rightward direction (i.e., for higher p), and similarly the iso-profit curve π^* have higher foreign profits in the upward direction (for higher q), as illustrated. To derive the properties of the reaction curves, it is helpful to simplify the elasticities of demand. If the demand functions d and d^* have income elasticities of unity, it turns out that a change in income I does not affect the elasticities, nor the reaction curves. It follows that the elasticities can be written as functions of the *price ratio* of p and q , or $\eta^*(p/q)$ and $\eta(q/p)$.⁷

⁷See problem 8.5 to prove these properties of the elasticities.

The assumption that we used earlier to ensure that the ad valorem tariff led to a fall in the import price p^* , or a terms of trade gain, was that the elasticity of demand was decreasing in *quantity*. Since we are now thinking of the elasticities as a function of relative prices, the analogous assumption would be $\eta''(p/q) > 0$ and $\eta'(q/p) > 0$; that is, elasticities increase as the *relative price* of that good rises. Treating foreign and domestic marginal costs as constant in (8.28), this assumption will ensure that both reaction curves $p = r^*(q, \tau)$ and $q = r(p)$ slope upward, as shown in figure 8.6. As the relative price of the competing good rises, the elasticity falls, and each firm will charge a higher price for its own good. Furthermore, there is a *dampened* response of each price to that of the competing good, so that $\frac{dp}{dq} \frac{q}{p} = r_q^*(q, \tau) \frac{q}{p} < 1$, and $\frac{dq}{dp} \frac{p}{q} = r_p(p) \frac{p}{q} < 1$.⁸

Suppose that the *ad valorem* tariff τ increases. From the second-order conditions for profit maximization, we will have an increase in the tariff-inclusive import price p .

In figure 8.7, the foreign reaction curve shifts rightward to $R'R'$. This leads to an induced increase in the domestic price q , and a further increase in the import price p , until the new equilibrium is reached at point D . We are interested in whether the percentage increase in p is *less than* the amount of the tariff, that is, $(dp/d\tau)(1+\tau)/p < 1$, which will ensure that $dp^*/d\tau < 0$ so that there is a terms of trade gain.⁹ It turns out that the assumptions we have already made, that $\eta''(p/q) > 0$ and $\eta'(q/p) > 0$, are enough to guarantee this outcome. These conditions are analogous to what we used above to ensure that the ad valorem tariff on a foreign monopolist led to a fall in the import price p^* , or a terms of trade gain. Thus, whether there is a single foreign firm, or a duopoly in the import market, the key condition to ensure a terms of trade gain is that the elasticity of demand is increasing in price or decreasing in quantity. This condition holds for any demand curve that is “less convex” than a constant-elasticity demand curve.

To demonstrate the effect of the tariff on p^* , write the import price *net* of the tariff as $p^* = p/(1+\tau)$. From (8.28a) this equals

$$p^* \left(1 - \frac{1}{\eta^*} \right) = C^*. \quad (8.29)$$

Totally differentiate this expression, holding marginal cost constant, to obtain

$$\frac{dp^*}{d\tau} \left(1 - \frac{1}{\eta^*} \right) + p^* \left(\frac{\eta^{**}}{\eta^{*2}} \right) \frac{d(p/q)}{d\tau} = 0 \Rightarrow \frac{dp^*}{d\tau} = - \left[\frac{p^* \eta^{**}}{\eta^* (\eta^* - 1)} \right] \frac{d(p/q)}{d\tau}. \quad (8.30)$$

Since we have assumed that $\eta^{**} > 0$, we see that $dp^*/d\tau < 0$ provided that $d(p/q)/d\tau > 0$, that is, provided that the tariff leads to a higher increase in the price of the importable than the domestic good. But this condition is easily confirmed from figure 8.7. With a rightward shift of the foreign reaction curve, and a dampened response of the domestic price q , the new equilibrium given by point D will lie *below* a ray from the origin going through the initial equilibrium given by point B . In other words, the relative price of the importable (p/q) must rise, and it follows from (8.30) that the net-of-tariff foreign price falls, which is a terms of trade gain for the importer.

To calculate the change in social welfare under Bertrand duopoly, we rewrite the social welfare function as $W[p, q, L + \tau p^* x + qy - C(y)]$, including both the import

⁸ In problem 8.6 you are asked to demonstrate these properties of the reaction curves.

⁹ Notice that with $p = p^*(1+\tau)$, then $dp/d\tau = p^* + (1+\tau) dp^*/d\tau$, so $(dp/d\tau)(1+\tau)/p = 1 + (dp^*/d\tau)(1+\tau)/p^*$. Therefore, if $(dp/d\tau)(1+\tau)/p < 1$, then $dp^*/d\tau < 0$.

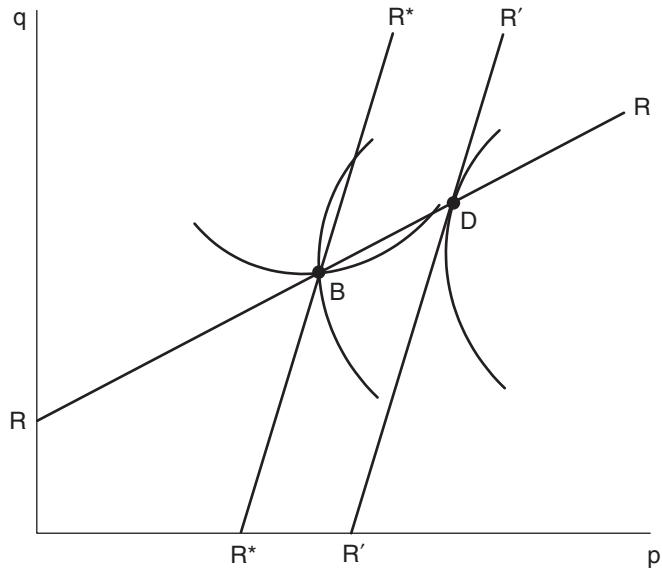


Figure 8.7

price $p = p^*(1 + \tau)$ and domestic price q . Totally differentiating this, and using steps similar to those used in (8.3), we find that the welfare change due to the change in the *ad valorem* tariff is

$$\frac{dW}{dt} = \tau p^* \frac{dx}{d\tau} - x \frac{dp^*}{d\tau} + [q - C(y)] \frac{dy}{d\tau}. \quad (8.31)$$

The first term on the right-hand side of (8.31) vanishes for small tariffs, and the second term is positive when $\eta'(p/q) > 0$ and $\eta'(q/p) > 0$, so there is a terms of trade gain. The third term in (8.31) depends on the change in equilibrium output from point B to point D , and is of ambiguous sign in general: the increase in prices from point B to point D reduces demand, while the fall in (p/q) shifts demand toward the domestic good, so the effect on domestic output depends on the relative strength of these effects. It turns out that if the increase in domestic price is not too large, then an increase in domestic output occurs. The condition to ensure this is that the elasticity of the home reaction curve does not exceed $\frac{dq}{dp} = r_p(p) \frac{p}{q} < (\frac{\eta-1}{\eta})$.¹⁰ In that case, domestic output increases due to the tariff, which generates a welfare gain additional to the terms of trade gain.

The finding that domestic output may rise due to the tariff depends on our maintained assumption that total expenditure I on the imported and domestic products is *constant*. This would apply, for example, with a Cobb-Douglas utility function between an aggregate of the products (d, d^*) and all other goods in the economy, but would not be true otherwise. Generally, total expenditure on the imported and domestic products can *rise or fall* due to the tariff, and the latter case could lead to a fall in domestic

¹⁰ See problem 8.7.

output and a welfare loss.¹¹ So the “profit shifting” motive for import protection is not that robust, as we also found under Cournot competition. Instead, the terms of trade motive for tariffs becomes the best indicator of gain or loss for the importing country. Under our assumptions that the elasticity of demand is *increasing* in price, there is a terms of trade gain for the importer, and the optimal tariff is positive for that reason. Conversely, with constant elasticities of demand, the optimal tariff is zero. Finally, in the somewhat unusual case where the elasticity of demand is *decreasing* in price, an import subsidy is optimal.

TARIFFS ON JAPANESE TRUCK AND MOTORCYCLES

We have found that with imperfect competition, a terms of trade argument for a tariff is likely but not guaranteed. The “strategic” response of foreign firms in adjusting their price p^* then becomes an interesting empirical question. Feenstra (1989) provides some empirical evidence on this topic from two U.S. tariffs initially applied during the early 1980s: a 25 percent tariff on imports of compact trucks, coming from Japan; and a temporary, declining tariff beginning at 45 percent on imports of heavyweight motorcycles from Japan. The history of how these tariffs came to be applied is itself an important lesson in trade policy.

In 1979, Paul Volcker was appointed as Chairman of the Board of Governors of the Federal Reserve Bank in the United States. Inflation exceeded 10 percent at that time, and showed signs of getting worse, and so Volcker was committed to bring it down. A period of tight monetary policy followed, which led to very high interest rates, a strong dollar, and a deep recession beginning in January 1980. There may well have been long-term gains resulting from this policy, since following its recovery from the recession in late 1982, the United States began nearly two decades of expansion, interrupted only by a mild recession in 1990–91. But at the same time, one short-term cost was high in terms of unemployment, and one of the sectors hardest hit was automobiles.

Accordingly, in June 1980 the United Automobile Workers applied to the ITC for protection, under Section 201 of U.S. trade laws. A similar petition was received in August 1980 from Ford Motor Company. As described at the beginning of this chapter, Section 201 protection can be given when increased imports are a “substantial cause of serious injury to the domestic industry,” where “substantial cause” must be “not less than any other cause.” In fact, the ITC determined that the U.S. recession was a more important cause of injury in autos than were increased imports. Accordingly, it *did not* recommend that the auto industry receive protection.

With this negative determination, several congressmen from midwestern states continued to pursue import limits by other means. For cars imported from Japan, this protection took the form of a “voluntary” restraint on the quantity of imports, as described in the next chapter. For trucks imported from Japan, however, another form of

¹¹The change in domestic output due to tariffs is investigated in a monopolistic competition model by Helpman (1990), using a utility function that allows for expenditure on the differentiated good to change. He confirms that output of the domestic good can rise or fall due to the tariff. The latter possibility was first suggested by Markusen (1990), who referred to the phenomena as a “de-rationalizing tariff.” This might hold, for example, if the imported goods are intermediate inputs, and the tariff leads to reduced sales of the industry using the inputs. This was found to hold in a simulation model of the Mexican automobile industry by Lopez de Silanes, Markusen, and Rutherford (1994).

protection was available. During the 1970s, Japan had exported an increasing number of compact trucks to the United States, most as cabs/chassis with some final assembly needed. These were classified as “parts of trucks,” which carried a tariff rate of 4 percent, whereas “complete or unfinished trucks” carried a tariff rate of 25 percent. That unusually high tariff was a result of the “Chicken War” between the United States and West Germany in 1962. At that time, Germany joined the European Economic Community (EEC) and was required to adjust its external tariffs to match those of the other EEC countries. This resulted in an increase in its tariff on poultry, which was imported from the United States. In retaliation, the United States increased its tariffs on “complete or unfinished trucks” and other products, and so the 25 percent tariff on trucks became a permanent item of the U.S. tariff code. With prodding from the Congress, in 1980 the U.S. Customs Service announced that effective August 21 import lightweight cabs/chassis would be reclassified as complete trucks. This raised the tariff rates on nearly all Japanese trucks from 4 to 25 percent, a rate that remains in effect today.¹²

A second increase in tariffs occurred during the 1980s on heavyweight motorcycles (i.e., over 700 cc displacement), produced by Harley-Davidson. In 1983 this company also applied to the ITC for Section 201 protection. Harley-Davidson was affected not so much by the U.S. recession as by a long period of lagging productivity combined with intense competition from Japanese producers. Two of these competitors (Honda and Kawasaki) had plants in the United States and also imported, whereas two others (Suzuki and Yamaha) produced and exported from Japan. In the early 1980s, these firms were engaged in a global price war that spilled over into the U.S. market. As a result, inventories of imported heavyweight cycles rose dramatically in the United States: the ITC estimated that inventories as of September 1982 exceeded actual U.S. consumption during January–September of that year (USITC 1983, 13).

From our description of Section 201 at the beginning of the chapter, recall that protection can be recommended if increased imports are a “substantial cause of serious injury, *or threat thereof*, to the domestic industry.” The very high level of inventories held by Japanese producers was judged by the ITC to be a threat to the domestic industry, and protection was indeed granted. This protection took the form of a five-year declining tariff schedule: 45 percent effective April 16, 1983, and then declining annually to 35, 20, 15 and 10 percent, and scheduled to end in April 1988.¹³ In fact, Harley-Davidson petitioned the ITC to end the tariff early, after the 15 percent rate expired in 1987, by which time it had cut costs and introduced new and very popular products, and therefore restored profitability.

U.S. imports of compact trucks and heavyweight cycles from Japan provide ideal industry cases to study the pass-through of the tariffs to U.S. prices. We are especially interested in whether Japanese producers absorbed part of the tariff, which would be a terms of trade gain for the United States, or whether they fully passed it through to U.S. prices. There are few studies of the pass-through of tariffs, but many studies of the pass-through of exchange rates, that is, the response of import prices to change in the

¹²Japanese producers sued in U.S. court to have the reclassification reversed, but lost that case.

¹³Actually, the protection took the form of a “tariff-rate quota,” which means that the declining tariffs rates were applied to imports of heavyweight motorcycles from each source country when imports *exceeded* a quota limit for the exporting country. The United States was also importing heavyweight cycles from BMW in Germany, but that country was granted a quota large enough that none of the imports were subject to the tariff. In contrast, Japan was allocated a quota (ranging from 6,000 to 10,000 units per year) that was less than its imports, so that declining tariffs applied to all imports in excess of this amount.

value of the exporter's currency. Burstein and Gopinath (2014) provide a recent survey of this topic. Most studies at either the aggregate or disaggregate level suggest that the pass-through of exchange rates is *less than* complete, and averages about 0.6 (Goldberg and Knetter 1997, 1250), though this depends on the industry being studied.¹⁴ We will argue below that the pass-through of tariffs and exchange rates ought to be "symmetric" in a given industry, and will test this hypothesis for trucks and motorcycles. In addition, we will include U.S. imports of Japanese cars before the "voluntary" export restraint, though for that product we will estimate only the pass-through of the exchange rate.

ESTIMATING THE TERMS OF TRADE EFFECT

To obtain an estimating equation, let us begin with the first-order condition of a typical foreign exporter, in (8.28a), where we treat foreign marginal costs C^* as constant, and rewrite this as c^* , in the foreign currency. We need to convert this cost to the domestic currency using an expected exchange rate e , so we rewrite (8.28a) as

$$p\left(1 - \frac{1}{\eta^*}\right) = (1 + \tau)ec^*. \quad (8.32)$$

Using the assumption that the income elasticity of demand is unity, we express the elasticity of import demand, $\eta^*(p/q)$, as a function of the import/domestic price ratio. More generally, $\eta^*(p, q, I)$ will depend on consumer income, too, but is homogeneous of degree zero in (p, q, I) .¹⁵ So treating the domestic price, income, tariff rate, exchange rate, and foreign marginal cost as parameters, (8.32) is one equation in one unknown—the import price p . We can therefore solve for the import prices as a function of the parameters:

$$p = f[(1 + \tau)ec^*, q, I].$$

With $\eta^*(p, q, I)$ homogeneous of degree zero in (p, q, I) , it is readily verified that $f[(1 + \tau)ec^*, q, I]$ is homogeneous of degree one in its arguments. This is our first testable hypothesis, which checks the overall specification of the pricing equation. Second, notice that we have written the tariff, exchange rate, and marginal costs as multiplied together in the arguments of $f[(1 + \tau)ec^*, q, I]$, because that is how they appear in the first-order condition (8.32). This is another testable hypothesis. To implement both these tests, let us specify the pricing equation as a log-linear function of its arguments, which are indexed by time t :

$$\ln p_t = \alpha_t + \sum_{i=0}^L \beta_i \ln(c_t^* s_{t-i}) + \beta \ln(1 + \tau_t) + \sum_{j=1}^M \gamma_j \ln q_{jt} + \delta \ln I_t + \varepsilon_t. \quad (8.33)$$

The first term on the right-hand side, α_t , is a simple quadratic function of time, which will allow import price to change in a smooth fashion for reasons not specified

¹⁴The disaggregate results of Knetter (1989, 1993) show that exchange rate pass-through varies quite substantially across products and source countries, with U.S. exporters absorbing the *least* of any exchange rate change. Campa and Goldberg (2005) investigate the pass-through of exchange rates in broad import sectors of the OECD countries, and find average long-run elasticities ranging from 0.61 to 0.89 across sectors.

¹⁵See problem 8.5.

elsewhere in the equation. To obtain the second term, we specify that the expected exchange rate e_t is a weighted average of lagged spot rates, s_{t-i} , $i = 0, 1, \dots, L$ (in dollar/yen). This allows us to write $\ln(c_t^* s_{t-i})$ as having the coefficients β_i , $i = 0, 1, \dots, L$, the sum of which will indicate the total pass-through of the exchange rate to import prices. Next, we write the *ad valorem* tariff $\ln(1 + \tau_t)$ as having the coefficient β , which is the pass-through of the tariff. The hypothesis of *symmetric pass-through* of the tariff and exchange rate is therefore tested as

$$\sum_{i=0}^L \beta_i = \beta. \quad (8.34a)$$

The next terms appearing on the right-hand side of (8.33) are the prices $\ln q_j$ of domestic or rival imported products, $j = 1, \dots, M$. For Japanese imports, this would include the average price of U.S. products, as well as German imported cars or motorcycles. Finally, total consumer expenditure on the broad category of transportation equipment is included as a measure of "income" I . Notice that if the income elasticity of demand is unity, then the foreign elasticity $\eta^*(p, q)$ does not depend on I , and neither will the optimal price p_t , so $\delta = 0$. In any case, the hypothesis that the whole equation is homogeneous of degree one can be tested by

$$\sum_{i=0}^L \beta_i + \sum_{j=1}^M \gamma_j + \delta = 1 \quad (8.34b)$$

The results of estimating (8.33) using quarterly data for U.S. imports are shown in table 8.2. The first row indicates the sample period of quarterly data used for each product. There are two distinct samples used for heavy-weight motorcycles. The first sample consists of *interview* data reported by the ITC (USITC 1983, table 8; 1983–84). The advantage of this data is that it gives the unit-value of imports for *consumption*, inclusive of duty, for the major Japanese importers (Honda, Suzuki, and Yamaha). However, the disadvantage is that the data end in 1984:4 and include a small amount of German heavyweight motorcycles within the reported unit-value. A second source was collected at the border by the U.S. Department of Commerce giving the unit-value of import *shipments*, distinguishing Japanese and German heavyweight cycles up to 1987:1.¹⁶ We experiment with both data sources and also the pooled sample.

In the first regression of table 8.2 for Japanese imported cars, the coefficients on the exchange rate terms sum to 0.71, which is the estimate of the pass-through elasticity. For Japanese imported trucks, the sum of coefficients on the exchange rates is 0.63, while the coefficient on the tariff is 0.57. The hypothesis (8.34a) that these are equal is easily accepted, as is hypothesis (8.34b) that the pricing equation is homogeneous of degree one. When both of these restrictions are imposed, the pass-through elasticity (for the exchange rate or the tariff) becomes 0.58. Thus, there is strong evidence that the increase in the truck tariff was only partially reflected in U.S. prices: of the 21 percent increase, about $0.58 \cdot 21 = 12$ percent was passed through to U.S. prices, whereas the other 9 percent was absorbed by Japanese producers, leading to a terms of trade gain for the United States.

For Japanese imports of heavyweight motorcycles, however, the story is quite different.

¹⁶ Import shipments include motorcycles going into inventory. These data were adjusted to include the tariff.

TABLE 8.2: Regressions for Japanese Imported Products with Import Price as Dependent Variable

	Cars	Trucks	Cycles (consumption)	Cycles (shipments)	Cycles (pooled)
Period:	74.1–81.1	77.1–87.1	78.1–84.4	78.1–87.1	78.1–87.1
$c_i^* s_t$	0.44*	0.28*	0.29	0.80	0.45*
	(0.11)	(0.06)	(0.26)	(0.72)	(0.21)
$c_i^* s_{t-1}$	0.32*	0.14*	0.17*	-0.042	0.10
	(0.04)	(0.03)	(0.10)	(0.28)	(0.09)
$c_i^* s_{t-2}$	0.17*	0.06	0.12	-0.34	-0.031
	(0.08)	(0.050)	(0.15)	(0.57)	(0.16)
$c_i^* s_{t-3}$	-0.01	0.05	0.12	-0.083	0.042
	(0.05)	(0.03)	(0.08)	(0.23)	(0.087)
$c_i^* s_{t-4}$	-0.21*	0.10	0.19	0.72	0.32
	(0.10)	(0.08)	(0.23)	(0.82)	(0.22)
Exchange rate ^a	0.71*	0.63*	0.89*	1.05*	0.89*
	(0.10)	(0.08)	(0.36)	(0.56)	(0.22)
Tariff	—	0.57*	0.95*	1.39*	1.13*
	—	(0.14)	(0.22)	(0.30)	(0.16)
U.S. price	1.00	0.03	0.68	1.14	0.57
	(0.93)	(0.40)	(0.60)	(2.17)	(0.59)
German price	0.08	—	0.06	0.12	0.06
	(0.09)	—	(0.11)	(0.23)	(0.11)
Income	-0.03	-0.03	-0.23	-0.23	0.02
	(0.12)	(0.06)	(1.69)	(0.65)	(0.01)
N, K^b	29, 9	41, 9	28, 13	37, 13	65, 13
Adjusted R^2	0.988	0.989	0.907	0.769	0.833
Durbin-Watson	2.43	1.75	2.73	1.69	—

Notes: Standard errors are in parentheses.

*Significant at the 95 percent level with conventional t -test.

a. Sum of coefficients for $\ln(c_i^* s_{t-1})$, $i = 0, 1, \dots, 4$, where c_i^* is an aggregate of foreign factor prices, and s_{t-1} is the spot exchange rate (\$/yen).

b. N is the number of observation, and K the number of independent variables. Coefficients for time trends and quarterly dummies are not reported.

Regardless of which sample is used (consumption, shipments, or pooled), we find that the pass-throughs of the exchange rate or the tariff are both insignificantly different from unity! So while the hypothesis (8.34a) of “symmetric” pass-through is still confirmed, as is (8.34b), it is no longer the case that the United States experienced a terms of trade gain in this product. Rather, the *full amount* of the tariff in each year was passed through to U.S. prices.

What explains the differing results for compact trucks and heavyweight motorcycles? In the case of trucks, we note that prior to the increase of the tariff in August 1980, nearly all compact trucks sold in the United States were produced by Japanese firms, some of which were marketed through American auto companies. But after the tariff was imposed, U.S. producers introduced their own compact truck models, with characteristics very similar to those of the Japanese imports (see Feenstra 1988b). The Japanese producers (Isuzu and Mitsubishi) that had formerly been selling to American firms began to market compact trucks independently. In this environment of relatively intense competition, we can expect that Japanese firms would be reluctant

to pass through the full amount of the tariff and risk losing more market share in the United States.

In heavyweight motorcycles, by contrast, recall that there was already a global price war, so that prices were likely close to marginal cost. This leaves little room for Japanese producers to absorb part of the tariff. This is reinforced by the fact that the tariff was temporary, and that U.S. inventories were high: there would be little reason to sell to the United States at a reduced price in one year, if instead sales could be made out of inventory and some exports delayed to a later year when the tariff was lower.¹⁷ By this logic, it is not surprising that the pass-through of the tariff in motorcycles was complete.

How should we assess the efficacy of the tariff in these two products? The tariff on trucks led to a terms of trade gain, and if this exceeds the deadweight loss, then it would generate a welfare gain for the United States. This is not the case for the tariff on motorcycles, however, where there is a deadweight loss but no offsetting terms of trade gain. By our conventional welfare criterion, then, the tariff on trucks looks better than the tariff on heavyweight cycles. But this criterion ignores the fact that the tariff on motorcycles was *temporary*, whereas that on trucks is still in place today. Furthermore, it is well documented (see Reid 1990) that Harley-Davidson was on the brink of bankruptcy in 1982–83 and was able to secure a bank loan only after receiving protection, so the tariff may well have contributed to its continued survival. Its near-bankrupt status was due to problems of poor management and lagging productivity, while its revival after 1983 was due to the introduction of improved products and production techniques. It cannot be argued that this broad change in company practices was *caused* by the tariff, but it appears that the temporary tariff bought the company some breathing room.¹⁸ In view of the improved products later offered by Harley-Davidson (which were emulated by its Japanese rivals), the temporary tariff may well have contributed to long-run welfare gains for consumers.¹⁹ In contrast, the compact trucks introduced by American firms after the tariff were quite similar to the existing Japanese models, and brought little additional welfare gain (Feenstra 1988b).

FUNCTIONAL FORMS

In our discussion above, we solved for the price of imports $p = f[(1 + \tau)ec^*, q, I]$ as a function of tariff-inclusive marginal costs abroad $(1 + \tau)ec^*$, the price(s) of competing goods q , and consumer expenditure I . In the estimation we imposed a log-linear form onto this function, as in (8.33). While this approach was acceptable in the early literature, today it is expected that the functional form of the pricing equation should arise from the underlying function form of consumer demand (or the utility function).

¹⁷ It was still the case that exports from Japan of heavyweight motorcycles (over 700 cc) were positive in every year of the tariff. In addition, Japanese firms began to sell a 699 cc motorcycle in the United States, which was a way to evade the tariff. See the discussion in Irwin (2002, 135–37).

¹⁸ This is the view expressed by the chief economist at the ITC at the time: “if the case of heavyweight motorcycles is to be considered the only successful escape-clause, it is because it caused little harm and it helped Harley-Davidson get a bank loan so it could diversify” (Suomela 1993, 135, as cited by Irwin 2002, 136–37).

¹⁹ Feenstra and Taylor (2014, chapter 9) compute the deadweight loss of the protection in heavyweight motorcycles and compares it to the present discounted value of the welfare gain (measured by the increase in the stock market value of Harley Davidson). They conclude that this protection can be justified as a case of “infant industry” tariff.

There are a number of functional forms that are particularly convenient in obtaining the pricing equation.

Demand Functions

An inverse demand equation proposed by Bulow and Pfeiderer (1983) is

$$p(x) = \alpha - \beta x^\delta, \quad \alpha, \beta, \delta > 0.$$

We substitute this demand equation into (8.16), treating marginal costs as constant so that $\pi''(x) = 2p'(x) + xp''(x)$ is the slope of marginal revenue. Then we find the pass-through

$$\frac{dp}{dt} = \frac{p'(x)}{\pi''(x)} = \frac{-\beta \delta x^{\delta-1}}{-2\beta \delta x^{\delta-1} - \beta \delta (\delta-1)x^{\delta-1}} = \frac{1}{(1+\delta)} < 1.$$

Thus, this demand curve gives a constant pass-through of a specific tariff in *level* form, rather than elasticity form. This functional form for demand has been used recently by Atkin and Donaldson (2014) to study how *intra-national* transport costs are passed through to prices in two developing countries.

Can we also achieve constant pass-through in elasticity form? To answer this question, let us return to (8.20), which gives the general response of the foreign price to the ad valorem tariff. Expressing this in elasticity form we obtain

$$\frac{d \ln p^*}{d \ln (1+\tau)} = -\frac{\eta'(x)x}{(\eta-1)^2} \frac{d \ln x}{d \ln (1+\tau)} \frac{C^*'}{p^*}.$$

Assuming constant marginal costs, we can simplify this expression using $d \ln x / d \ln (1+\tau) = -\eta d \ln p / d \ln (1+\tau)$, and also $d \ln p^* / d \ln (1+\tau) = d \ln p / d \ln (1+\tau) - 1$ and the first-order condition (8.19), to obtain

$$\frac{d \ln p}{d \ln (1+\tau)} = \left[1 - \frac{\eta'(x)x}{(\eta-1)} \right]^{-1}.$$

It is immediately apparent from this expression that $d \ln p^* / d \ln (1+\tau)$ is constant and lies between 0 and unity if and only if $\eta'(x)x / (\eta-1)$ is a negative constant. We can express that condition as

$$\frac{d \ln [\eta(x)-1]}{d \ln x} = \frac{\eta'(x)x}{(\eta-1)} = -\beta, \quad \beta > 0.$$

Integrating this expression, it must be that $\eta(x) = 1 + \alpha x^{-\beta}$, $\alpha > 0$. If we can in turn integrate this expression to get a demand curve, then the pass-through in elasticity form will indeed be constant. To fix ideas, suppose we work with the additively separable utility function $\sum_i v(x_i)$ from Krugman (1979), discussed in chapter 5, where $\eta(x_i) = -v'(x_i) / [x_i v''(x_i)]$. Then the condition for constant pass-through is

$$-\frac{v'(x_i)}{x_i v''(x_i)} = 1 + \alpha x_i^{-\beta} \Leftrightarrow -\frac{v'(x_i)}{v''(x_i)} = x_i + \alpha x_i^{1-\beta}.$$

Consider the special case $\beta = 1$. Then we find that up to a constant, $v(x_i) = \ln(x_i + \alpha)$, so a logarithmic utility function with a displaced origin leads to constant pass-through in elasticity form. This utility function has been used by Saure (2009) and Simonovska (2015), for example. Substituting $\beta = 1$ into the expression above, we find that the pass-through equals one-half. But now we also know that demand curves can be generated with constant pass-through not equal to one-half, by solving the differential equation above for $\beta \neq 1$ to obtain $v(x_i)$.

More generally, rather than keeping the pass-through constant we might be interested in obtaining markups and a degree of pass-through that varies systematically with the price or quantity of firms. The procedure above can be used to obtain such demand curves. We are also interested in the properties of markups over the business cycle, where it is usually thought that markups are pro-cyclical: rising during a boom in the economy and falling during a recession. Kimball (1995) proposes a framework to obtain markups that vary systematically in a macroeconomic model, as is applied by Klenow and Willis (2006); see also the treatment in Burstein and Gopinath (2014, 427–28). Flexible families of expenditure and demand functions that lead to nonconstant markups and nonconstant pass-through in international trade models are explored by Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012), Fabinger and Weyl (2014), and Feenstra (2014). Specific examples are given in the next two sections.

Nested CES Utility Function with Finite Number of Firms

Atkeson and Burstein (2008) use a nested CES utility function with a finite number of firms to obtain incomplete pass-through.²⁰ They assume that N^* foreign firms i export the quantity x_{in} to the home country in sector n , while N domestic firms j sell the quantity y_{jn} in that sector.²¹ These quantities are combined in a CES fashion with elasticity σ ,

$$X_n = \left(\sum_{i=1}^{N^*} x_{in}^{(\sigma-1)/\sigma} + \sum_{i=1}^N y_{jn}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)} \quad \sigma > 1. \quad (8.35a)$$

The CES aggregates X_n in each sector are then used to obtain utility

$$U = \left(\int_0^1 X_n^{(\varepsilon-1)/\varepsilon} dn \right)^{\varepsilon/(\varepsilon-1)} \quad \sigma > \varepsilon > 1, \quad (8.35b)$$

where $\varepsilon < \sigma$ is the elasticity of substitution across the sectors.

To obtain the demand for each differentiated export consumed at home, let us start at the sector level. Analogous to equation (6.1), demand in each sector is $X_n = (Y/P)(P_n/P)^{-\varepsilon}$, where Y denotes total expenditure, P is the overall price index, and

²⁰ Atkeson and Burstein (2008) and Burstein and Gopinath (2014, 428) attribute this method for obtaining incomplete pass-through to Dornbusch (1987). An early example is Feenstra, Gagnon, and Knetter (1996), who use CES preferences over a discrete number of products to obtain a U-shaped relationship between pass-through and the exporting country's share in the destination market; a more recent example is Auer and Schoenle (2012). Alessandria (2004) introduces search costs for consumers, and obtains a pass-through function that is also U-shaped in the firm's market share. Garetto (2012) expands on a similar framework to encompass heterogeneous firms.

²¹The numbers of firms N and N^* can differ across sectors, but we do not make this explicit.

P_n is the price index in sector n . As in chapter 5, equation (5.14), the theoretical price index P_n equals the expenditure function in each sector and is given by

$$P_n = \left(\sum_{i=1}^{N^*} p_{in}^{1-\sigma} + \sum_{j=1}^N q_{jn}^{1-\sigma} \right)^{1/(1-\sigma)},$$

where p_{in} and q_{jn} are the prices of the imported and domestic goods, respectively. Then demand for each foreign variety equals $x_{in} = (p_{in}/P_n)^{-\sigma} X_n$, which is analogous to (5.24) in chapter 5. Combining these various equations, the spending on each foreign variety is $p_{in} x_{in} = (p_{in}/P_n)^{1-\sigma} P_n X_n = (p_{in}/P_n)^{1-\sigma} (P_n/P)^{1-\varepsilon} Y$.

We are now in a position to compute the elasticity of demand for a foreign variety. Atkeson and Burstein (2008) assume that there are a finite number of firms producing in each sector (i.e., small N and N^*), so even if there was a *single* sector then the elasticity of demand for a given variety would not equal σ . Rather, the elasticity of demand for each foreign variety would equal $-\partial \ln x_i / \partial \ln p_i = \sigma(1 - s_i) + s_i$, where s_i equals the share of demand for that variety in the sector.²² With a continuum of sectors, the elasticity is slightly more complicated. Because there is a finite number of firms, the change in the price of any one foreign product *will* affect the sector price index P_n ; but because there is a continuum of sectors, there is no effect on the overall price index P (or on total expenditure Y). To compute the elasticity of demand, we use spending $\ln(p_{in} x_{in}) = (1 - \sigma) \ln p_{in} + (\sigma - \varepsilon) \ln P_n + (\varepsilon - 1) \ln Y$, and differentiate to obtain

$$\begin{aligned} \eta_{in} &= -\frac{\partial \ln x_{in}}{\partial \ln p_{in}} = 1 - \frac{d \ln(p_{in} x_{in})}{d \ln p_{in}} \\ &= 1 - (1 - \sigma) - (\sigma - \varepsilon) \frac{\partial \ln P_n}{\partial \ln p_{in}} \\ &= \sigma(1 - s_{in}) + \varepsilon s_{in}, \end{aligned}$$

where $s_{in} = \partial \ln P_n / \partial \ln p_{in}$ is the share of expenditure for foreign variety i *within* sector n , obtained as the elasticity of the expenditure function.²³ Simply stated, s_{in} is the sector market share for the foreign firm exporting product i .

Our assumption that $\sigma > \varepsilon$ implies that as the market share s_{in} rises, the elasticity of demand η_{in} will fall, and so the markup of the foreign firm rises. So in contrast to the simple CES case with a large number (or a continuum) of firms, the case of a finite number of firms allows for markups that are rising for firms with larger market shares. Furthermore, using the equilibrium price for the foreign exporter, $p_{in} = [\eta_{in}/(\eta_{in} - 1)] c_{in}(1 + \tau_n)$, we can substitute for the elasticity η_{in} to obtain

$$p_{in} = \left[1 + \frac{1}{\sigma(1 - s_{in}) + \varepsilon s_{in} - 1} \right] c_{in}(1 + \tau_n). \quad (8.36)$$

Because $\sigma > \varepsilon$, it is apparent that the markup in brackets is *increasing* in the market share of the firm. A foreign firm facing a decrease in its tariff or a depreciation of its currency will gain market share through lowering its price, and therefore increase

²²See problem 5.2.

²³The derivative of the theoretical price index (or the expenditure function) equals quantity, so that $\partial \ln P_n / \ln p_{in} = (\partial P_n / p_{in})(p_{in}/P_n) = s_{in}$.

its markup. In addition, as shown by Burstein and Gopinath (2014, 428–29) and by Amiti, Itskhoki, and Konings (2012), the (positive) elasticity of the markup with respect to the relative price of a firm is also increasing in the market share of the firm.²⁴ Therefore, large exporting firms will increase their markups even more in response to a depreciation of their exchange rate.

Amiti, Itskhoki, and Konings (2012) log-linearize the above pricing equation to obtain an estimating equation where the market share of each exporting firm appears on its own in the estimating equation, and also interacts with the exchange rate change. Firms with higher market share are expected to have a *lower* pass-through of changes in exchange rates into destination prices, because they instead adjust their markups by more. They find that this hypothesis is accepted using Belgian firm-level data. In addition, they find that the *import intensity* of the firms are important for determining the degree of pass-through: firms that import more intermediate inputs suffer an increase in costs when their exchange rate depreciates, so they are able to pass-through less of that depreciation to lower prices. Specifically, they find that small exporters with no imported inputs have a nearly complete pass-through, of over 90 percent. In contrast, the larger exporters have both high market shares abroad and high imported inputs, and these firms have a pass-through of about one-half.

The sensitivity of pass-through to the use of intermediate inputs was earlier emphasized by Gron and Swenson (1996). Differing degrees of pass-through to tariff reductions, depending on whether the tariffs are being reduced in outputs or on inputs, has been found recently by De Loecker, Goldberg, Khandelwal, and Pavcnik (2014), as will be discussed in chapter 11.

Translog Expenditure Function

The translog expenditure function with symmetry imposed across the parameters has the share equation for good i ,²⁵

$$s_i = (1/N) + \gamma(\bar{\ln p} - \ln p_i), \quad \gamma > 0,$$

where N is the number or mass of products and $\bar{\ln p}$ is the average over the log prices. If N is large (or if there is a continuum of products), then $\bar{\ln p}$ will hardly be affected by the change in the price of a single product, and so treating that effect as zero, the elasticity of demand is

$$\eta_i = 1 - \frac{d \ln s_i}{d \ln p_i} = 1 + \frac{\gamma}{s_i}.$$

Thus, the elasticity of demand is inversely related to a product's market share. It follows that products with a higher market demand will also have lower elasticities and higher markups.

Suppose that a foreign firm faces an ad valorem tariff for sales into the home country. Treating the prices of all other firms as fixed under Bertrand competition, and denoting its constant marginal costs by c , its tariff-inclusive price is

²⁴See problem 8.8.

²⁵See problem 5.6.

$$p_i = \left(\frac{\eta_i}{\eta_i - 1} \right) c_i (1 + \tau) = \left(1 + \frac{s_i}{\gamma} \right) c_i (1 + \tau). \quad (8.37)$$

Totally differentiating this expression, we obtain

$$\frac{d \ln p_i}{d \ln (1 + \tau)} = \left(\frac{\gamma + s_i}{2\gamma + s_i} \right) \approx \frac{1}{2} \text{ for } N \text{ large.}$$

The approximation above holds for large N , in which case the share of each product is small. We therefore obtain a pass-through in elasticity form of approximately one-half.

Without using this approximation, we see from the above equations that: (i) the markup of a firm is increasing in its market share, which is the same result as what we found for the nested CES case, that is, comparing (8.36) and (8.37); (ii) the pass-through of a change in tariffs (or exchange rate) is *increasing* in the market share of the exporting firm for the translog case.²⁶ That result is the opposite of what we found for the nested CES case, where firms with higher market share are expected to have a *lower* pass-through. From an efficiency point of view, it is desirable that a drop in tariffs due to trade liberalization leads to the greatest reduction in markups for the (domestic and foreign) firms with the greatest market shares and highest markups. Feenstra and Weinstein (2010) use the translog expenditure function to estimate the gains to the United States from the reduction in markups as domestic firms face import competition and reduce their markups. They find that the welfare gain from the reduction in markups are of similar magnitude to the welfare gain from increased import variety, and that the *sum* of these two gains in the translog case is comparable to the welfare gain from import variety alone in the CES case (as analyzed by Broda and Weinstein 2006).

INFANT INDUSTRY PROTECTION

Our discussion of the motorcycle tariff in the United States suggests a situation where protection may have allowed the domestic industry to avoid bankruptcy. Without arguing that this was definitely the case, let us use it as a possible example of “infant industry” protection. Theoretically, infant industry protection occurs when a tariff in one period leads to an increase in output, and therefore to a reduction in future costs, sufficient to allow the firm to survive, whereas otherwise it would not. Such protection is a very old idea, dating back to Hamilton (1791), List (1856), and Mill (1909).²⁷ An essential assumption of the argument is that the firm needs to earn positive profits each period to avoid bankruptcy. That is, there must be some reason that the capital market does not allow the industry to cover current losses by borrowing against future profits. A model of infant industry protection is developed by Dasgupta and Stiglitz (1988), and more recent treatment is in Melitz (2005). A historical example from the U.S. steel rail industry is provided by Head (1994).

An infant industry is an example of *declining marginal costs*, that is, when the future marginal costs are a decreasing function of current output. We have pretty much

²⁶See problem 8.8.

²⁷Cited by Baldwin (1969), who provides a number of reasons why the “infant industry” argument is *not* valid. See also the survey by Corden (1984, 91–92).

ignored this case throughout the chapter and have treated the marginal costs as either constant or increasing in output. When marginal costs are declining, however, then there may be additional scope for “strategic” trade policies. Krugman (1984) uses a model of declining marginal costs to argue that import promotion might act as export promotion: that protecting an import industry today might turn it into an export industry tomorrow. This intriguing idea is investigated for the production of random access memory chips by Baldwin and Krugman (1988a), and further empirical work would be desirable.

DUMPING

If there is a case to be made for infant industry protection, whereby an increase in the import price allows a firm to survive, then the reverse should also be true: a decrease in the import prices might lead a firm to shut down. This would be an example of “predatory dumping,” in which a foreign exporter lowers its prices in anticipation of driving rivals in the domestic country out of business. A model of predatory dumping is developed by Hartigan (1996), and like the infant industry argument, it relies on a capital market imperfection that prevents the home firm from surviving a period of negative profits.

Predatory dumping is presumably rare. In contrast, allegations of dumping together form a widespread phenomenon and are growing ever more common. Furthermore, charges of dumping are often made against trading partners in the same industry; for example, the United States will charge European countries and Japan with dumping steel in the United States, and likewise, those other countries will charge the United States with dumping steel there! This does not sound like “predatory dumping,” but must have some other rationale.

In his classic list of reasons for dumping, Jacob Viner referred to “long-run” or “continuous” dumping, to “maintain full production from existing facilities without cutting prices” (1966, 23, cited by Staiger and Wolak 1992, 266). This can occur in markets with oligopolistic competition and excess capacity. Ethier (1982b) presents a model emphasizing demand uncertainty and excess capacity, leading to dumping. But subsequent literature has focused on a simpler framework without uncertainty, where dumping is a natural occurrence under imperfect competition as oligopolists enter each other’s markets. This is demonstrated in the next section, using the “reciprocal dumping” model of Brander (1981) and Brander and Krugman (1983).

RECIPROCAL DUMPING

To model the oligopolistic competition between foreign and domestic firms, we return again to the Cournot model used earlier in the chapter, but now allow for a number of firms N in the home country and N^* abroad. The home firms sell y_i in the home market and export $y_i^*, i = 1, \dots, N$, while the foreign firms sell x_j in the home country and x_j^* in their own local market, $j = 1, \dots, N^*$. The equilibrium price in the home market will be denoted by $p(z)$, where $z = \sum_{i=1}^N y_i + \sum_{j=1}^{N^*} x_j$, and the foreign price is $p^*(z^*)$, where $z^* = \sum_{i=1}^N y_i^* + \sum_{j=1}^{N^*} x_j^*$.

Let us also assume that there are “iceberg” transportation costs involved in delivering the product from one country to the other, as in chapters 5 and 6, but for clarity we

now use T rather than τ to denote these costs: the amount $T > 1$ of the product must be shipped in order for one unit to arrive, so that $(T - 1)$ “melts” along the way. The home price $p(z)$ denotes the c.i.f. (cost, insurance, freight) price, while the f.o.b. (free on board) price received by the foreign exporter per unit shipped is p/T . This can be compared to p^* , which is the price received by the foreign firms in its own market. Notice that when computing prices to assess allegations of dumping, the U.S. Department of Commerce (DOC) works with these f.o.b. prices; that is, it *deducts* transportation costs from import prices before assessing whether dumping has occurred. So it would conclude that dumping has occurred at home if $p/T < p^*$. Similarly, the f.o.b. price received by the home firm per unit exported is p^*/T , and it could be charged with dumping in the foreign market if $p^*/T < p$.

From this setup, it is apparent that if we had an equilibrium where the c.i.f. prices were equal across the countries, $p = p^*$, then “reciprocal dumping” must be occurring, that is, $p/T < p^*$ and $p^*/T < p$. This was the outcome obtained by Brander and Krugman (1983) in a symmetric model, where the home and foreign markets were equal in size and had identical marginal costs. More generally, we are interested in whether reciprocal dumping will occur more in markets that are not necessarily the same size. Let us suppose that the home and foreign firms have identical marginal costs, but that the demand curves $p(z)$ and $p^*(z^*)$ are not the same. When will we observe reciprocal dumping?

To answer this question, denote the marginal cost of producing in either country by c . The fixed costs of production are α . A firm located in country i and selling to country j will face “iceberg” transport costs, so the marginal cost of delivering one unit abroad is cT . The home firms solve the profit maximization problem:

$$\max_{y_i, y_i^*} \pi = [p(z) - c]y_i + [p^*(z^*) - cT]y_i^* - \alpha,$$

with the first-order conditions

$$\pi_{y_i} = p(z) + y_i p'(z) - c = 0,$$

$$\pi_{y_i^*} = p^*(z^*) + y_i^* p^{*'}(z^*) - cT = 0.$$

Let us impose the symmetry condition that home firms are selling the same amount in the home market, $y_i = y$, and in the foreign market, $y_i^* = y^*$. Then the first-order conditions are written as

$$p\left(1 - \frac{(y/z)}{\eta}\right) = c, \quad p^*\left(1 - \frac{(y^*/z^*)}{\eta^*}\right) = cT, \quad (8.38)$$

where $\eta = z/p'(z)$ and $\eta^* = z^*/p^{*'}(z^*)$ p^* are the home and foreign demand elasticities.

In a similar manner, we can derive the first-order conditions for the foreign firms,

$$p^*\left(1 - \frac{(x^*/z^*)}{\eta^*}\right) = c, \quad p\left(1 - \frac{(x/z)}{\eta}\right) = cT, \quad (8.39)$$

where these firms are selling x^* in their own market and x in the home market.

By adding up the market shares of all firms selling in each country, we must obtain

$$\begin{aligned} N(y/z) + N^*(x/z) &= 1, \\ N(y^*/z^*) + N^*(x^*/z^*) &= 1. \end{aligned} \quad (8.40)$$

Notice that we can write this simple system in matrix notation as

$$(N, N^*) \begin{bmatrix} (y/z) & (y^*/z^*) \\ (x/z) & (x^*/z^*) \end{bmatrix} = (1, 1), \quad (8.40')$$

which can be inverted to solve for the number of firms in each country:

$$(N, N^*) = \frac{(1, 1)}{D} \begin{bmatrix} (x^*/z^*) & -(y^*/z^*) \\ -(x/z) & (y/z) \end{bmatrix} \Rightarrow \begin{aligned} N &= \frac{1}{D} \left(\frac{x^*}{z^*} - \frac{x}{z} \right), \\ N^* &= \frac{1}{D} \left(\frac{y}{z} - \frac{y^*}{z^*} \right), \end{aligned} \quad (8.41)$$

where D is the determinant of the two-by-two matrix of market shares shown in (8.40'). From the first-order conditions (8.38) and (8.39), it is clear that $(y/z) > (x/z)$ and $(x^*/z^*) > (y^*/z^*)$, so that local sales in each county exceed imports (due to the assumption of equal marginal costs but positive transport costs). This ensures that $D > 0$.

Notice that the solutions for N and N^* from (8.41) are not guaranteed to be positive, though this occurs if each firm sells more in its local market than in its export market, $(y/z) > (y^*/z^*)$ and $(x^*/z^*) > (x/z)$. To determine when this will be the case, let us make the simplifying assumption that the elasticity of demand is equal in the two markets, $\eta = \eta^*$. Then we can substitute the market shares from the first-order conditions (8.38) and (8.39) into (8.41) and use $\eta = \eta^*$ to obtain

$$N = \frac{\eta}{D} \left(\frac{cT}{p} - \frac{c}{p^*} \right) > 0 \text{ if and only if } p/T < p^*, \text{ and} \quad (8.42a)$$

$$N^* = \frac{\eta}{D} \left(\frac{cT}{p^*} - \frac{c}{p} \right) > 0 \text{ if and only if } p^*/T < p. \quad (8.42b)$$

Thus, we have obtained the following result.

THEOREM (WEINSTEIN 1992; FEENSTRA, MARKUSEN, AND ROSE 1998)

When the elasticities of demand and marginal costs of production are equal across countries, and firms in both countries are selling in both markets, then reciprocal dumping necessarily occurs: $p/T < p^*$ and $p^*/T < p$.

Remarkably, reciprocal dumping is guaranteed whenever the firms are selling in both markets! Our hypothesis that the elasticities of demand are equal across countries is a simplifying assumption: Weinstein (1992) establishes the theorem without this assumption (but using the weak condition (8.17'), that the market's marginal revenue

curve is steeper than the demand curve). Weinstein also discusses what happens as the number of firms in one country grows. As N^* grows, the equilibrium price p^* in the foreign market falls. For a sufficiently high number of foreign firms, the inequality $p/T < p^*$ becomes an equality, and at this point the number of domestic firms from (8.42a) is zero: import competition has eliminated the domestic industry. However, note from (8.42b) that we still have $p^*/T < p$, so that the foreign firms are dumping into the domestic market. Thus, even when the number of firms in one country has grown large enough to eliminate competition in the other, unilateral dumping still occurs.

WELFARE EFFECT OF TRADE

This model of reciprocal dumping is an alternative explanation for “intra-industry” trade, which we studied in chapter 5 under a monopolistic competition model. In that case, we argued that trade would bring extra gains, due to the increased variety of products and also economies of scale. But in the Cournot model we are studying, the firms are selling a homogeneous product so that there are no gains due to product variety, but instead losses due to wasted transportation costs in “cross-hauling.” On the other hand, trade brings a reduction in the exercise of monopoly power, due to the competition introduced from foreign firms. So this raises the question of which effect dominates: the social loss due to wasted transportation, or the social gain due to reduced monopoly power?

Brander and Krugman (1983) argue that with a *fixed* number of firms, moving from autarky to free trade has an *ambiguous* effect on global welfare. This is illustrated in figure 8.8. for a single home and foreign firm. Autarky occurs at the monopoly price p_0 . Provided that this price is above the marginal costs of exporting, cT , for the foreign firm, then with free trade the foreign firm will indeed sell to the home market.²⁸ Suppose that imports are $(z_1 - y_1)$, and these reduce the home price to p_1 . This brings a rise in global welfare of the area shown (the top triangle of this area is consumer surplus in the importing country, and the remainder is profits for the foreign exporter), but a loss in global welfare due to added transport costs shown by the rectangle. A similar diagram would apply in the foreign country. The relative sizes of these areas are ambiguous, so we do not know whether trade raises or lowers global welfare.

However, with free entry of firms in both countries, Brander and Krugman (1983) argue that global welfare necessarily improves due to free trade. The welfare function with multiple firms is $W(p, L + tm + N\pi)$ where π is the profits of each home firm, but with profits equal to zero this becomes $W(p, L + tm)$. Furthermore, comparing autarky and free trade, we have zero tariff revenue in each case, so welfare is just $W(p, L)$. Thus, welfare will rise going from autarky to free trade if and only if the import price falls. Brander and Krugman argue that this is indeed the case: with import competition, prices unambiguously fall, and so welfare in both countries improves. The fall in prices is also reflected in a fall in average costs (since price equals average costs from zero profits), which implies the output of each firm increases going from autarky to free trade. So similar to the monopolistic competition model of chapter 5, we have

²⁸As noted by Brander and Krugman (1983), the condition to observe trade with a single firm in each country is that the monopoly price, $p_0 = c\eta/(\eta-1)$ exceeds marginal cost of cT . This is written as $\eta/(\eta-1) > T$, or alternatively, $\eta < T/(T-1)$. So transport costs cannot be too high relative to the elasticity.

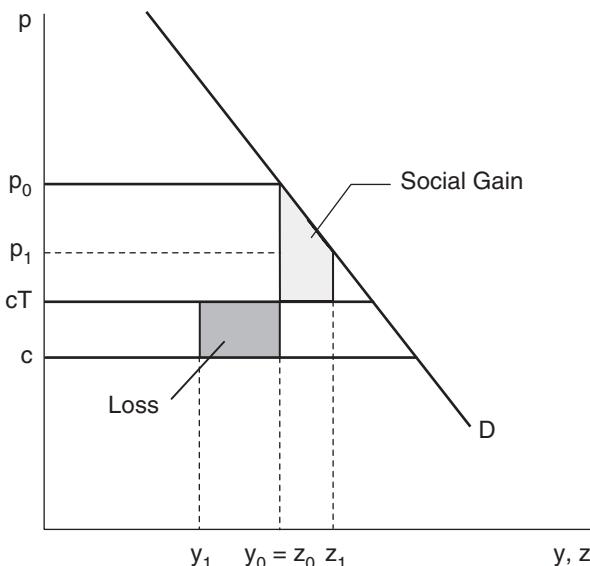


Figure 8.8

that trade brings gains due to the exercise of economies of scale. But unlike those in the monopolistic competition model, these gains come despite the fact that trade is intrinsically wasteful due to the homogeneous product and transportation costs.

IMPACT OF ANTIDUMPING DUTIES

Suppose that instead of comparing welfare under free trade and autarky, we instead consider the effect of an antidumping duty imposed by one country. When the number of firms is fixed at, say, one in each country, this corresponds to the case of Cournot duopoly analyzed earlier in the chapter. We found there that a small duty was likely to improve the terms of trade for the importing country, and may also contribute to an efficiency gain through increasing the output of the home firm. For both reasons, a small tariff was welfare improving for the importer. Should we think of antidumping duties in the same way?

The answer is no. Rather than leading to an improvement in the terms of trade, the application of antidumping law can often lead to a *worsening* of the terms of trade and a welfare loss for the importing country. To understand this, we need to review the administration of these laws in the United States. Antidumping actions go through several distinct phases (Staiger and Wolak 1994): an initial investigation by the DOC, which determines whether or not the imported product is being sold at “less than fair value”; followed by an initial investigation by the ITC to determine whether or not the domestic industry is “materially injured”; followed by a final determination of both agencies; followed by the application of duties if both findings are affirmative and the case is not withdrawn; and then also by an annual administrative review in cases where duties are imposed (Blonigen and Haynes 2002). Let us consider how each of these stages affects prices.

Initial Investigation

In the first stage of an investigation, which we call period 1, the DOC compares the prices of imported products with their prices (or costs) abroad. This calculation involves comparing the f.o.b. price received by the foreign exporter, which we denote by p_1/T , with its own home price of p^* . If $p_1/T < p^*$ then there is pricing at “less than fair value.” In nearly 95 percent of cases brought before the DOC, it concludes that this occurs and recommends duties in period 2 of $(1 + \tau_2) = p^*/(p_1/T) > 1$.

Notice that the duty applied in period 2 *depends* on the price charged by the exporter in period 1. In this sense, we must treat the antidumping duty as *endogenous*, and the exporter will have an incentive to raise its period 1 price so as to lower the period 2 duty.²⁹ Furthermore, this increase in price occurs even *before* the duty has been imposed, so that the importing country does not collect any tariff revenues. This increase in the import price amounts to a pure terms of trade loss for the importer. Because there is no tariff revenue collected, and still allowing for free entry of home firms, the welfare criterion is modified from $W(p, L + tm)$ to $W(p, L)$. Thus, any increase in the import price leads to a welfare loss.

There is evidence that this investigation stage of dumping actions does indeed lead to an impact on imports, though it is often measured by the effect on quantities rather than prices. Using a sample of all antidumping cases in the United States from 1980 to 1985, Staiger and Wolak (1992) find that the initiation of an investigation has a substantial impact on imports, reducing them by about one-half as much as that would have occurred under duties. The implication is that import prices must increase, leading to a loss for the importer.

Withdrawal of Cases

While the DOC finds evidence of “pricing at less than fair value” in most cases, the ITC rules in favor of “material injury” in only about one-half of the cases it considers. Thus, of some 1,150 cases in the United States during 1980–2011, about 450 were rejected by the ITC and another 550 had duties levied. As shown in table 8.1, the remaining 150 cases or 13 percent were *withdrawn* prior to a ruling by the ITC. In these cases the U.S. firms can negotiate with the foreign firm on the level of prices and market shares (Prusa 1991, 1992). Prusa finds that the withdrawal and settlement of cases has about the same impact on reducing import quantity as do actual duties, which is also found by Staiger and Wolak (1992).³⁰ Again, the implication is that import prices rise, with a direct welfare loss for the importing country.

Continuing Investigations

Consider now the case where an antidumping duty has been imposed, and denote the period 2 import price (inclusive of duty) by p_2 . Suppose that there is an administrative review to recalculate the amount of duty imposed in period 3. In calculating dumping margins during the administrative review, the DOC *removes* the tariff from the import price p_2 , and also removes any transport costs, obtaining $p_2/[T(1 + \tau_2)]$ as the f.o.b.

²⁹See problems 8.9 and 8.10 to demonstrate this in a model with Bertrand competition.

³⁰There are some cases that are withdrawn for procedural reasons and are *not* settled between the home and foreign firms. Taylor (2004) argues that these have little or no impact on import quantity or price.

price for the foreign exporter. This is compared to the exporter's home price of p^* to determine whether the foreign firm is continuing in its practice of "pricing at less than fair value." Whenever $p_2/[T(1+\tau_2)] < p^*$, duties of $(1+\tau_3) = p^*/[p_2/T(1+\tau_2)] > 1$ are imposed in period 3.

Note that these duties can be rewritten as $(1+\tau_3) = (1+\tau_2)[p^*/(p_2/T)]$, and so there is a built-in "continuation" of the antidumping duties: even if the foreign firm chooses $p^* = p_2/T$, so there will have been no initial evidence of dumping, the firm would still be faced with a tariff of $(1+\tau_3) = (1+\tau_2)$ in period 3. In other words, to avoid a charge of continued dumping during the administrative review, the foreign firm would not only need to increase its period 2 price above what was charged in period 1, it would need to further increase the period 2 price by the *full amount* of the period 2 duty. In the above notation, a charge of dumping is avoided in period 2 if and only if

$$p_2/[T(1+\tau_2)] \geq p^* \Rightarrow p_2 \geq p^* T(1+\tau_2) > p_1(1+\tau_2), \quad (8.43)$$

where the last inequality is obtained because dumping occurred in period 1, $p_1/T < p^*$.

Using the last inequalities in (8.43), we see that to avoid continuing antidumping duties the import price must rise between period by the amount $p_2/p_1 > (1+\tau_2)$, so there is *more than complete pass-through* of the period 2 duty. This hypothesis receives strong support from Blonigen and Haynes (2002). They estimate that the pass-through of the initial antidumping duties can be as high as 160 percent for the cases that are subject to an administrative review. This occurs because, as explained above, the DOC removes existing tariffs from import prices when computing the dumping margin. So in order to avoid duties after an administrative review, the tariff-inclusive price must rise by *more than* the amount of duties.

From a welfare point of view, this result very likely violates the criterion for welfare improvement. With the pass-through of duties exceeding unity, then $dp^*/dt > 0$, and so there is a terms of trade loss. Therefore, the only way that welfare can rise is if there is a large increase in domestic output with the associated efficiency gain (i.e., if the final term in (8.3) is positive). However, this source of gain vanishes if there is free entry into the domestic industry, so that profits are zero. In that case, the criterion for welfare gain is (8.26) rather than (8.3), and welfare can rise only if $dp/dt < m/d(p)$, meaning that the pass-through of the duty is less than the import share. This is clearly violated when the pass-through is greater than unity. Thus, in all three stages we have considered—initial investigation, withdrawal and settlement of cases, and continuing investigations—the presumption is that antidumping duties will lead to a welfare loss for the importer. This occurs because the duties being imposed are treated by the exporter as endogenous, and create an incentive to raise prices so as to avoid the duties.

Ruhl (2014) quantifies the magnitude of the welfare loss due to antidumping policy for the United States in 1992. In the calibrated model, he finds that antidumping policy reduces U.S. consumption (and therefore welfare) by 3.2 percent. As in our discussion above, that welfare loss comes not only from higher prices charged by firms subject to antidumping duties, but also from the higher prices that exporting firms charge in order to minimize the probability of being found guilty of dumping (i.e., pricing at "less than fair value") in the initial stage of the investigation. If we apply Ruhl's estimate of a 3.2 percent loss in consumption to U.S. GDP of \$6.3 trillion in 1992, we obtain a welfare cost of about \$200 billion in that year. If instead we focus on U.S. consumption of \$4.2 trillion in 1992, then the loss of 3.2 percent amounts to \$134 billion. That estimate is thirty-four times greater than the welfare cost of antidumping

and countervailing duties in the United States, which, Gallaway, Blonigen, and Flynn (1999) estimated at \$4 billion in 1993. In contrast to those authors, Ruhl builds into his model the price increases due to firms *avoiding* antidumping duties, whereas Gallaway, Blonigen, and Flynn consider only the *actual* antidumping and countervailing duties in effect in 1993. In addition, Ruhl allows for firm heterogeneity, which is important because antidumping duties apply at the firm-level and discriminate against the most productive (and hence lowest price) firms. It is plausible that these features result in a considerably larger estimate of the welfare cost from antidumping legislation, as found by Ruhl.

CONCLUSIONS

Since this chapter has covered a large number of different models, a summary is appropriate. Consider tariffs that are *exogenously* imposed, such as under the escape clause provision of the GATT/WTO. In general, this will have three welfare effects on the importing country: (a) a deadweight loss; (b) a terms of trade effect; (c) a reduction in the monopoly distortion if the output of home firms increase (without leading to inefficient entry). The deadweight loss always harms the importer, but is of the second order of smalls for a small tariff. The change in the output of home firms, sometimes referred to as a “profit shifting” effect, is of ambiguous sign and cannot be relied upon to generate gains for the importer. Accordingly, the best indicator of the welfare impact of small tariffs is the terms of trade effect.

It is sometimes argued that countries must be very large in their sales or purchases of a product on world markets for their tariff to have an impact on the terms of trade. Even the United States, it is argued, should be treated as “small” in nearly all markets.³¹ I disagree with this assessment. Under imperfect competition, it is quite plausible that foreign exporters will absorb part of the tariff so the import price does not rise by its full amount; our detailed analysis of the Cournot and Bertrand cases has identified the exact conditions under which this occurs. This depends on exporters treating their foreign markets as segmented (i.e., charging different prices to each), and holds regardless of the size of the importing country. The abundant evidence on partial pass-through of exchange rates (surveyed in Goldberg and Knetter 1997) suggests that the same is true for tariffs, though the exact magnitude of pass-through will depend on the industry being considered. For the comparison of the U.S. tariffs on truck and motorcycles we found quite different pass-through behavior, with the tariff on trucks—but not the tariff on motorcycles—generating a terms of trade gain.

Even if we accept that in many industries there will be a terms of trade and welfare gain due to a small tariff, this hardly justifies their use: the terms of trade gain comes *entirely* at the expense of foreign exports, so this is a “beggar thy neighbor” policy. The use of tariffs can therefore invite retaliation from trading partners. An example of such retaliation occurred in the steel industry in the United States, which received Section 201 protection in March 2002.³² Declining tariffs ranging from 8 to 30 percent were applied for three years to a range of steel products and various

³¹ See Irwin (2002, 63).

³² This was the first instance of a Section 201 action that was initiated by a U.S. president (and followed from a campaign promise made by George W. Bush while campaigning in steel-producing states).

importers.³³ Shortly after these tariffs were imposed, the European Union, Japan, and Korea announced plans to file protests with the WTO and retaliate with tariffs of their own applied against U.S. products. Such retaliation means that the use of tariffs becomes a negative-sum game; that is, they can end up harming both trading partners.

What can prevent a country from attempting to move the terms of trade in its favor? One view of the GATT is that it was designed with exactly this goal in mind. Bagwell and Staiger (1999, 2002) argue that the provisions of GATT such as “reciprocity” (i.e., tariff reductions should be reciprocal across countries) are effective in preventing countries from using tariffs to improve their terms of trade. We will discuss this further in chapter 10. By a similar logic, I think that the escape clause provision of the GATT should be viewed as a way to promote free trade while allowing for exceptions in specific circumstances. The linkage was made explicit in the Section 201 protection to the steel industry in the United States, with which the George W. Bush administration hoped to secure additional votes for “fast track” authority.³⁴ So Article XIX of the GATT and Section 201 of U.S. trade law need to be assessed in a broader political-economic context of securing industry support for free trade. Given that these provisions are used rarely, and only temporarily, they can be viewed as rather effective.

The same is not true of the antidumping provisions of GATT. As we have argued, the GATT definition of dumping will apply in many markets subject to imperfect competition, and there has been an increasingly widespread use of the dumping provisions. The antidumping duties that are imposed must be treated as *endogenous*: they depend on the prices chosen by the exporter prior to, and during, the dumping action. This creates an incentive for exporters to raise their prices before dumping duties are imposed, and to raise them even further before an administrative review of the dumping case. These actions correspond to a terms of trade loss for the importer, as indicated by the empirical evidence of Blonigen and Haynes (2002). This can also be reflected in an increase in the price of import-competing products, of course, so the antidumping duties promote more collusive behavior. In the next chapter, we will find that import quotas also can lead to more collusive behavior among firms, and so along with antidumping legislation, these policies must be judged as very costly.

PROBLEMS

- 8.1 Derive the optimal demand for the consumer with the quasi-linear utility function given by $c_0^h + U^h(c^h)$, where c_0^h is the consumption of a numeraire good and c^h is the consumption vector of all other goods, with U^h increasing and strictly concave. The consumer maximizes utility subject to the budget constraint $c_0^h + p'c^h \leq I^h$.

³³Some of the tariffs took the form of “tariff-rate quotas,” which means that the tariffs are applied only for imports from each source country *in excess* of some amount. These quotas can be varied preferentially across source countries without violating the MFN principle of the GATT. Canada, Mexico, Israel, and Jordan are entirely exempt from the tariffs because of their free trade agreements with the United States.

³⁴“Fast track” authority refers to the ability of the executive branch to present to the U.S. Congress a proposal for an expansion of free trade, for an “up or down” vote without allowing amendments. The executive branch had this authority from 1975 to 1994 and again from 2002 to 2007.

- 8.2 Derive the optimal specific tariff from (8.12) in the case where there is a single foreign firm, the demand curve $d(p)$ is linear, and marginal costs are constant, $C''(y)=0$.
- 8.3 Derive conditions under which the Cournot reaction curves in figure 8.4 are both downward sloping.
- 8.4 By totally differentiating the equilibrium condition (8.22), show that
- $dy/dt > 0$ provided that $p'(x+y) + yp''(x+y) < 0$;
 - The total change in home profits π due to the tariff is positive.
- 8.5 In this problem we derive the properties of the elasticities used with Bertrand competition.
- First review problem 1.2 of chapter 1 and show that if a function $y=f(v)$ is homogeneous of degree α , meaning that for all $\lambda > 0$, $f(\lambda v) = \lambda^\alpha f(v)$, then its first derivative is homogeneous of degree $\alpha - 1$.
 - Thus, if demand $d(p, q, I)$ is homogeneous of degree zero in (p, q, I) , then $d_q(p, q, I)$ is homogeneous of degree minus one. Use this to show that $\eta(p, q, I) \equiv -qd_q(p, q, I)/d(p, q, I)$ is homogeneous of degree zero in (p, q, I) .
 - Now assume that the income elasticity is unity, so that demand is written as $d(p, q, I) = \varphi(p, q)I$. Show that the elasticity η does not depend on I . Therefore, show that the elasticity η can be written as a function of q/p .
- 8.6 Using the results from problem 8.5, assume that $\eta''(p/q) > 0$ and $\eta'(q/p) > 0$.
- Then by differentiating the first-order conditions (8.28), show that the reaction curves $p = r^*(q, \tau)$ and $q = r(p)$ both slope upward, as shown in figure 8.6, with elasticities less than unity.
 - Also compute the elasticity of the foreign reaction curve $p = r^*(q, \tau)$ with respect to τ .
 - Use the results in (a) and (b) to solve for the total change in q and p due to the tariff.
- 8.7 Output of the home firm is $y = d(q, p, I)$.
- Use the fact that $d(q, p, I)$ is homogeneous of degree zero in (q, p, I) to show that $d_q q + d_p p + d_I I = 0$. Then use the assumption that the income elasticity of demand is unity to show that $d_p p / d = (\eta - 1)$, where $\eta \equiv -qd_q/d$.
 - Using part (a), compute the total change in home output $y = d(q, p, I)$ as p increases along the home reaction curve. Show that output increases if $0 < q\eta'/p\eta < (\eta - 1)^2$.
- 8.8 (a) Consider the markup shown in brackets in (8.36), where the market share of foreign firm i in sector n is given by $s_{in} = (p_{in}x_{in}/P_n X_n) = (p_{in}/P_n)^{1-\sigma}$. Compute the derivative of the natural log of the markup in (8.36) with respect to $\ln(p_{in}/P_n)$. Reversing the sign so that this elasticity is positive, is the elasticity of the markup increasing in the share of the firm?
- (b) Repeat part (a) but using the translog markup shown in (8.37). That is, compute the derivative of the natural log of the markup in (8.37) with respect to $\ln p_{in}$. Reversing the sign so that this elasticity is positive, is the elasticity increasing in the share of the firm?
- 8.9 Under Bertrand competition, show that foreign profits fall due to the tariff when $\eta''(p/q) > 0$ and $\eta'(q/p) > 0$.
- 8.10 Consider the problem of an exporting firm facing the threat of an antidumping duty. Given the period 1 import price chosen by the exporting firm, the

duty imposed in period 2 equals $(1+\tau_2) = p^*/p_1$ whenever $p_1 < p^*$. This tariff is imposed with probability θ , and conversely, with probability $(1-\theta)$ there will be no duty. Denote the value of period 2 profits when the duty is imposed by $\pi_2^*(\tau_2)$, where from problem 8.9 we have that $\pi_2^{*\prime}(\tau_2) < 0$. Let $\pi_2^*(0)$ denote the maximized value of period 2 profits for the foreign firm in the case of zero duty. Then the period 1 problem can be stated as

$$\max_{p_1} \pi^*(p_1, q_1) + \delta[\theta\pi_2^*(\tau_2) + (1-\theta)\pi_2^*(0)] \text{ s.t. } (1+\tau_2) = p^*/p_1. \quad (8.44)$$

We suppose that the home firm chooses q_1 , under Bertrand competition. Derive the first-order conditions for the home and foreign firms, and show that the threatened duty leads to an *increase* in the foreign price p_1 .

EMPIRICAL EXERCISE

In this exercise, you are asked to reproduce some of the empirical results from Feenstra (1989). To complete the exercise, the files “cars.csv, trucks.csv, cycon.csv, cyship.csv, cypool.csv” should be stored in the directory c:\Empirical_Exercise\Chapter_8\. Each of these can be used in STATA programs “cars.do, trucks.do, cycon.do, cyship.do, cypool.do” to create a dataset with the variables described in “Documentation_Ch8.doc.” Use these “do” programs to answer the following:

- 8.1 Replicate table 8.2, that is, run the specifications of (8.33) without imposing the tests of symmetry or homogeneity. Duplicate all of the coefficients that are reported in this table, except the Durbin-Watson statistics.
- 8.2 Then replicate Feenstra’s table 8.2 by imposing the tests of homogeneity and symmetry, shown in (8.34a) and (8.34b). Instead of conducting the Wald test, as done in Feenstra 1989, instead conduct the analogous F-test. Do you accept or reject the hypotheses of symmetry and homogeneity?

Import Quotas and Export Subsidies

In addition to import tariffs, quotas and subsidies are widely used forms of trade policy. Quotas and subsidies can in principle be applied to either imports or exports, but we will focus here on *import* quotas and *export* subsidies. How do these policy instruments differ from tariffs, and does this difference depend on the type of competition in the market? These are questions we shall investigate in this chapter.

We begin with the result of Bhagwati (1965) on the “equivalence” of tariffs and quotas. This means that under *perfect* competition, applying a quota that limits the number of units imported will have essentially the same effects as applying a certain level of tariff: for each quota, there is an “equivalent” tariff. This equivalence result no longer holds under *imperfect* competition, however. Bhagwati (1965) considered a monopolistic home importer subject to a quota, and that analysis was later extended to duopoly with a home firm and a foreign firm that is subject to an import quota (Harris 1985; Krishna 1989). As we shall see, a quota and tariff that have comparable effects on the level of imports can then have quite different effects on the import price, and therefore on welfare in the importing country.

Another reason why tariffs and quotas differ occurs when foreign firms can choose the *quality* of the good that they export. In that case, a limitation on the number of units exported (as under a quota) will have quite a different impact from a tariff on the value of exports (as with an ad valorem tariff). The importance of this case became apparent with the quota applied to Japanese auto exports to the United States during the 1980s. The evidence strongly supports the hypothesis that Japanese firms changed the characteristics of the cars they sold, shifting toward higher-quality models. We shall describe a theoretical model of quality choice and the empirical evidence for autos (Feenstra 1988a) and other industries (Feenstra and Romalis 2014).

The finding that trade policies have differing effects depending on the market structure carries over to the case of export subsidies. In the conventional two-sector model, there is no reason to use export subsidies: they will lead to a deadweight loss (analogous to a tariff) for a small country, and have an additional terms of trade loss (the opposite of a tariff) for a large country. This seems to contradict the fact that many countries have used export subsidies to support their industries at some time. To explain this, we need to go beyond the two-sector model with perfect competition. Adding more goods turns out to make a difference, and there

is a potential role for targeted export subsidies in models with many goods (Feenstra 1986; Itoh and Kiyono 1987).

Dropping perfect competition and instead allowing for duopoly between a home and foreign firm exporting to a third market also makes a difference. The question then is whether the home government can give its own firm a “strategic” advantage by subsidizing it. We shall require that the subsidy also be in the *national* interest, which means that profits for the exporter need to rise by more than the amount of the subsidy itself. Initial analysis of this problem (Brander and Spencer 1985) suggests that such an advantage would indeed occur, at least under Cournot-Nash competition. Later work (Eaton and Grossman 1986), however, shows that this advantage would be *reversed* under Bertrand competition. The conclusion is that export subsidies are in the national interest only under some forms of market competition, but not generally. We conclude the chapter by investigating the effects of subsidies to the production of commercial aircraft, produced by Boeing in the United States and Airbus in Europe. Irwin and Pavcnik (2004) have investigated the effects of subsidies on prices and profits in this industry, as we shall discuss.

EQUIVALENCE OF TARIFFS AND QUOTAS

The analysis of quotas under perfect competition proceeds much like our graphical treatment of tariffs, in the previous chapter. In panel (a) of figure 9.1 we show the domestic demand D and supply curve S , together with the constant world price p^* . Under free trade, domestic demand is at c_0 and supply at y_0 , so imports are $m_0 = c_0 - y_0$. The import demand curve $M = D - S$ is shown in panel (b), together with the horizontal export supply curve X at the world price p^* . These intersect at the equilibrium imports m_0 .

Now suppose that an import quota of \bar{X} is imposed, meaning that the quantity imported cannot exceed this amount. This effectively establishes a *vertical* export supply curve denoted by \bar{X} in panel (b), which now intersects import demand at m_1 with the price of p_1 . In panel (a), the price of p_1 leads to increased supply of y_1 and reduced demand of d_1 . Notice that there would be an equivalent effect on price, consumption, and production if instead the government had imposed a tariff of $t \equiv p_1 - p^*$. In this sense, for every quota there is an *equivalent* tariff.

The welfare change from having the quota in figure 9.1 is $-(a + b + c + d)$ consumer surplus loss + (a) producer surplus gain. However, the area c , which would be collected as revenue under a tariff, now needs to be considered more carefully. This area is the difference between the domestic price p_1 and the world price p^* . Therefore, whoever is actually importing the good stands to collect the pure profits or “rents” equal to $(p_1 - p^*)m_1$, which is area c . There are four possible ways that these quota rents can be allocated.

First, the quota licenses can be given to home firms, which are then able to import at the world price p^* and sell domestically at p_1 , earning the difference between these as rents. An example of this is the dairy industry in the United States, where U.S. producers of cheese also receive licenses to import from abroad.¹ With home firms earning the rents c , the net loss in domestic welfare due to the quota is $-(a + b + c + d)$

¹ As described in Hornig, Boisvert, and Blandford (1990), there is actually some sharing of the quota rents between importers and exporters and the U.S. government.

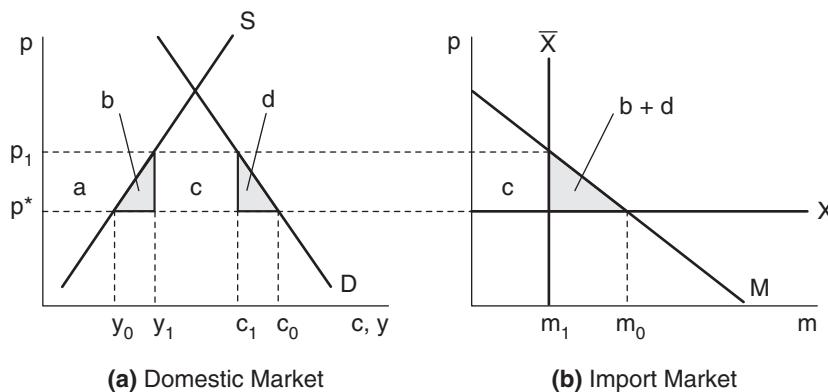


Figure 9.1

consumer surplus loss $+(a + c)$ producer surplus gain $= -(b + d)$, which is the same loss as we found for a tariff.

However, a second possibility is that while the quota licenses are given to home firms, these firms engage in some kind of inefficient activities in order to obtain the licenses. For example, if licenses for some imported input are allocated in proportion to the previous year's production of a final good, then firms will produce more final goods than they can sell (and probably of lower quality) just to obtain the import licenses. Krueger (1974) calls this kind of activity "rent seeking," and it can include lobbying and other resource-using activities. It has been suggested that the waste of resources devoted to rent-seeking activities could be as large as the value of rents themselves, so that area c would be wasted rather than accrue to domestic firms. In that case, the welfare loss due to the quota would be $-(a + b + c + d)$ consumer surplus loss $+(a)$ producer surplus gain $= -(b + d + c)$, which is greater than for the tariff.²

Third, the quota licenses can be auctioned by the government of the importing country. This has occurred in New Zealand, for example.³ In a well-organized auction, we might expect that the revenue collected would equal the value of the rents, so that area c accrues to the home government. In that case, the net loss in domestic welfare due to the quota is $-(a + b + c + d)$ consumer surplus loss $+(a)$ producer surplus gain $+(c)$ auction revenue $= -(b + d)$, which is again the same loss as we found for a tariff.

Finally, the government of the importing country can give authority for implementing the quota to the government of the *exporting country*. Because the exporting country allocates the quota among its own producers, this is sometimes called a “voluntary export restraint” (VER), or “voluntary restraint agreement” (VRA). This arrangement has occurred for quotas that the United States has used in automobiles, steel, textiles, and other industries. In automobile imports for Japan after 1981, for example, the Ministry of International Trade and Industry (MITI) told each auto manufacturer in Japan how much it could export to the United States. We have discussed the background to this case in chapter 8 and will analyze it in detail later in this chapter. At this point, we note that the quota rents are then earned by *foreign producers*, so the

²Bhagwati and Srinivasan (1980) have suggested, however, that there might also be "profit-seeking" activities for tariff revenue, in which case we ought to take this into account for a tariff, too.

³See Bergsten et al. (1987).

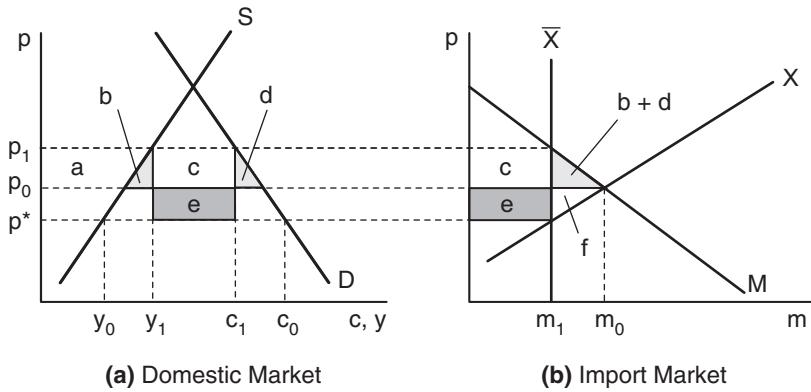


Figure 9.2

TABLE 9.1: Annual Cost of U.S. Import Protection (billion dollars, years around 1985)

	U.S. Deadweight Loss (area $b + d$)	Quota rents (c or $c + e$)	Foreign Deadweight Loss (area f)
Automobiles	0.2 – 1.2	2.2 – 7.9	0 – 3
Dairy	1.4	0.25	0.02
Steel	0.1 – 0.3	0.7 – 2.0	0.1
Sugar	0.1	0.4 – 1.3	0.2
Textiles and Apparel	4.9 – 5.9	4.0 – 6.1	4 – 15.5
Average Tariffs	1.2 – 3.4	0	—
Total ^a	7.9 – 12.3	7.3 – 17.3	4.3 – 18.8

Source: Feenstra (1992), with detailed sources cited there.

a. In dairy the quota rents are earned by U.S. importers, foreign exporters, and the U.S. government, and so are not included in the total.

loss in domestic welfare equals $-(a + b + c + d)$ consumer surplus loss +(a) producer surplus gain $= -(b + d + c)$, which is larger than for a tariff.⁴

In figure 9.2 we extend the analysis of the quota to a large country. In this case the quota rents might be measured as c or $(c + e)$, and these accrue to the *foreign producers* under a VER. The deadweight loss for the importing country is still measured by area $(b + d)$, and it can be argued that the area f represents an analogous deadweight loss for the *foreign* country. In table 9.1 we present some estimates of the home and foreign deadweight loss, along with the rents, for major U.S. quotas around 1985. In all cases except diary, these rents were earned by foreign exporters. We will be discussing the case of automobiles later in the chapter, and estimates of the quota rents earned by foreigners range from \$2 to 8 billion. Textiles and apparel also have very large quota rents and U.S. deadweight losses (about \$5 billion each), under the global system of quotas known as the Multi-Fiber Arrangement (MFA). The *foreign* deadweight losses due

⁴This raises the question of why a country would use a VER, since it gives the rents to foreigners. One answer is that this helps to prevent retaliation by the foreign countries. In addition, Feenstra and Lewis (1991b) argue that transferring the rents to foreigners serves as an “incentive compatibility” device, so that the importing country will use the VER only if the domestic industry is genuinely harmed.

to the MFA are even larger, with estimates ranging from \$4 to \$16 billion. The upper end of these estimates is obtained from models that include very costly rent-seeking activities by foreign producers in developing countries. A recent study by Khandelwal, Schott, and Wei (2013) confirms that the MFA quotas in China were allocated in an inefficient manner.

Summing the costs shown in table 9.1, the total deadweight loss due to these quotas was in the range of \$8 to \$12 billion annually in the mid-1980s, while the quota rents transferred to foreigners were another \$7 to \$17 billion annually. Many of these costs for the United States are no longer relevant today: the quota in automobiles ceased to be binding after 1987, as discussed later in the chapter, and under the Uruguay Round of the WTO the MFA was dismantled. The quotas in the steel industry, whose costs are summarized in table 9.1, were replaced by temporary tariffs (as described at the end of chapter 8). In addition, the “average tariffs” included in the last row of table 9.1 reflect only conventional tariffs, and do not include the additional cost due to U.S. antidumping duties (estimated at \$4 billion annually or considerably more, also discussed at the end of chapter 8).

NONEQUIVALENCE DUE TO IMPERFECT COMPETITION

We have shown the equivalence of tariffs and quotas, in the sense that when each instrument is chosen to result in the same import level, it also has the same impact on prices. With imperfect competition in the domestic market, however, this result quickly breaks down. The reason is that the quota creates a “sheltered” market for the domestic firm, which will lead to higher prices and lower sales than under a tariff with same level of imports. The case of a domestic monopoly, as analyzed by Bhagwati (1965), is demonstrated in figure 9.3.

With the fixed world price of p^* , the demand curve facing the domestic monopolist is essentially horizontal at that price, so marginal revenue is also horizontal and the profit-maximizing quantity is y_0 , where marginal revenue equals marginal cost. Note that this is the *same quantity* that a competitive firm or industry would produce if

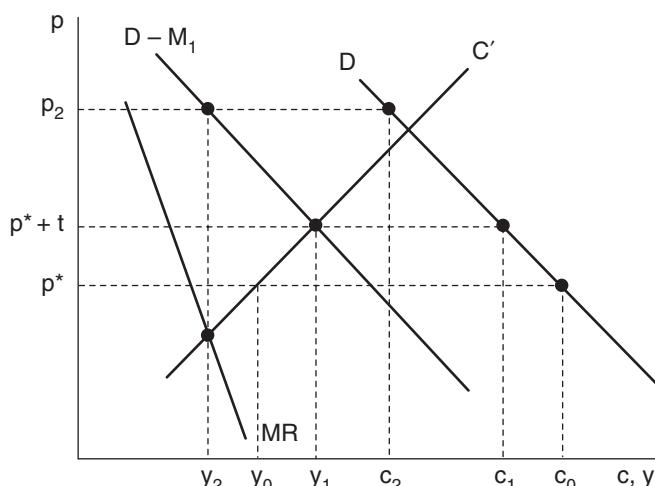


Figure 9.3: Domestic Monopoly

it had the same marginal costs as the monopolist. In this sense, free trade in a small country eliminates the market power of the monopolist, that is, eliminates its ability to restrict supply and raise price. This is a potential extra source of gains from trade.

Suppose initially that a *tariff* of t is applied to imports. In that case, the domestic monopolist can charge as much as $p^* + t$, but no more, so the demand curve is now horizontal at that price. This means that the marginal revenue curve is also horizontal at $p^* + t$, so the profit-maximizing quantity is where that price equals marginal cost, which is y_1 in figure 9.3. Consumption at that price is c_1 , so imports are $m_1 = c_1 - y_1$.

Now suppose that instead of the tariff, a quota of m_1 is applied. This means that for any price above p^* , the fixed amount m_1 will be imported, so the demand curve facing the monopolist is the initial demand curve D less the amount m_1 , which we illustrate by $D - M_1$. Unlike the situation under the tariff, the monopolist retains the ability to influence the domestic price: it can choose the optimal price and quantity along $D - M_1$. So we graph the marginal revenue curve below $D - M_1$, which is denoted by MR , and find the profit-maximizing price and quantity where this intersects marginal cost, at p_2 and y_2 . It will definitely be the case that $p_2 > p^* + t$, so that the quota *with the same level of imports as the tariff* has led to a higher domestic price. Therefore, the tariff and quota are no longer equivalent: the quota enables the monopolist to exercise market power, and has a higher welfare cost for this reason.

It is also the case that the quota might even lead to a *fall* in domestic output as compared to free trade. This is illustrated in figure 9.3 by $y_2 < y_0$. This is not a necessary result, and we could instead have drawn the MR curve so that $y_2 > y_0$. It is rather surprising that $y_2 < y_0$ is even possible, because it suggests that workers in the industry would *fail to be protected* by the quota: that is, employment could fall due to the reduction in output under the quota. We see, then, that this policy instrument has quite undesirable results as compared to a tariff.

These results of Bhagwati (1965) give us good insight into the effect of a tariff versus a quota under imperfect competition, but the assumption that the domestic monopoly does not interact at all with foreign firms is not satisfactory. Presumably, the domestic firm competes with foreign firms, under Cournot or Bertrand competition. In the first case, Cournot-Nash competition, the import quota has rather predictable effects: restricting the foreign firms to sell less than its free trade amount will shift sales toward the domestic firm, under the normal shape of reaction curves.⁵ In order to obtain more dramatic results, we instead consider an import quota (restricting quantity) in a game where the two firms are initially engaged in Bertrand competition (taking each other's prices as given). This case has been analyzed by Harris (1985) and Krishna (1989), and we follow the exposition of Harris.

As in the previous chapter, we assume that the imported and domestic products are imperfect substitutes, and denote their prices by p and q , respectively. The amount exported by the foreign firm is x , and the amount sold by the domestic firm is y . Under the assumption that the elasticity of demand for each product is increasing in the relative price of that product, the reaction curves $p = r^*(q)$ and $q = r(p)$ both slope upward, with profits increasing in the directions shown in figure 9.4(a). As the relative price of the competing good rises, the elasticity falls, and each firm will charge a higher price for its own good. Furthermore, there is a *dampened* response of each price to that of the competing good, so that $\frac{dp}{dq} \frac{q}{p} = r^*(q) \frac{q}{p} < 1$, and $\frac{dq}{dp} \frac{p}{q} = r(p) \frac{p}{q} < 1$.⁶

⁵See problem 8.1.

⁶See problem 8.6.

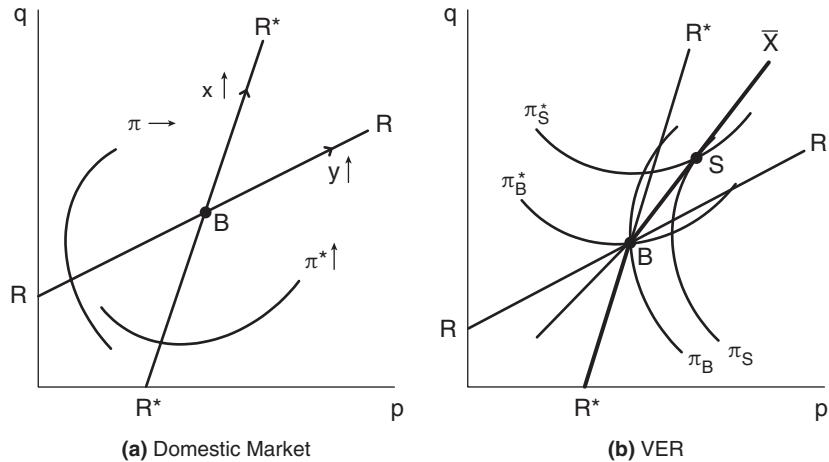


Figure 9.4: Bertrand Duopoly

In addition, under weak additional conditions the demand for each good will increase along its own reaction curve, as also illustrated in figure 9.4(a).⁷

The Bertrand equilibrium under free trade occurs at point B . Suppose then that the foreign firm faces a “voluntary” export restraint (VER) on sales to the domestic market, of the amount $x(p, q) \leq \bar{x}$. To make the argument as clear as possible, suppose further that the quota limit \bar{x} is set at the free trade level of exports, obtained at point B . Then what will be the response of the domestic and foreign firms to this export restraint? To answer this question, let us graph the prices at which the constraint $x(p, q) \leq \bar{x}$ holds, as done in figure 9.4(b). This constraint holds by choice of \bar{x} at the Bertrand equilibrium. If we moved up the foreign reaction curve R^*R^* , then export sales x increase, so the constraint is violated. Accordingly, the constraint will hold at *higher* values of p , as illustrated by the line $B\bar{X}$ in figure 9.4(b).

The fact that the foreign firm is restricted to sell $x(p, q) \leq \bar{x}$ is common knowledge among all firms. In this situation, it is no longer appropriate to suppose that the firms engage in Bertrand competition, treating the prices of the other firm as fixed. In particular, the domestic firm will realize that if it *raises* its price and shifts export sales toward the foreign firm, thereby leading to $x(p, q) > \bar{x}$, the foreign firm will have no choice but to increase its own price to restore $x(p, q) = \bar{x}$. The reaction curve of the foreign firm is therefore $R^*B\bar{X}$. We can expect the domestic firm to take advantage of the export restraint and increase its prices, knowing that the foreign firm will have to respond. Essentially, the VER gives the domestic firm a “first mover advantage” in selecting its price, so it can act like a Stackelberg leader relative to the reaction curve of the foreign firm.

By this reasoning, the optimal action by the domestic firm will be to increase its price until an iso-profit curve is tangent to the constraint $B\bar{X}$, as shown at point S in figure 9.4(b). This movement clearly increases the profits of the home firm, from π_B to π_S . Furthermore, the discrete increase in prices from point B to point S is much different from what we would expect from a tariff: since imports are equal at points

⁷See problem 8.7.

B and *S*, a zero tariff would have the same impact on imports as the export restraint, but obviously no impact at all on prices. So as in the analysis of Bhagwati (1965), the quota has a much greater impact on import and domestic prices than would a tariff at the same level of imports.

What is new in the duopoly model is the realization that the *foreign* firm also benefits from the export restraint. This is clear from figure 9.4(b) when exports are limited to the free-trade level: profits of the foreign firm increase from π_B^* to π_S^* . Harris (1985) refers to this as a reason why export restraints are “voluntary,” and Krishna (1989) refers to the export restraint as a “facilitating practice”; that is, it facilitates a more collusive outcome between the firms. Notice, however, that if we consider a quota at *less than* the free-trade level of exports, the profits of the foreign firm do not necessarily go up.⁸ In any case, the losers from the export restraint are home consumers who are paying a higher price for the domestic good and imports, without the benefits of tariff revenue. This begs the question of why a government would ever use a VER rather than a tariff. One answer is that the VER may prevent retaliation from the foreign government, since its own firms may be gaining from the trade restriction.

NONEQUIVALENCE DUE TO QUALITY CHOICE

We turn now to a second reason why tariffs and quotas may differ, and this is due to the choice of product quality for the exporting firms. We think of quality as a product characteristic that brings higher utility for consumers but also has higher costs for the firm. If an exporting firm is constrained in the *quantity* that it exports, how will this affect its choice of quality or product characteristics? Our argument in this section is that, under certain conditions, it will be optimal for the firm to increase product quality in response to the quota, whereas it would not do so with an ad valorem tariff.

To develop this argument, we will suppose that there are $i=1,\dots,N$ varieties of a differentiated product, each of which has characteristics given by the vector z_i . As in our treatment of monopolistic competition in chapter 5, we will suppose that the demand for these product varieties arises from an aggregate utility function, now given by

$$U(z_1 c_1, \dots, z_N c_N), \quad (9.1)$$

where c_i denotes the consumption of each variety, $i=1,\dots,N$, and z_i denotes the quality of each product, which then multiplies consumption. Several features of the aggregate utility function deserve comment.

First, if we recognize that the product varieties are actually bought by *different* individuals, is it still legitimate to use an aggregate utility function like (9.1)? This question has been answered in the affirmative by McFadden (1978, 1981) and Anderson, de Palma, and Thisse (1992). McFadden (1981) provides a very general aggregation theorem that applies to the behavior of individuals making choices over a discrete number of products, whereby they each purchase one unit of their optimal variety. It turns out that by summing over their quantity purchased of each product, their aggregate decisions are *equivalent* to those made by a representative consumer. We discuss

⁸See problem 9.2.

McFadden's aggregation theorem and extensions in appendix B, which allow us to use an aggregate utility function like (9.1).

Specific functional forms can also be obtained. One case that we shall consider is the CES aggregate utility function

$$U = \sum_{i=1}^N (z_i c_i)^{(\sigma-1)/\sigma}, \quad (9.2)$$

again defined over the consumption c_i of each good times its quality z_i . The fact that the CES utility function can arise from a discrete choice problem by individuals is demonstrated by Anderson, de Palma, and Thisse (1989; 1992, 85–90), provided that individuals consume continuous amounts of the differentiated good. A more conventional assumption is that they must consume zero or one unit, and in that case, the aggregate demands obtained are the “logit” formulation. We shall use the alternative demand functions at the end of the chapter, and an application of the logit demand in the context of product quality by Fajgelbaum, Grossman, and Helpman (2011a,b).

Given that an aggregate utility function exists, however, it is still rather special to suppose that the product characteristics and consumption enter in the *separable* form shown in (9.1), that is, in the form $z_i c_i$, which we can interpret as “quality times quantity.” This functional form implies that every unit consumed of good i is evaluated as having the same quality. This form of the utility function was used by Swan (1970), but criticized as being too special by Spence (1975). Spence uses a more general formulation like $f(z_p, c_i)$, under which the marginal utility of characteristics per unit consumed, $f_z(z_p, c_i)/c_i$, can vary with the level of consumption c_i . If we interpret demand as coming from different individuals, then $f_z(z_p, c_i)/c_i$ depending on c_i means that consumers have differing valuations of characteristics. We will use the special form $z_i c_i$ for convenience, but as we shall repeatedly emphasize, the results that we obtain on quality upgrading will not necessarily carry over to more general functional forms.

Free Trade Equilibrium

We begin by analyzing the problem for consumers and firms in the absence of any quota restriction. Consumers are presented with a set of $i=1, \dots, N$ varieties, with fixed characteristics z_i and prices p_i , and then choose the optimal quantity of each variety. It will be convenient to work with the “quality adjusted” prices, which are defined by $q_i \equiv p_i/z_i$. That is, the higher the overall product quality z_p , the lower the quality-adjusted prices q_i . The aggregate consumer maximizes utility in (9.1), subject to the budget constraint $\sum_{i=1}^N p_i c_i \leq I$. The Lagrangian for this problem is

$$\begin{aligned} L &= U(z_1 c_1, \dots, z_N c_N) + \lambda \left(I - \sum_{i=1}^N p_i c_i \right) \\ &= U(d_1, \dots, d_N) + \lambda \left(I - \sum_{i=1}^N q_i d_i \right), \end{aligned} \quad (9.3)$$

where the second line of (9.3) follows by defining $d_i \equiv z_i c_i$ as the effective “quality-adjusted” demand, and also by using the quality-adjusted prices $q_i \equiv p_i/z_i$. This rewriting of the Lagrangian makes it clear that instead of choosing c_i given prices p_i and quality z_p , we can instead think of the aggregate consumer as choosing d_i given quality-adjusted prices q_i , $i=1, \dots, N$. Let us denote the solution to problem (9.3) by $d_i(q, I)$, where q is the vector of quality-adjusted prices.

Varieties of the differentiated product can be sold by domestic and foreign firms, which we do not distinguish in the notation. Producing one unit of product i with quality z_i requires unit-costs of $g_i(z_i)$, where we are treating factor prices as constant and suppress them in this cost function. For simplicity, we are assuming constant returns to scale, so that the costs $g_i(z_i)$ do not depend on the level of output, but we do assume that $g_i(z_i)$ is convex in z_i . We will allow a firm to produce multiple varieties of a product, and suppose that the firm in question simultaneously chooses price and quality for the first $i = 1, \dots, M$ varieties. Furthermore, suppose that there is a specific transport cost of T_i to ship each unit to its destination market. Denoted the f.o.b. (free on board) price that the firms receives by p_i^* , then the c.i.f. (cost, insurance, freight) price that the consumer pays is $p_i = (p_i^* + T_i)$ and the quality-adjusted price is $q_i = p_i/z_i$.

Since consumption of each of the varieties is $c_i = d_i(q, I)/z_i$, the profit-maximization problem for this firm is then

$$\begin{aligned} \max_{p_i^*, z_i} \sum_{i=1}^M [p_i^* - g_i(z_i)] c_i &= \max_{p_i, z_i} \sum_{i=1}^M [p_i - g_i(z_i) - T_i] c_i \\ &= \max_{q_i, z_i} \sum_{i=1}^M \left[q_i - \frac{g_i(z_i) + T_i}{z_i} \right] d_i(q, I). \end{aligned} \quad (9.4)$$

In the first equality we add and subtract the specific transport costs T_i and use the c.i.f. prices $p_i = (p_i^* + T_i)$, and in the second equality we have changed the variables from nominal prices p_i to quality-adjusted prices q_i . This transformation relies on our assumption that *prices and characteristics are chosen simultaneously*.

It is immediately clear that to maximize profits in (9.4), the firms must minimize $[g_i(z_i) + T_i]/z_i$ for each variety, which is interpreted as the average costs per unit of quality (inclusive of transport costs) for good i . Taking logs and minimizing this over the choice of characteristics z_i leads to the first-order condition

$$\frac{g_i(z_i) + T_i}{z_i} = \frac{\partial g_i}{\partial z_i}, \quad i = 1, \dots, M. \quad (9.5)$$

Thus, we obtain equality between the average and marginal costs of quality. Notice that this condition holds despite the fact that the firm will be choosing its prices in a monopolistically competitive fashion, provided that price and quality are chosen simultaneously.⁹

Let us defer the optimal choice of prices for a moment to study the choice of quality arising from this first-order condition. Feenstra and Romalis (2014) adopt a specific form for the production function for quality, under which a firm with productivity φ_i that devotes labor of L_i per unit of output will obtain the quality of

$$z_i = (L_i \varphi_i)^\theta, \quad 0 < \theta < 1,$$

for all units of product i . For lower values of the parameter θ , the firm experiences greater diminishing returns in increasing quality. The corresponding cost per unit of output is

$$g_i(z_i) = w_i L_i = w_i z_i^{1/\theta} / \varphi_i. \quad (9.6)$$

⁹This first-order condition is also obtained by Rodriguez (1979) in a perfectly competitive model.

Using this cost function in the first-order condition (9.5), we quickly solve for log quality

$$\ln z_i = \theta[\ln T_i - \ln(w_i/\varphi_i) + \ln(\theta/(1-\theta))]. \quad (9.7)$$

Thus, quality is increasing in the productivity of the firm φ_i , decreasing in its wage, and most importantly, *increasing* in the specific transport costs T_i ; the firm will optimally choose *higher quality* for products that are being sent to more distant markets. This result is an example of the “Washington apples” effect, named after an example given by Alchian and Allen (1964, 74–75).¹⁰ In this example, it is noted that while the state of Washington grows apples of many varieties, the *best apples* were shipped the farthest distance, to East Coast markets. This is explained by treating the transport costs as the same for each apple. If there are two apples with prices $p_1 > p_2$, whereby the first apple is of the higher quality, then adding a specific transport cost of T will result in $(p_1 + T)/(p_2 + T) < p_1/p_2$, and therefore the *relative* price of the higher-quality apple will fall. Accordingly, we would expect the relative demand for the higher-quality apple to be greater in a more distant market. The result in (9.7) is a more sophisticated version of the Washington apples effect, with firms endogenously raising the quality that they ship to more distant markets.¹¹

Pricing by the Multiproduct firm

The multiproduct firm we are analyzing could be selling in different export markets, in which case its price in one market would not affect its demand in another. But for the sake of argument, let us suppose that the sales occur in the same market. In that case, charging a lower price for one product will pull demand away from another product that the firm is selling, in other words, it will “cannibalize” its own sales by charging a lower price. As a result, multiproduct firms in general charge higher prices than do single-product firms. We will give a simple illustration of this “cannibalization” effect in this section.

Maximizing (9.4) over the choice of prices q_i , we obtain the first-order conditions

$$d_i(q, I) + \sum_{j=1}^M \left[q_j - \frac{g_j(z_j) + T_j}{z_j} \right] \frac{\partial d_j}{\partial q_i} = 0, \quad i = 1, \dots, M. \quad (9.8)$$

This expression is more complicated than for a single-product firm because the firm is selling all the products $j = 1, \dots, M$, and therefore must take into account the effect of a change in its price q_i on all these products. To simplify this expression, let us consider the case of a CES utility function in (9.2). Then it can be confirmed that the demand derivatives are symmetric, $\partial d_i / \partial q_j = \partial d_j / \partial q_i$.¹² Using this in (9.8) and dividing by effective demand d_i , we can reexpress it as

$$1 + \sum_{j=1}^M \left[1 - \frac{g_j(z_j) + T_j}{p_j} \right] \frac{\partial \ln d_i}{\partial \ln q_j} = 0, \quad i = 1, \dots, M. \quad (9.9)$$

¹⁰Actually, the example given by Alchian and Allen (1964, 74–75) appears to refer in different editions of that textbook to the highest quality *California grapes or oranges* being sent to more distant markets. It was Borcherding and Silberberg (1978) who popularized the example of Washington apples.

¹¹Hummels and Skiba (2004) argue that this effect can be observed in trade data across countries.

¹²See problem 9.3 to derive this and also expression (9.11) below.

Let us denote the ratio of price to marginal cost for each product by $\mu_j = p_j / [g_j(z_j) + T_j] \geq 1$. We can see that the expression in brackets in (9.9) equals $(\mu_j - 1)/\mu_j = [p_j - g_j(z_j) - T_j]/p_j \geq 0$, which is the difference between price and marginal cost measured relative to price. This is the “Lerner index” of monopoly power for a single-product firm, and with price chosen optimally will equal the inverse of its elasticity of demand. To see how this Lerner pricing rule is modified with multiproduct firms, let us conjecture a solution where the price-cost ratios are *constant* across all products sold by the firm in question, $\mu_j = \mu$. Then it is immediately clear that the solution to (9.9) is

$$\left(\frac{\mu - 1}{\mu} \right) = - \left(\sum_{j=1}^M \frac{\partial \ln d_j}{\partial \ln q_j} \right)^{-1}. \quad (9.10)$$

Expression (9.10) says that the Lerner index for the firm equals the inverse of the *sum of* demand elasticities (rather than just the inverse of the own demand elasticity, as occurs for a single-product firm). In order for this solution to be valid, however, we need to have that the sum of elasticities on the right of (9.10) be *independent of good i*, because we have assumed that the markup is common across goods. This independence holds if an equiproportional increase in all prices charged by a firm needs to lead to the same percentage drop in demand for *any* product sold by that firm. It turns out that this condition is satisfied for CES demands, in which case the sum of elasticities is

$$-\left(\sum_{j=1}^M \frac{\partial \ln d_j}{\partial \ln q_j} \right) = \sigma + (1 - \sigma) \left(\sum_{j=1}^M \frac{p_j c_j}{I} \right), \quad (9.11)$$

where $\sigma > 1$ is the elasticity of substitution between products. The expression on the right-hand side of (9.12) is the sum of sales over all products sold by the firm, measured relative to total expenditure I on the differentiated product. In other words, the expression on the right is the *total market share* of the firm. As the market share grows, the sum of elasticities falls, and the markup of the firm will rise in (9.10).

Using the solution for μ from (9.10) and (9.11), the optimal prices for the firm are

$$p_i = \mu g_i(z_i), \quad i = 1, \dots, M. \quad (9.12)$$

Thus, in this CES case the firm charges the same markups over all products that it sells. This is certainly special and would not carry over to other functional forms. Bresnahan (1981) was the first to estimate a discrete choice model for autos that allowed for differing markups over product sold, and found that higher-priced models tended to have higher percentage markups. Similarly Berry, Levinsohn, and Pakes (1995, 1999), Feenstra and Levinsohn (1995), and Goldberg (1995) use functional forms more general than CES that allow markups to differ across products. These discussions go beyond the scope of this chapter, though the theory behind them is discussed in appendix B, and we shall comment on their empirical findings below.

While our solution for prices in (9.12) is special because the markup is constant across product varieties, it is still general enough to incorporate the cannibalization effect, meaning that the markup in (9.10)–(9.11) becomes greater as the market share of the multiproduct firm rises. The analysis of multiproduct firms in a monopolistic competition model is an ongoing topic of research.¹³ Working initially with homogeneous

¹³The papers we describe here all assume CES preferences. Alternative models using quadratic preferences are explored by Eckel and Neary (2009), Ottaviano and Thisse (2011), and Dhingra (2013).

firms, Feenstra and Ma (2008) develop the equilibrium conditions with CES preferences and firms choosing their optimal range of product varieties. They find that the integration of two markets—such as that due to the opening of trade—leads to an expansion in the range of varieties produced by each firm, and to an exit within each country, but with a larger number of firms in total supplying to consumers. In other words, even with CES preferences as in Krugman (1980), the model has similar features of firm exit within each country and the expansion of varieties available to consumers, as in Krugman (1979).

Feenstra and Ma then examine the equilibrium with heterogeneous firms, and find that it is technically difficult: even with a continuum of products, there will be a discrete number of potentially large firms in the market, so that precise zero profits for the borderline firm cannot be assured. Still, they are able to show that the relationship between product variety within a firm and productivity is non-monotonic: as small firms become more productive they will add product varieties, but at some point the new varieties “cannibalize” too much in sales, and so the largest firms reduce their range of product variety. This hypothesis of a U-shaped relationship between productivity and the range of varieties is confirmed empirically by Raff and Wagner (2013).

Rather than assuming a CES function over all varieties, it is convenient to assume instead a nested CES function as in the previous chapter, (8.35), where the varieties in each nest are produced by a single firm. With an elasticity of substitution $\sigma > 1$ between the varieties sold by a firm within a nest, and the elasticity ε with $1 < \varepsilon < \sigma$ between one nest and another, the elasticity of demand is as shown by (8.36). But because each firm supplies all the varieties within a nest, then its share s_{in} within the nest is unity and we see that the elasticity in (8.36) is simply $\varepsilon > 1$. In other words, the multiproduct firm will charge the fixed markup $\varepsilon/(\varepsilon - 1) > \sigma/(\sigma - 1)$, which reflects the cannibalization effect in a very simple way. This approach is taken by Allanson and Montagna (2005), Arkolakis, Ganapati, and Muendler (2014), and Ma (2009).¹⁴ Bernard, Redding, and Schott (2011) also assume a nested CES structure, but in contrast to these other authors, suppose that each firm produces only one product in each nest, thereby eliminating any effect since firms do not compete with their own products.

The presence of higher prices for multiproduct firms is tested by Hottman, Redding, and Weinstein (2014). Using barcode data on a large number of products and firms, they find that constant markups as with a CES model work quite well for many firms. But the largest multiproduct firms that account for a good deal of sale have much higher markups. They find that cannibalization is an important source of the higher markups for these firms.

Voluntary Export Restraint

Suppose now that the exports of a multiproduct foreign firm are restricted to the amount $\sum_{i=1}^M c_i = \sum_{i=1}^M (d_i/z_i) \leq \bar{X}$, where \bar{X} is the quota amount. Under the VER on Japanese autos, for example, the amounts \bar{X} were allocated to each Japanese firm by the Ministry of International Trade and Industry (MITI). We do not introduce a subscript to denote each foreign firm, but analyze the profit maximization problem for any such firm. The Lagrangian is

¹⁴Allanson and Montagna (2005) incorrectly state that the firm will use the markup $\sigma/(\sigma - 1)$, corresponding to the firm-level nest, while Arkolakis, Ganapati, and Muendler (2014) do not distinguish the elasticities of the firm-level and cross-firm nests in their main analysis. Raff and Schmitt (2016) model cannibalization as in the non-nested CES structure of Feenstra and Ma (2008).

$$\begin{aligned} L &= \max_{q_i, z_i} \sum_{i=1}^M \left[q_i - \frac{g_i(z_i) + T_i}{z_i} \right] d_i(q_i, I) + \lambda \left(\bar{X} - \sum_{i=1}^M \frac{d_i(q_i, I)}{z_i} \right) \\ &= \max_{q_i, z_i} \sum_{i=1}^M \left[q_i - \frac{g_i(z_i) + T_i + \lambda}{z_i} \right] d_i(q_i, I) + \lambda \bar{X}, \end{aligned}$$

where the second line follows by simply combining terms. Thus, to maximize profits, it is now the case that the foreign firms will choose the characteristic z_i to minimize $[g_i(z_i) + T_i + \lambda]/z_i$. In other words, the first-order condition (9.5) continues to hold by simply adding the Lagrange multiplier λ to the specific transport costs for each product:

$$\frac{g_i(z_i) + T_i + \lambda}{z_i} = \frac{\partial g_i}{\partial z_i}, \quad i = 1, \dots, M. \quad (9.5')$$

Higher values of the Lagrange multiplier λ will be associated with higher levels of quality chosen by the firm. To see this, totally differentiate (9.5') with respect to z_i and λ to readily obtain

$$\frac{dz_i}{d\lambda} = \left(\frac{\partial^2 g_i}{\partial z_i^2} \right)^{-1} > 0,$$

where the sign is obtained from our assumption that $g_i(z_i)$ is convex in z_i . The fact that quality rises with the quota is also apparent when we adopt the specific form for the cost function in (9.6), leading to the solution for quality in (9.7), rewritten here as

$$\ln z_i = \theta [\ln(T_i + \lambda) - \ln(w_i/\varphi_i) + \ln(\theta/(1-\theta))]. \quad (9.7')$$

Again, we see that an increase in the Lagrange multiplier on the quota constraint will lead to an increase in quality, just like an increase in the specific transport cost in the “Washington apples” effect.¹⁵ Notice that this increase in quality would *not occur* under an ad valorem tariff τ . In that case, we denote the tariff-inclusive quality-adjusted price by $p_i = (1+\tau)(p_i^* + T_i)$, so that with $q_i \equiv p_i/z_i$ the profit-maximization problem (9.4) would be rewritten as

$$\max_{p_i, z_i} \sum_{i=1}^M [p_i^* - g_i(z_i)] c_i = \max_{q_i, z_i} \sum_{i=1}^M \left[\frac{q_i}{(1+\tau)} - \frac{g_i(z_i) + T_i}{z_i} \right] d_i(q_i, I), \quad (9.4')$$

Changes in the tariff have no effect on the optimal choice of quality, since (9.5) still holds.

¹⁵We stress again that this result depends on the special functional forms we have used. Krishna (1987) uses a more general utility function like $f(z_i, c_i)$ as in Spence (1975), and finds that the change in product quality is ambiguous. Similarly, Feenstra (1988a) uses marginal costs for the firm that depend on quantity, in which case the change in product quality is again ambiguous. An alternative formulation of the quality-choice problem, using an “address” specification, is in Das and Donnenfeld (1987, 1989). They also find that the impact of the quota on quality is ambiguous in general (1989) but leads to quality upgrading under specific functional forms (1987).

MEASURING THE QUALITY OF TRADED GOODS

We turn now to two applications of the theory outlined above. The first application is from Feenstra and Romalis (2014), and deals with the measurement of import and export quality when we do not observe the characteristics of products. The second application focuses on the VER imposed on Japanese automobile imports by the United States in the 1980s, in which case we have abundant data on the characteristics of car models.

Measuring Quality without Characteristics

For most products, characteristics data will not be available and yet we might want an estimate of product quality so as to correct the nominal price p_i to obtain a quality-adjusted price q_i . This is the goal of a number of studies, including Hallak (2006), Hallak and Schott (2011), Johnson (2012), Khandelwal (2010), Martin and Méjean (2012), and Schott (2004, 2008). A recent study by Feenstra and Romalis (2014) aims to obtain quality-adjusted export and import prices for a very large set of countries, for incorporation into calculations of “real GDP” in the Penn World Table, and we shall describe their results. Feenstra and Romalis assume the cost function for quality shown in (9.6), $g_i(z_i) = w_i z_i^{1/\theta} / \varphi_i$, resulting in the quality choice in (9.7). Substituting (9.7) into (9.6) we readily obtain $g_i(z_i) = [\theta/(1-\theta)]T_i$. Thus, the marginal costs of production are proportional to the specific trade costs, as we shall use below.

We treat firms as selling to *separate* export markets, so there is no cannibalization effect in their choice of prices. We add the superscript k to denote the destination market. In the CES case, the optimal choice of the c.i.f. price for firm i selling to destination k is

$$(p_i^{*k} + T_i^k) = [g_i(z_i^k) + T_i^k] \left(\frac{\sigma}{\sigma - 1} \right).$$

Then using $g_i(z_{ij}^k) = [\theta/(1-\theta)]T_i^k$ from above, we solve for the f.o.b. and c.i.f. prices as

$$\begin{aligned} p_i^{*k} &= T_i^k \left[\left(\frac{1}{1-\theta} \right) \left(\frac{\sigma}{\sigma - 1} \right) - 1 \right], \text{ and} \\ p_i^k &= T_i^k \left[\left(\frac{1}{1-\theta} \right) \left(\frac{\sigma}{\sigma - 1} \right) \right]. \end{aligned}$$

Notice that these solutions for the f.o.b. and c.i.f. prices are independent of the productivity of the firm, φ_i , and also independent of wages. That result occurs because a firm with an increase in productivity (or reduction in wages) will have lower costs and therefore choose a higher quality, and that increase in quality just offsets the reduction in costs so that the nominal price charged by the firm is *unchanged*.¹⁶ While the cost function $g_i(z_i) = w_i z_i^{1/\theta} / \varphi_i$ used by Feenstra and Romalis (2014) is special because it leads to this “knife-edge” case, that is also a very convenient result because firm-level export prices are not available for many countries. Rather, Feenstra and Romalis interpret the prices shown above as the *average* export f.o.b. and import c.i.f. prices from

¹⁶For studies that empirically examines the relationship between the prices charged by firms and their quality, see Baldwin and Ito (2011), Crozet, Head, and Mayer (2012), Johnson (2012), and Mandel (2010).

each source country i to each destination country k , as measured by the corresponding unit-values in trade reported by the exporting and importing countries, respectively. The goal of their analysis is to correct these unit-values for quality.

To see how this correction is done, we first substitute the f.o.b. price above into the formula for quality in (9.7) to obtain

$$\ln z_i^k = \theta[\ln(\kappa p_i^{*k}) - \ln(w_i/\varphi_i)], \quad \kappa \equiv \frac{\theta(\sigma-1)}{1+\theta(\sigma-1)}. \quad (9.13)$$

We see that log quality for firm i selling to destination k is proportional to the natural log of the f.o.b. price, though it also depends on the wage of firm i adjusted for its productivity. This result will allow us to use the f.o.b. price to substitute for quality in an import demand equation. Consistent with our earlier finding that the price set by the firm does not depend on its productivity, let us focus on a homogeneous firm model of monopolistic competition, so that φ_i is the productivity of *every* firm in country i and w_i is the country i wage.¹⁷

The CES import-demand equation in country k is identical to what we have used in earlier chapters, but now expressed using the quality-adjusted prices

$$\frac{X_i^k}{X^k} = N_i \left(\frac{p_i^k/z_i^k}{P^k} \right)^{1-\sigma} \text{ with } P^k = \left(\sum_{j=1}^C N_j (p_j^k/z_j^k)^{1-\sigma} \right)^{1/(1-\sigma)}, \quad (9.14)$$

where X_i^k is the value of exports from country i to k , and N_i is the product variety of exports from country i to every destination. Inverting this demand equation, we obtain the quality-adjusted price

$$q_i^k \equiv \left(\frac{p_i^k}{z_i^k} \right) = P^k \left(\frac{X_i^k/N_i}{X^k} \right)^{1/(1-\sigma)}. \quad (9.15)$$

Noting that X_i^k/N_i equals the value of exports *per firm*, this equation represents a very simple “demand side” solution for quality: the higher the exports per firm, holding fixed the relative price p_i^k/P^k , then the higher the product quality and the lower the quality-adjusted price. This is the intuition that Hallak and Schott (2011) and Khandelwal (2010) rely on to estimate product quality. What we can see, however, is that this procedure is not independent of the number of exporting firms N_i that appears in (9.15). Khandelwal uses country population, for example, to proxy for the number of exporters, but that choice certainly influences his results.

So the demand side of the model alone cannot be relied on to solve for product quality, and the supply-side solution for quality in (9.13) also depends on additional data beyond the f.o.b. price: namely, on country wages and productivity. Since the goal of Feenstra and Romalis is to obtain quality-adjusted prices for a wide range of products (i.e., all SITC four-digit products, numbering over 700) and bilateral trade between all countries, they did not rely on a particular choice of data for these variables. Rather, they used the *zero profit* condition from monopolistic competition to

¹⁷Feenstra and Romalis (2014) show that the results to follow, and in particular the formula for the quality-adjusted export prices, do not depend on this assumption but continue to hold with heterogeneous firms that differ in their productivities.

pin down the productivity-adjusted wages, and to obtain an estimate of product quality that is independent of wages and independent of the number of exporters N_i .

To develop the zero-profit condition, suppose that the fixed costs of exporting from country i to k are $(w_i f_i^k / \varphi_i)$, that is, these costs are inversely related to country productivity.¹⁸ The value of exports per firm equals X_i^k / N_i , and in the CES case the profit earned by a firm is then $(X_i^k / N_i)^\sigma$. It follows that the zero-profit condition for exporters from country i to k is

$$\frac{(X_i^k / N_i)}{\sigma} = \frac{w_i f_i^k}{\varphi_i}.$$

If we combine this zero-profit condition with quality in (9.13), we obtain an alternative solution for the quality-adjusted price that comes from the supply side of the model

$$q_i^k \equiv \left(\frac{p_i^k}{z_i^k} \right) = \left(p_i^k / (\kappa p_i^{*k})^\theta \right) \left(\frac{X_i^k}{\sigma N_i f_i^k} \right)^\theta. \quad (9.16)$$

Notice that given the number of firms N_i , exports X_i^k are *positively* related to the quality-adjusted price, in contrast to the demand-side equation shown above in (9.15), where higher exports must indicate a lower quality-adjusted price. The positive supply-side relation occurs because when comparing exports from two countries to the same destination market, higher exports per firm are associated with higher productivity-adjusted wages from the above zero-profit condition, and hence, quality is lower in (9.13) and the quality-adjusted price is higher.

We can combine the demand and supply-side relations (9.15) and (9.16) to eliminate exports per firm and obtain a solution for the quality-adjusted export prices. Consider two countries i and j selling to the same market k . Then from (9.15) and (9.16) we obtain the ratio of their quality-adjusted export prices

$$\frac{q_i^k}{q_j^k} = \left(\frac{p_i^k / (p_i^{*k} f_i^k)^\theta}{p_j^k / (p_j^{*k} f_j^k)^\theta} \right)^{\frac{1}{1+\theta(\sigma-1)}}. \quad (9.17)$$

This solution for the relative export prices depends only on the c.i.f. and f.o.b. prices, as well as on the fixed costs of exporting and the model parameters θ and σ . To obtain these parameters, Feenstra and Romalis (2014) take the natural log of exports in (9.14) and use (9.13) to obtain

$$\ln X_i^k = \underbrace{[\ln N_i - (\sigma-1) \ln (w_i / \varphi_i)]}_{\beta_i \delta_i} + \underbrace{[\ln X^k + (\sigma-1) \ln P^k]}_{\beta^k \delta^k} - (\sigma-1) [\ln p_i^k - \theta \ln (\kappa p_i^{*k})].$$

This relation can be estimated as a gravity equation, including the source and destination fixed effects δ_i and δ^k and with *two* prices appearing on the right: the c.i.f. price appearing with coefficients $(1-\sigma) < 0$, and the f.o.b. price appearing with coefficient $\theta(\sigma-1) > 0$. The positive sign on the f.o.b. price reflects the fact that it is being used to substitute for quality. Feenstra and Romalis use an extension of the Feenstra (1994)

¹⁸Bilbiie, Ghironi, and Melitz (2012) also assume that fixed costs are inversely related to productivity, equal across firms. In addition, Feenstra and Romalis (2014) allow the fixed costs of exporting to depend on the real expenditure in the destination market, which follows from Arkolakis (2010). Feenstra and Romalis further rely on several measures of language similarity between any two countries to measure the fixed costs f_i^k .

technique (discussed in the appendix to chapter 5) to obtain the two coefficients σ and θ . They find a median estimate over some 700 SITC products of 6.07 for σ and 0.61 for θ , with only a handful of products violating the conditions that $\sigma > 1$ and $0 < \theta < 1$.

In figure 9.5 (top panel) we show an index of the unit-values for 185 countries in 2007, aggregated across all SITC products, followed by an index of quality estimates

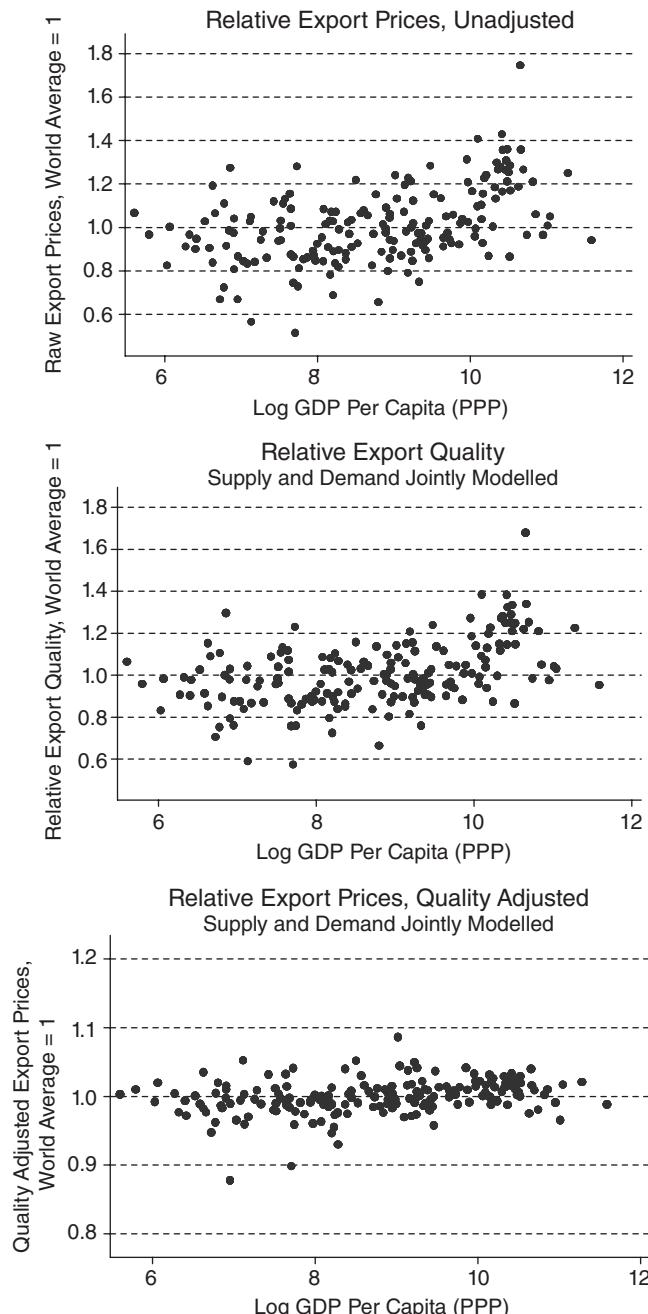


Figure 9.5: Export Prices, Export Quality, and Quality-Adjusted Export Prices, 2007

(second panel) and then an index of the quality-adjusted prices (third panel). All of the panels in figure 9.5 use per capita real GDP on the horizontal axis. Notice that richer countries tend to export more expensive goods (top panel), and these goods are estimated to be of higher-than-average quality (second panel). The quality-adjusted price (third panel), tends to be only slightly higher for richer countries, indicating that most of the higher export price for developed countries is explained by quality. Comparing the top panel with the second, it is visually apparent that the quality indexes are similar to the unit-value indexes, and as a result, the quality-adjusted prices (third panel) show much less variation than the raw unit-value. In other words, most of the differences in unit-values across countries are due to quality differences, and not due to “true” quality-adjusted price differences.

Feenstra and Romalis (2014) also obtain results for the indexes of import prices across countries, as shown in figure 9.6. They find that richer countries import more expensive items (top panel) that are of higher quality (second panel). But the quality-adjusted *import* prices still increase noticeably with the importing country’s GDP per capita (as shown in the third panel of figure 9.6), and more so than the quality-adjusted *export* prices (as shown in the third panel of figure 9.5). One reason this increase occurs in their model is that the poorer countries are often smaller in economic size, and so only more efficient firms are able to overcome the fixed costs of exporting to them. As a result, the quality of products exported to poorer countries does not decline with income as fast as the price does; therefore, the quality-adjusted import prices (as shown in the third panel of figure 9.6) still increase with income.

In figure 9.7, we divide the export price index for each country by its import prices index to obtain the terms of trade. In the top panel of figure 9.7 we show the terms of trade without any adjustment for quality. There is evidently no relationship between the terms of trade and country income. In the second panel, we divide quality-adjusted export prices by quality-adjusted import prices, and we obtain the quality-adjusted terms of trade that is *negatively* related to the level of income, at least weakly. In other words, the “true” terms of trade faced by countries is lower for rich countries than for poor countries. This rather surprising result is due in part to the lower unit-value of imports for poor countries, but it also relies on the supply side of the model: countries with lower imports (because they are poor or just small) buy from more efficient foreign firms that can overcome the fixed costs of exporting, and these firms sell higher quality. This result lends support to the proposition of Fajgelbaum, Grossman, and Helpman (2011a) that poorer countries are net importers of higher-quality goods, because these goods are not produced locally.

Measuring Quality in Automobiles

We turn now to a second application of product quality, and that is to the VER applied to U.S. automobile imports during the 1980s. With unemployment in the U.S. automobile industry growing during 1979–80 and profits falling, Ford Motor Company and the UAW applied to the ITC for Section 201 protection. As discussed in chapter 8, this application was denied on the grounds that the *U.S. recession* appeared to be a greater cause of unemployment than rising imports from Japan. Faced with this negative outcome, several congressmen from midwestern states pursued protection for automobiles via other means. Senators Danforth and Bentsen introduced a bill to restrict imports, which would have restricted Japanese automobiles imports into the United States to 1.6 million units during 1981, 1982, and 1983. Indeed, this bill was

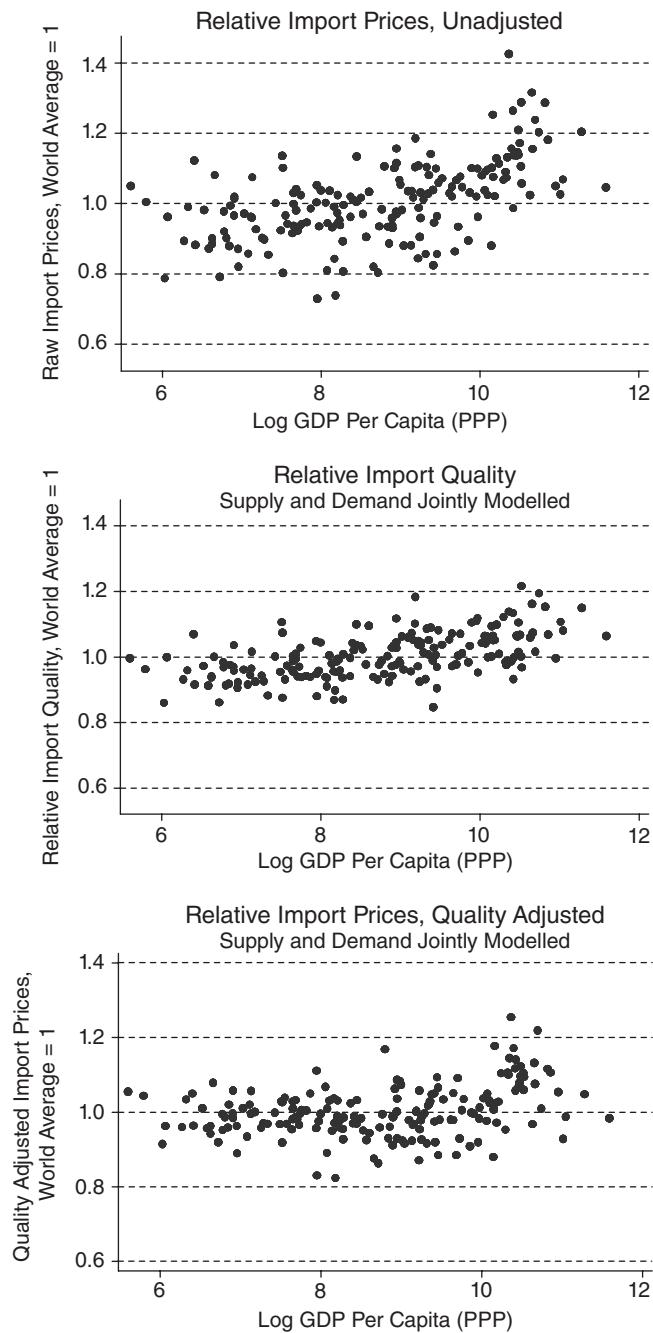


Figure 9.6: Import Prices, Import Quality and Quality-Adjusted Import Prices, 2007

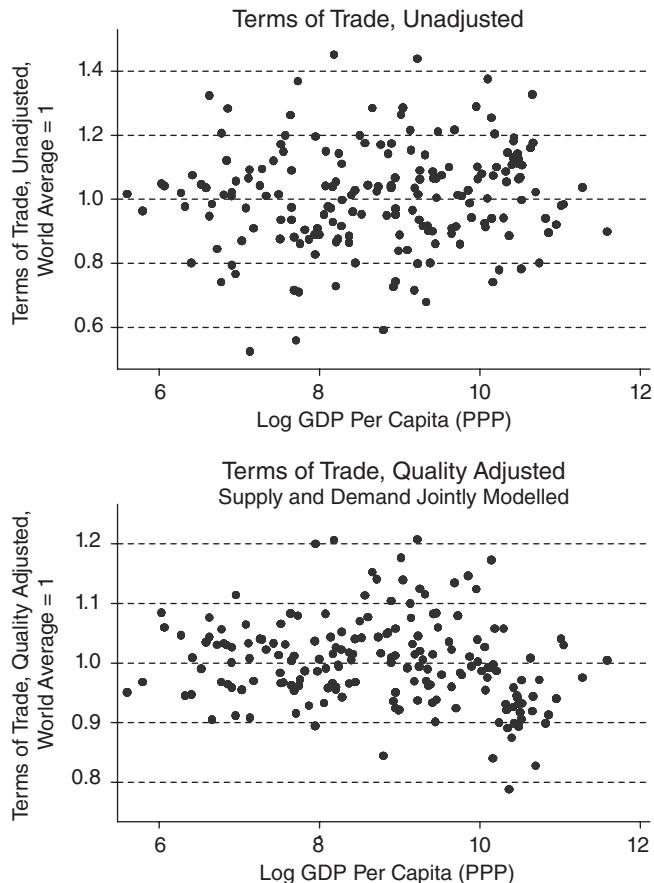


Figure 9.7: Terms of Trade: Unadjusted and Quality Adjusted, 2007

scheduled for revision in the Senate Finance Committee on May 12, 1981. Clearly aware of this pending legislation, the Japanese government announced on May 1 that it would “voluntarily” limit the sales of automobiles in the U.S. market.

For the period April 1981–March 1982, this limit was set at 1.83 million autos, including 1.68 million cars exported to the United States, 82,500 utility vehicles, and 70,000 vehicles exported to Puerto Rico. This limit was in place until March 1984, after which it was raised to 2.02 and then 2.51 million vehicles annually until March 1992, as shown in table 9.2, where we compare the actual imports with the VER limits. It is evident that the VER was nearly exactly satisfied in early years, and remained binding until 1987. After that, however, actual imports were below the VER, so that by March 1994 it was discontinued. The reason that imports fell below the VER after 1987 was that Japanese companies had begun assembling cars in the United States, so imports naturally fell.

Our interest, then, is in the effects of the VER during the early years. To estimate the change in product quality, initially consider the case where there is no VER. Assuming that the costs to produce each model take the form $\ln g_i(z_{it}) = \beta_t + \beta' z_{it} + \xi_{it}$, where z_{it} denotes a vector of characteristics, ξ_{it} is a random error that reflects unmeasured characteristics, and we add the year fixed-effect β_t to denote any other factors

TABLE 9.2: Japanese Car Imports and VER limit, 1979–1994

Time period	U.S. Imports from Japan	VER limit	Difference (Imports – VER)
April 1979–March 1980	1,779,497	none	
April 1980–March 1981	2,011,822	none	
April 1981–March 1982	1,833,313	1,832,500 ^a	813
April 1981–March 1983	1,831,198	1,832,500	-1,302
April 1983–March 1984	1,851,694	1,832,500	19,194
April 1994–March 1985	2,031,250	2,016,000 ^b	15,250
April 1985–March 1986	2,605,407	2,506,000 ^c	99,407
April 1986–March 1987	2,518,707	2,506,000	12,707
April 1987–March 1988	2,377,383	2,506,000	-128,617
April 1988–March 1989	2,115,304	2,506,000	-390,696
April 1989–March 1990	2,015,920	2,506,000	-490,080
April 1990–March 1991	1,911,828	2,506,000	-594,172
April 1991–March 1992	1,728,446	2,506,000	-777,554
April 1992–March 1993	1,637,519	1,650,000	-12,481
April 1993–March 1994	1,549,587	1,650,000	-100,413

Source: Table 2 from “The U.S. Automobile Industry: Monthly Report on Selected Economic Indicators,” U.S. International Trade Commission, May issue of each year, cited in Berry, Levinsohn, and Pakes (1999, 402).

Notes: This table reports U.S. imports of cars and sport-utility vehicles from Japan, including imports to Puerto Rico.

a. Computed as 1.68 million cars to the United States, 82,500 utility vehicles to the United States, and 70,000 vehicles to Puerto Rico.

b. Of the 2.106 million total, 1.85 million were cars sold to the United States.

c. Of the 2.506 million total, 2.30 million were cars sold to the United States.

(such as wages) that affect costs equi-proportionately across models. Adding the time subscript, optimal prices from (9.12) are $p_{it} = \mu_t g_i(z_{it})$, where for convenience here we ignore the specific transport costs of export (these will later be incorporated along with the VER). Prices are then written in logs as

$$\ln p_{it} = \alpha_t + \beta' z_{it} + \xi_{it}, \quad i=1, \dots, M, \quad (9.18)$$

where $\alpha_t = \beta_t + \ln \mu_t$ reflects year-specific factors in costs and the markups. We should also be indexing α_t by firms (since the costs and markups differ across firms), but omit this notation. Equation (9.18) is a log-linear regression of prices on characteristics and is called a “hedonic regression.” Our derivation shows that this is a valid representation of the equilibrium under monopolistic competition provided that markups are the same across products (as in the CES case).

Under the VER, optimal prices now equals $p_{it} = \mu_t [g_i(z_{it}) + \lambda_t]$, where λ_t is the Lagrange multiplier on the VER constraint. For the same specification of marginal costs, $\ln g_i(z_{it}) = \beta_t + \beta' z_{it} + \xi_{it}$, prices are rewritten as

$$p_{it} = \exp(\alpha_t + \beta' z_{it} + \xi_{it}) + s_t, \quad i=1, \dots, M, \quad (9.19)$$

where $\alpha_t = \beta_t + \ln \mu_t$ as before and $s_t = \mu_t \lambda_t$ are the specific price increases due to the VER.

To convert the random error to an additive form, define

$$\begin{aligned}\varepsilon_{it} &= \exp(\alpha_t + \beta' z_{it} + \xi_{it}) - \exp(\alpha_t + \beta' z_{it}) \\ &= [\exp(\xi_{it}) - 1] \exp(\alpha_t + \beta' z_{it}).\end{aligned}$$

Then (9.19) is rewritten as

$$p_{it} = \exp(\alpha_t + \beta' z_{it}) + s_t + \varepsilon_{it}, \quad i=1, \dots, M. \quad (9.20)$$

Thus, under the VER we need to allow for *specific* price increases s_t from year to year, as well as the *proportional* price changes captured by α_t .

Feenstra (1988a) estimates these hedonic regressions using a sample of U.S. imports of Japanese cars and trucks. Only cars and utility vehicles were subject to the VER, whereas trucks had a higher tariff imposed in August 1980, as discussed in the previous chapter. Summary statistics for the sample over 1979–85 are shown in table 9.3. Notice that the unit-value (i.e., average price) of Japanese cars increased by 20 percent from 1980 to 1981, which was unusually high, and a further 10 percent from 1981 to 1982. We should not attribute this entire amount to the VER, however. The nominal increase in prices can also reflect general inflation as well as increases in the quality of each product. Using the estimates for a hedonic regression, we can define quality as $(\alpha_t + \beta' z_{it})$, which will increase whenever the characteristics z_{it} rise. Observing whether this occurred for Japanese autos under the VER, as predicted by the theory, is the goal of the empirical work, as well as to estimate the impact of the VER on prices as measured by the specific price terms s_t .

An initial estimate of (9.20) for 1979–85, using the base version of each Japanese car model sold in the United States, is shown in column (1) of table 9.4.¹⁹ The coefficients at the top of the table are estimates of β_k for each characteristic k . For example, an increase in width of one foot is estimated to raise the price of a car by $\exp(0.36) = 1.43$, while an increase in horsepower of 100 will increase price by $\exp(0.69) = 2$, or double the price. In the lower part of the table we report the estimates of α_t , referred to as the *proportional* price increases (measured relative to 1979), and then the estimates of s_t , referred to as the *specific* price increases (measured relative to 1980). Except for 1985, we see that the specific price increases are positive but have high standard errors. This is due to the multicollinearity between the terms α_t and s_t , both of which are measuring annual price increases.

To address this multicollinearity, we consider pooling data for U.S. imports of Japanese cars and trucks. Denoting trucks with an asterisk, we rewrite the hedonic regression (9.17) as

$$\ln p_{it}^* = \alpha_t^* + \beta^{**} z_{it}^* + \xi_{it}^*, \quad i=1, \dots, M. \quad (9.21)$$

Thus, we are allowing trucks to have different coefficients on their characteristics from those on cars, and also differing proportional price increases α_t^* . Because trucks were

¹⁹ Assuming that the random errors ξ_{it} have equal variance, then from (9.19) we see that the variance of ω_{it} is proportional to $\exp(\alpha_t + \beta' z_{it})^2$. Accordingly, we obtain preliminary estimates of α_t and β and then use the inverse of $\exp(\alpha_t + \beta' z_{it})^2$ as weights when estimating (9.20). See the empirical exercises to reproduce the results in tables 9.3 and 9.4.

TABLE 9.3: Japanese Cars and Trucks, and U.S. Small Cars, 1979–85

	1979	1980	1981	1982	1983	1984	1985
Japanese Cars							
Number of models	21	24	24	24	26	29	31
Unit-value (\$)	4,946	5,175	6,211	6,834	7,069	7,518	8,153
(percent change)		(4.6)	(20.0)	(10.0)	(3.4)	(6.4)	(8.4)
Price index	96.6	100	119.8	128.9	131.3	138.9	147.8
(percent change)		(3.5)	(19.8)	(7.6)	(1.9)	(5.8)	(6.4)
Unit-quality (\$)	4,361	4,473	4,867	5,253	5,637	5,862	6,130
(percent change)		(2.6)	(8.8)	(7.9)	(7.3)	(4.0)	(4.6)
Quality index	98.7	100	108.6	115.2	121.3	126.6	130.6
(percent change)		(1.3)	(8.6)	(8.6)	(5.3)	(4.4)	(3.2)
Japanese Trucks^a							
Number of models	10	10	11	11	10	11	12
Unit-value (\$)	4,804	4,937	6,134	6,426	6,134	6,247	6,250
(percent change)		(2.8)	(24.2)	(4.8)	(−4.5)	(1.8)	(0.05)
Price index	97.02	100	127.8	130.9	121	123.1	125.2
(percent change)		(3.1)	(27.8)	(2.4)	(−7.6)	(1.7)	(1.7)
Unit-quality (\$)	4,627	4,638	4,791	4,930	4,997	5,010	5,433
(percent change)		(0.2)	(3.3)	(2.9)	(1.4)	(0.3)	(8.4)
Quality index	99.6	100	103.5	105.9	105.7	105.3	116.0
(percent change)		(0.4)	(3.5)	(2.3)	(−0.2)	(−0.4)	(10.2)
U.S. small cars							
Number of Models	24	22	23	27	33	34	
Unit-value (\$)	4186	5067	5915	6446	6581	6781	
(percent change)		(21.0)	(16.7)	(9.0)	(2.1)	(3.0)	
Price Index	81.9	100.0	116.8	125.6	125.4	127.5	
(percent change)		(22.1)	(16.8)	(7.5)	(−0.2)	(1.7)	
Unit-Quality (\$)	4195	4132	4183	4351	4497	4563	
(percent change)		(−1.5)	(1.2)	(4.0)	(3.4)	(1.5)	
Quality Index	100.5	100.0	102.2	104.4	105.7	106.1	
(percent change)		(−0.5)	(2.2)	(2.1)	(1.3)	(0.4)	

Source: Revised from Feenstra (1985, 1988a); empirical exercises 8.1 and 8.4.

Notes: a. Includes utility vehicles.

not subject to the VER, we do not include any specific price increases in this regression. Estimates of (9.21) using the data on trucks are shown in regression (2) of table 9.4.²⁰

We now consider testing the hypothesis that the year-to-year price changes in Japanese cars and truck exports to the United States would be the same, *after correcting for the trade policies in each product*. Cars were subject to the VER after April 1981, whereas trucks imported from Japan were subject to an increase in the tariff from 4 to 25 percent after August 1980, as discussed in the previous chapter. Feenstra (1988a) argues that the 21 percent increase in the wholesale price of trucks as they cross the border

²⁰Utility vehicles are given the same hedonic coefficients β^* as trucks, but use the year coefficients s_t and α_t of cars, since utility vehicles were subject to the VER.

TABLE 9.4: Hedonic Regression for Japanese Cars and Trucks, 1979–85, with Prices as Dependent Variable

	(1) Cars	(2) Trucks	(3) Cars	(4) Trucks
Constant	6.33*	7.71*	5.83*	6.63*
	(0.58)	(0.50)	(0.86)	(0.77)
Weight (tons)	0.03	0.41*	0.05	0.48*
	(0.12)	(0.08)	(0.11)	(0.12)
Width (feet)	0.36*	0.01	0.39*	0.21
	(0.11)	(0.10)	(0.14)	(0.14)
Height (feet)	-0.06	0.01	-0.06*	-0.03
	(0.06)	(0.05)	(0.06)	(0.07)
Horsepower (100)	0.69*	0.20	0.81*	0.24
	(0.09)	(0.10)	(0.10)	(0.15)
Transmission	0.14*	0.03	0.18*	0.05
(5-speed or auto)	(0.02)	(0.03)	(0.04)	(0.04)
Power steering	0.06*	0.09	0.07*	0.06
	(0.03)	(0.04)	(0.03)	(0.06)
Air conditioning	0.15*		0.16*	
	(0.03)		(0.04)	
Four-wheel drive		0.22*		0.30*
		(0.03)		(0.06)
<hr/>				
Proportional:				
Year 1980	0.01	0.004	0.01	0.01
	(0.03)	(0.033)	(0.03)	(0.03)
Year 1981	0.05	0.21*	0.07	0.23*
	(0.11)	(0.03)	(0.04)	(0.04)
Year 1982	0.09	0.23*	0.07	0.23*
	(0.11)	(0.03)	(0.04)	(0.04)
Year 1983	0.02	0.17*	-0.001	0.16*
	(0.11)	(0.04)	(0.05)	(0.05)
Year 1984	0.09	0.19*	0.02	0.18*
	(0.10)	(0.04)	(0.05)	(0.05)
Year 1985	0.22*	0.18*	0.10	0.11
	(0.10)	(0.04)	(0.06)	(0.06)
<hr/>				
Specific:				
Year 1981	410		367	
	(619)		(239)	
Year 1982	378		600*	
	(650)		(254)	
Year 1983	770		887*	
	(601)		(243)	
Year 1984	624		1,104*	
	(618)		(251)	
Year 1985	-123		856	
	(719)		(366)	
<hr/>				
Observations	179	75	254	
R ²	0.990	0.996	0.992	

Source: Revised from Feenstra (1988a); empirical exercises 9.2 and 9.3.

Notes: Standard errors are in parentheses.

* Significant at 95 percent level.

would be reflected by at most a 16 percent increase in the retail price.²¹ Accordingly, we consider testing the hypothesis that the proportional year increase for trucks, α_t^* , is 0.16 greater than the proportional year increase for cars, α_t , for 1981 and later years:

$$\alpha_t^* = \alpha_t + 0.16 \text{ for } t = 1981, 82, \dots, 85. \quad (9.22)$$

Feenstra (1988a) finds that this hypothesis is accepted in all years except 1985, and so it is imposed on the subsequent estimation except for that year.

In columns (3) and (4) of table 9.4, we reestimate the regressions for cars and trucks, while imposing the cross-equation restraint (9.22) for $t = 1981, \dots, 84$. Notice that this reduces the standard errors on the specific price increases for cars substantially. By 1984–85, the impact of the VER is to increase the price of cars by about \$1,000 on average. This is an estimate of the pure price increase due to the quota, after correcting for changes in the characteristics of Japanese cars. Multiplying this by the average 1984–85 imports of some 2.2 million, we obtain an estimate of the quota rents of \$2.2 billion annually in 1984–85.

This estimate equals the lower end of the quota rents for autos shown in table 9.1. Where do the higher estimates come from? It turns out that along with the increase in Japanese car prices, European producers selling in the United States also raised their prices substantially—by nearly one-third, as estimated by Dinopoulos and Kreinin (1988). They calculate a U.S. welfare loss due to this increase in European prices of some \$3.4 billion in 1984, which exceeds the loss due to quota rents transferred to Japan. This nicely illustrates the result argued by Harris (1985) and Krishna (1989) that a VER will act as a “facilitating practice” and leads to more collusive behavior in the market where it is imposed. For American small cars we also show a substantial increase in prices at the bottom of table 9.3: by 21 percent from 1979–80 and 17 percent from 1980–81. Some of this increase occurred before the VER, but it appears that this quota allowed American producers to maintain and continue large price increases.

What about the change in product quality for Japanese imports? In fact, there was a considerable upgrading of model characteristics at just the time that the VER was imposed. In table 9.3, we show the “unit-quality,” which is computed as $(\alpha_0 + \beta' z_i)$ and then averaged across all models. The unit-quality increases by 8.8 percent from 1980 to 1981, and another 7.9 percent over 1981–82, and 7.3 percent over 1982–83. The dollar increase in quality over 1980–85 is \$1,650, which reflects higher values of the characteristics applied to each model. This vastly exceeds the quality change for trucks or U.S. small cars as shown in table 9.3. These observations strongly support the theoretical prediction that Japanese producers raised characteristics to increase the quality of cars due to the VER, but this did not occur under the ad valorem tariff in trucks.

Our estimate of the quota rents over 1980–85 can be compared to estimates in other studies. Ries (1993) studies the change in stock prices of Japanese auto producers and parts suppliers during the weeks in 1981 when this policy was being formulated. He finds strong evidence of positive returns to the stocks of Japanese auto producers during this period, though only after it became clear that MITI would administer the quotas to each producer (so that the producers would capture the rents). In contrast, only the largest parts suppliers obtain abnormal stock during this period.

²¹This can be thought of as a pass-through of 0.75 from the 21 percent increase in tariff to a 16 percent increase in retail price. This is somewhat higher than the pass-through of 0.58 for the truck tariff estimated in Feenstra (1989) and discussed in the previous chapter.

Further supportive evidence is obtained by Goldberg (1995), who finds that the VER was binding in 1983–84 and 1986–87, with the highest rents in the earlier years.

The only study to question this conclusion is Berry, Levinsohn, and Pakes (1999) (hereafter BLP 1999), who argue that the VER was *not binding* in the early years and became binding in 1986. They obtain the highest estimates for the price effect of the quota (like s_t in our notation) in the years 1987–90. This is a puzzling finding, since from table 9.2 we know that Japanese imports fell short of the VER limit during March 1987–April 1988 and in later years. For the remainder of this section, we discuss possible explanations for the differing findings of Feenstra (1988a) and BLP (1999).

We begin by extending our sample from 1979–85, as used in the regressions of table 9.4, to 1979–90. The new estimates of the car and truck regressions (9.20) and (9.21) are shown in table 9.5. In columns (1) and (2) we report the car and truck regressions without any cross-equation constraints, while in columns (3) and (4) we make use of the constraint in (9.22) for $t = 1981, \dots, 84$.²² As before, making use of this cross-equation constraint is effective in reducing the standard errors of the specific price increases.

Including the additional five years of data turns out to have a substantial impact on some of the hedonic coefficients: for example, the coefficient on width in the car regression is much higher than in table 9.4, while the coefficient on horsepower is smaller. With this change in the hedonic coefficients, the specific price effect of the VER in the *early years* is *reduced* somewhat: in 1984 we estimate a price increase due to the VER of \$812 in column (3) of table 9.5, as compared to \$1,104 in column (3) of table 9.4. This price impact is small in 1985 in both tables, but the effect of the VER grows after that and peaks in 1987, where we estimate quota rents of \$1,164 per vehicle in column (3) of table 9.5. So by extending the sample, we find that the VER is still binding in 1987. After that year, however, our estimates give highly imprecise estimates of the price impact of the VER, ranging from positive in column (1) to negative in column (3), but in all cases *insignificantly different from zero*. Consistent with the fact that Japanese exports to the United States fell short of the VER limit, we do not identify any significant impact of this trade restriction on prices after 1987.

In contrast, BLP (1999) continue to find a significant price impact of the VER through 1990, but no significant effect before 1986. I believe that the reason for this has to do with the way in which BLP model the supply side of the market. They model marginal costs for Japanese producers as determined by time trends and wages, with the exchange rate entering insignificantly. This formulation predicts that a steady increase in prices would have occurred over 1980–85 *even without* the VER, and so the VER does not appear to be binding in their estimates. But this predicted increase in prices is contradicted by the *actual* movement in truck prices, which rose due to the tariff in 1980–81 but were quite stable or fell for several years afterward (see table 9.3). The free-trade change in car prices that we have used the same as for trucks, correcting for the tariff as in (9.22). Therefore, we are using a *lower* increase in free-trade car prices than BLP, which accounts for differences in our findings through 1985.

We find, as do BLP (1999), that the VER is binding in 1986 and 1987, but after this time we begin to estimate a rather large *proportional* increase in car prices, and so the *specific* price effect of the VER becomes smaller and insignificant. In contrast, the free-trade increase in prices allowed by BLP is still dominated by the time trends and wages entering marginal costs, and it turns out that these are not enough to explain the

²²As in note 19, utility vehicles are given the same hedonic coefficients β^* as trucks, but using the year coefficients s_t and α_t of cars, since utility vehicles were subject to the VER.

TABLE 9.5: Hedonic Regression for Japanese Cars and Trucks, 1979–90, with Prices as Dependent Variable

	(1) Cars	(2) Trucks	(3) Cars	(4) Trucks
Constant	4.05*	6.96*	4.61*	6.58*
	(1.25)	(1.25)	(0.77)	(0.42)
Weight (tons)	0.48*	0.49*	0.37*	0.48*
	(0.20)	(0.09)	(0.14)	(0.09)
Width (feet)	0.82*	0.07	0.73*	0.16*
	(0.21)	(0.06)	(0.13)	(0.08)
Height (feet)	-0.27*	-0.03	-0.22*	-10.07
	(0.10)	(0.05)	(0.07)	(0.06)
Horsepower (100)	0.42*	0.36*	0.42*	0.35*
	(0.09)	(0.08)	(0.07)	(0.09)
Transmission	0.16*	0.09	0.14*	0.09*
(5-speed or auto)	(0.05)	(0.02)	(0.04)	(0.02)
Power steering	0.16*	0.01	0.15*	-0.002
	(0.05)	(0.06)	(0.03)	(0.05)
Air conditioning	0.15*	0.15*		
	(0.05)	(0.04)		
Four-wheel drive		0.09		0.30*
		(0.10)		(0.06)
Year Dummy:	Cars Proportional	Cars Specific	Trucks Proportional	Cars Proportional
Year 1980	-0.04 (0.07)	-0.01 (0.02)	-0.03 (0.03)	-0.03 (0.03)
Year 1981	0.06 (0.13)	256 (472)	0.31* (0.06)	0.16* (0.05)
Year 1982	0.09 (0.13)	360 (512)	0.31* (0.06)	0.17 (0.05)
Year 1983	0.05 (0.13)	510 (482)	0.19* (0.04)	0.06 (0.05)
Year 1984	0.09 (0.12)	695 (477)	0.20* (0.05)	0.07 (0.05)
Year 1985	0.23* (0.07)	60 (93)	0.05* (0.02)	0.22* (0.05)
Year 1986	0.28* (0.13)	672 (602)	0.21* (0.05)	0.26* (0.09)
Year 1987	0.36* (0.13)	1317* (583)	0.19* (0.05)	0.34* (0.09)
Year 1988	0.41* (0.14)	859 (1067)	0.05 (0.15)	0.43* (0.10)
Year 1989	0.40* (0.15)	852 (1134)	0.11 (0.16)	0.47* (0.10)
Year 1990	0.32* (0.14)	850 (1123)	0.21 (0.17)	0.42* (0.10)
Observations	404	404		
R ²	0.995	0.993		

Sources: Updated from Feenstra (1988a), as in empirical exercise 9.5.

Notes: Standard errors are in parentheses.

*Significant at 95 percent level.

actual increase in Japanese prices: hence, they estimate that the VER is binding after 1987, even though actual exports fell short of the VER limit.²³

EXPORT SUBSIDIES

We turn now to our second major topic for the chapter, which is the application of export subsidies. These are used by developing and industrial countries alike to support their industries. There is little doubt that these subsidies are welcomed by the firms receiving them, and this might explain their use in a political economy model, as we examine in the next chapter. But in this chapter we are concerned with social welfare, which means that the *revenue cost* of the subsidies must be counted against any increase in profits of the receiving industries: only if the revenue cost is *less than* the net change in producer plus consumer surplus are the subsidies in the national interest. As we shall see, this cannot occur in a two-good model with perfect competition, and after presenting these results we will consider other frameworks.

PERFECT COMPETITION

We consider first the case of a small country, in figure 9.8. In panel (a) we show the domestic demand D and supply curves S , together with the constant world price p^* . Under free trade, domestic demand is at c_0 and supply at y_0 , so exports are $x_0 = y_0 - c_0$. This is shown in panel (b), which graphs the domestic export supply curve $X = S - D$. We can think of the fixed world price p^* as establishing a horizontal import demand curve from the rest of the world, which intersects X at the equilibrium exports x_0 .

Now suppose that firms in the home industry are given a subsidy of s dollars per unit exported. With this subsidy, the home industry would be able to earn $p^* + s$ on all quantities that are exported. Accordingly, firms in this industry would be unwilling to sell at home for anything less than that amount, and the domestic price must also rise to $p^* + s$. At this price, domestic demand falls to c_1 and supply rises to y_1 , so exports rise to the amount $x_1 = y_1 - c_1$. This corresponds to a *rightward shift* of the domestic export supply curve in panel (b), because at the same international price of p^* , exports have increased from x_0 to x_1 . Equivalently, the export supply curve shifts downward, from X to X' , by the amount of the subsidy s .²⁴

The welfare effect of the subsidy can be readily calculated in panel (a) as $-(a + b)$ consumer surplus loss $+(a + b + c)$ producer surplus gain $-(b + c + d)$ subsidy cost $= -(b + d)$, which is always negative. Area $(b + d)$ is the deadweight loss of the subsidy, and it is also shown as the triangle under the export supply curve in panel (b). This area is analogous to the deadweight loss that we found for a tariff, in chapter 8. Therefore, the small country also loses due to an export subsidy.

Consider now the case of a large country, in figure 9.9. In panel (b), we show again the rightward shift of the domestic export supply curve, from X to X' . Because this

²³ Another explanation for the unusual pattern of VER price-effects estimated by BLP (1999), which show an impact of the VER after 1987 even though it was not binding, is that the estimation method they are using did not converge to the optimal coefficients. Knittel and Metaxoglou (2014) find that difficulties in convergence are a persistent problem with the BLP estimator.

²⁴ This downward shift of the export supply curve occurs because we are measuring the international price p^* on the vertical axis of panel (b), in both figures 9.6 and 9.7.

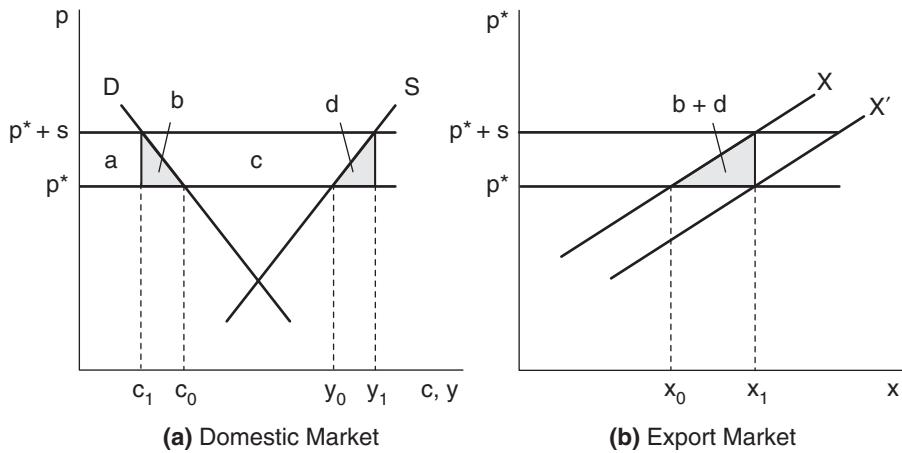


Figure 9.8: Small Country

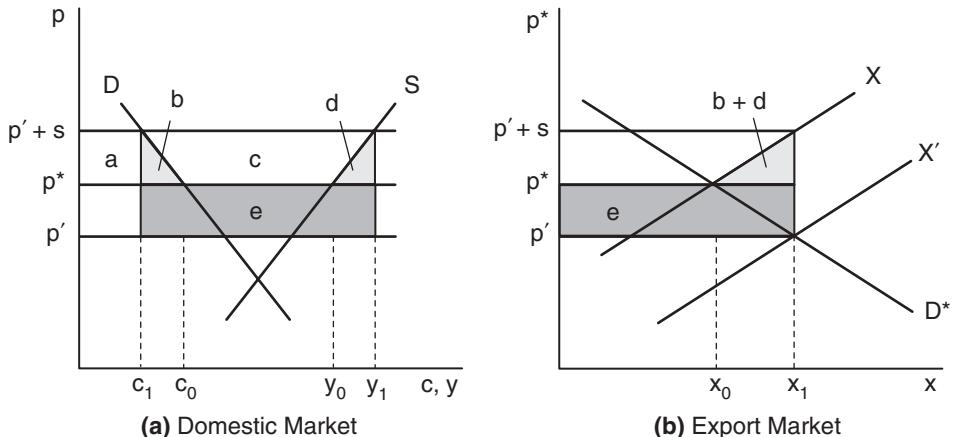


Figure 9.9: Large Country

country is a large producer in world markets, it faces a downward-sloping demand curve D^* from the rest of the world. Therefore, the rightward shift of its export supply curve leads to a *fall* in the price of its exports, from p^* to p' . The exporting country experiences a decline in its terms of trade, which leads to a *further* welfare loss. In figure 9.9, the deadweight loss of the export subsidy is still shown by the triangle $(b + d)$, while the terms of trade loss is shown by the rectangle (e) , and therefore the welfare loss is the sum of these. So unlike the case of an import tariff, which improves the terms of trade gain and potentially raises welfare for a large country under perfect competition, we see that the export subsidy leads to an unambiguous welfare loss.

MANY GOODS

The foregoing demonstration that export subsidies cannot be welfare improving implicitly assumes two goods: the export good in question and another good used to balance trade in the background (for example, the additively separable numeraire good

introduced at the beginning of chapter 8). If there are other goods in the economy, then their prices are being held fixed (in terms of the numeraire). But what if there are other goods whose international prices *also change* with the export subsidy on the good in question. Is it possible for these other goods to have a terms of trade gain, which might overcome the terms of trade loss on the good being subsidized? If so, then the exporting country could gain from the subsidy.

Feenstra (1986) considers a three-good model where a subsidy on one export good, which lowers its price, can nevertheless improve the overall terms of trade for a country and therefore be welfare improving. The condition to ensure this is that the subsidized good and the other export are sufficiently strong *complements*. In that case, the subsidy can lead to an increase in demand and price for the other export good that is enough to offset the terms of trade loss on the subsidized good. A simple example (not drawn from international trade) is the subsidy to help cover travel, lodging, and food offered by many casinos as a way to attract customers: the casino recovers more than enough from gambling revenues to make this subsidy profitable.

The difficulty with the three-good example is that the results are very sensitive to the cross-elasticities in demand and production, above which it is difficult for researchers (let alone a government) to have accurate information. So is there a case where subsidies might be welfare improving even without relying on detailed elasticity information? It turns out that there is, and this is due to Itoh and Kiyono (1987). These authors consider a Ricardian model with a continuum of goods, based on Dornbusch, Fischer, and Samuelson (1977) as discussed in chapter 3. Without presenting this model again in detail, we will summarize the results found by Itoh and Kiyono.

The assumption of the Ricardian model is that countries differ in their technologies. Denote the continuum of goods by $z \in [0, 1]$, and let $a^*(z)$ denote the labor needed per unit of production abroad, while $a(z)$ is the labor needed per unit of production at home. We order these goods such that $a^*(z)/a(z)$ is *declining* in z , meaning that the home country has a comparative advantage in the goods of lowest index z . In equilibrium, the “borderline” good z' will have prices that are equal across countries, so that $wa(z') = w^*a^*(z')$, where w and w^* are wages at home and abroad. Then the home country will be producing and exporting goods in the range $[0, z']$, while the foreign country will be producing and exporting goods in the range $(z', 1]$. Notice that relative wages are determined in equilibrium by $w/w^* = a^*(z')/a(z')$.

Itoh and Kiyono (1987) first consider a home export subsidy of s to *all goods exported*. This has the effect of expanding the range of exports from $[0, z']$ to $[0, z_1]$, with $z_1 > z'$. The new borderline good satisfies $[wa(z_1) - s] = w^*a^*(z_1)$, where $wa(z_1)$ is the home price for that good and $[wa(z_1) - s]$ is the subsidized export price. When we simplify this equation, relative wages are

$$\frac{w}{w^*} - \left(\frac{s}{w^*a(z_1)} \right) = \frac{a^*(z_1)}{a(z_1)}. \quad (9.23)$$

As suggested by this equation, the direct effect of an export subsidy is to *raise* the relative home wage. There is also an indirect effect through the increase in z_1 , which lowers $a^*(z_1)/a(z_1)$, but the net outcome is still an increase in w/w^* due to the subsidy $s > 0$.

However, with a subsidy granted to *all* home exports, the increase in relative wages does not lead to a welfare gain. The reason is that the revenue cost of the subsidy is too great: this cost would be a tax on labor, so that home labor earns $w - [s/a(z_1)]$ in its after-tax wage. The ratio of this after-tax wage to the wage of foreign workers is shown

on the left side of (9.23). But since the right side of (9.23) is declining in z_1 , then so is the relative after-tax wage on the left. So despite the fact that gross wages are increasing at home, the *net after-tax wage* falls relative to that abroad with the expansion of z_1 , and so the uniform export subsidy fails to raise home welfare.

Now consider instead a *targeted* export subsidy $s(z)$, which applies only to those goods in a range $[z_0, z_1]$, with $z_0 < z' < z_1$. Home workers still receive an increase in their relative wages w/w^* , as suggested by (9.23). But now the revenue cost of the subsidy can be lowered by suitable choice of z_0 . All goods below this cutoff do not receive any export subsidy, but since relative home wages have been increased, they will sell in international markets at a higher relative price. Therefore, the increase in home wages corresponds to a *terms of trade increase* for those goods not receiving the subsidy. Indeed, Itoh and Kiyono (1987) demonstrate that by choosing z_0 and z_1 sufficiently close to z' , the revenue cost of the subsidy is of the second order of smalls, whereas the terms of trade gain (which applies to all goods in the range $[0, z_0]$), is of the first order.²⁵ Therefore, the exporting country is better off.

The idea that a country might gain through *targeted* export subsidies is an intriguing one, which deserves more attention. Rodrik (1995b) is among the few papers that takes targeted subsidies seriously, and he explores the ability of these policies to work in various countries.²⁶ That paper is recommended for further reading.

SUBSIDIES WITH IMPERFECT COMPETITION

We next consider the effects of export subsidies under imperfect competition. In particular, suppose that a single home firm and a single foreign firm sell to a third market. The standard example here is sales of Boeing versus Airbus to a third market (such as China). We are interested in whether the home government can give its own firm a “strategic” advantage by subsidizing it. We shall require that the subsidy also be in the *national* interest, which means that profits for the exporter need to rise by more than the amount of the subsidy itself. We initially assume Cournot-Nash competition, as in Brander and Spencer (1985). Eaton and Grossman (1986) contrast the results from that case to Bertrand competition, as we shall also discuss.

Cournot Duopoly

We let x denote the sales of the home firm and x^* the sales of the foreign firm to the third market. We will treat x and x^* as differentiated products, so the home firm earns the price $p(x, x^*)$ and the foreign firm earns $p^*(x, x^*)$. It will be sufficient for our purposes to study the profit maximization of the home firm, as the foreign firm’s problem is similar. Home profits from exporting are

$$\pi = p(x, x^*)x - C(x). \quad (9.24)$$

²⁵The precise form of the export subsidy needed to raise welfare is described in Itoh and Kiyono (1987). Note that goods below z' need to be subsidized since otherwise, with the increase in w , these goods would no longer be exported. Having such “gaps” in the range of exported goods needs to be avoided to raise welfare.

²⁶See also World Bank (1993), which makes a strong distinction between export promotion programs (or “contests”) that are run efficiently and those that are not.

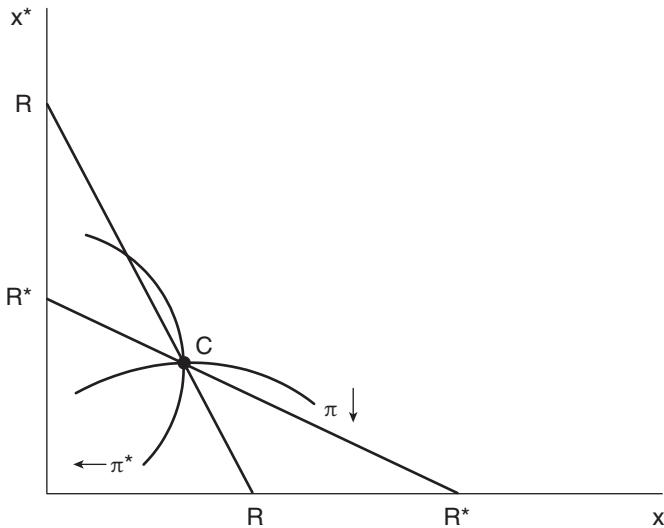


Figure 9.10

Maximizing this over the choice of x , the first-order condition is

$$\pi_x = p(x, x^*) + xp_x - C'(x) = 0, \quad (9.25)$$

and the second-order condition is $\pi_{xx} = 2p_x + xp_{xx} - C'' < 0$.

Using (9.25), we also solve for home exports x as a function of foreign sales x^* , written as the reaction curve $x = r^*(x)$. This reaction curve and the foreign counterpart $x^* = r^*(x)$ are graphed as RR and $R'R^*$ in figure 9.10, and their intersection determines the Cournot equilibrium denoted by point C . The typical property of these reaction curves is that they are both downward sloping. For stability, the home reaction curve RR needs to cut the foreign reaction curve $R'R^*$ from above, as illustrated.²⁷ Note that the iso-profit curves of π have higher profits in the downward direction (i.e., for reduced x^*), and similarly the iso-profit curves π^* have higher foreign profits in the leftward direction (for reduced x), as illustrated.

Now consider the impact of an export subsidy. Letting $p(x, x^*)$ denote the price paid by the buyer and s denote the *specific* subsidy, then $p(x, x^*) + s$ is the subsidized price received by the home firm. Home profits are, therefore,

$$\pi = [p(x, x^*) + s]x - C(x). \quad (9.26)$$

Maximizing this over the choice of x , the first-order condition is

$$\pi_x = p(x, x^*) + s + xp_x - C'(x) = 0, \quad (9.27)$$

where (9.27) defines a new reaction curve $x = r(x^*, s)$. To check how this reaction curve depends on the subsidy s , totally differentiate (9.27) to obtain

²⁷In problem 8.3, you are asked to derive these properties of the reaction curves.

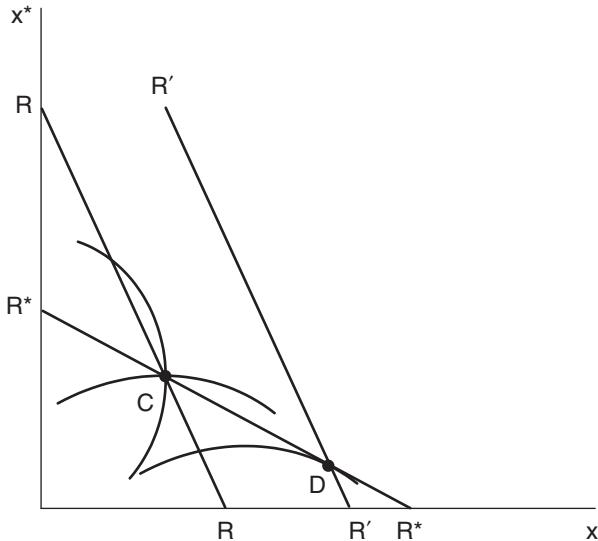


Figure 9.11

$$\frac{dx}{ds} = -\frac{1}{\pi_{xx}} > 0,$$

where this sign is obtained from the second-order condition for maximizing (9.26).

Thus, in figure 9.11 the export subsidy unambiguously shifts the home reaction curve to the right, from RR to $R'R'$, so the equilibrium moves from point C to point D . Under the case we have shown with both reaction curves downward sloping, this will increase home exports x and reduce foreign exports x^* . It is readily confirmed that this will raise home profits, which is not surprising. Our interest is not just in the profits of the home firm, however, but also in *social welfare*, defined as the profits earned by the home firm in the third market minus the revenue cost of the subsidy.²⁸ Since the subsidy cost is the amount sx , welfare becomes

$$W = [p(x, x^*) + s]x - C(x) - sx = p(x, x^*)x - C(x). \quad (9.28)$$

Notice that the expression on the right side of (9.28) is identical to the *original* expression for home profits in (9.24). That is, by taking the subsidized profits in (9.26), and subtracting the amount of subsidy, we get back the initial profits. This means that the original iso-profit curves for the home firm, as illustrated by π in figure 9.10, now measure social welfare.

The difference between profits in (9.24) and social welfare in (9.28), however, is that they are evaluated at *different equilibrium quantities* (x, x^*). Whereas initial profits are evaluated at point C , the subsidy moves the equilibrium to point D . This shift clearly *increases* home welfare (since the iso-profits and welfare contours are increasing in the downward direction). Indeed, an *optimal* subsidy would be one for which the welfare

²⁸Since the home and foreign firms are both operating in a third market, there is no impact on home consumers, and so we do not need to take this into account in social welfare.

contour is just tangent to the foreign reaction curve R^*R^* , as illustrated in figure 9.9 at point D . We see that under Cournot-Nash competition, the export subsidy by the home government leads to a welfare improvement.

Bertrand Duopoly

We now contrast the results obtained above to a Bertrand duopoly, when the home and foreign firms are each choosing prices, rather than quantities, as the strategic variable. Denote exports of the home firms to the third market by $x(p, p^*)$, where p is the price paid by the buyer. Letting s denote the *specific* subsidy, then $(p + s)$ is the subsidized price received by the home firm. Home profits are, therefore,

$$\pi = (p + s)x(p, p^*) - C[x(p, p^*)]. \quad (9.29)$$

Maximizing this over the choice of p , the first-order condition is

$$\pi_p = x(p, p^*) + (p + s)x_p - C'(x)x_p = 0. \quad (9.30)$$

Given the foreign price p^* and the subsidy, we can use (9.30) to solve for home export price p , obtaining the reaction curve $p = r(p^*, s)$. The analogous condition for the foreign exporter can be used to solve for foreign reaction curve $p^* = r^*(p)$. These reaction curves are illustrated in figure 9.12, with the typical properties that they are upward sloping and have a *dampened* response of each price to that of the competing good, so that $\frac{dp}{dp} \frac{p}{p} = r_p(p^*, s) \frac{p}{p} < 1$, and $\frac{dp}{dp} \frac{p}{p} = r_p^*(p) \frac{p}{p} < 1$.²⁹

With a zero subsidy, the intersection of these reaction curves determines the Bertrand equilibrium, at point B in figure 9.12. The iso-profit curves of π have higher profits in the rightward direction (i.e., for higher p^*), and similarly the iso-profit curve π^* have higher foreign profits in the upward direction (for higher p), as illustrated.

Now consider how the application of an export subsidy shifts the home reaction curve. Totally differentiating (9.30), we obtain

$$\frac{dp}{ds} = -\frac{x_p}{\pi_{pp}} < 0, \quad (9.31)$$

where the sign of (9.31) is obtained because $x_p < 0$, and $\pi_{pp} < 0$ from the second-order condition for profit maximization. Thus, (9.31) shows that the export subsidy will unambiguously *lower* the price charged by the home firm for its exports (given the foreign price). In figure 9.13, the equilibrium is moved from point B to a point like D .

To see how this affects home welfare, we again subtract the cost of the subsidy from (9.29) to obtain home welfare:

$$W = (p + s)x(p, p^*) - C[x(p, p^*)] - sx(p, p^*) = px(p, p^*) - C[x(p, p^*)]. \quad (9.32)$$

In (9.32), we once again obtain a measure of welfare that is identical to home profits in the absence of the export subsidy. This is shown by the iso-profit curve labeled π in figure 9.12 and W in figure 9.13, and increases in the rightward direction, that is,

²⁹In problem 8.6 you are asked to demonstrate these properties of the reaction curves.

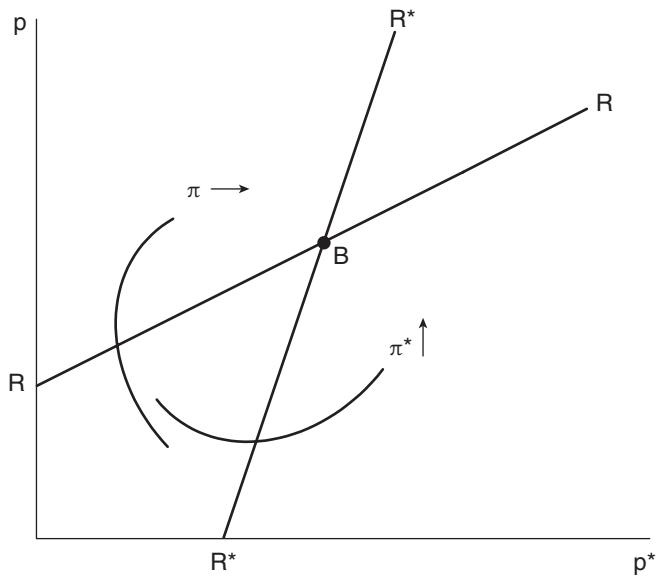


Figure 9.12

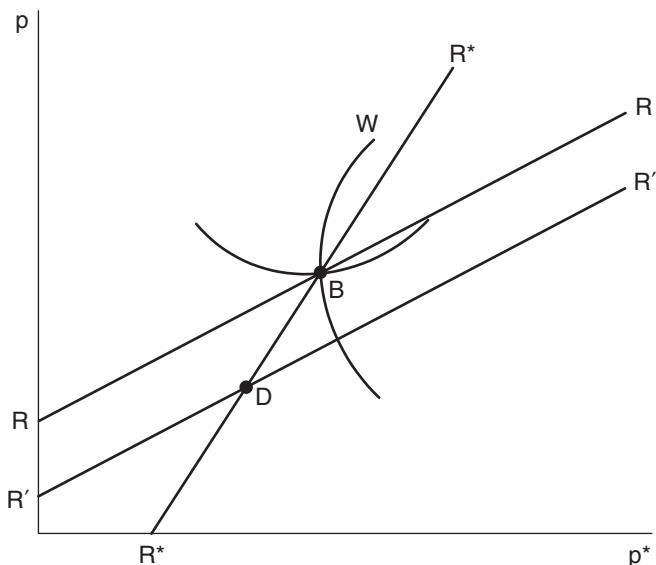


Figure 9.13

when the foreign price p^* increases. It is clear from figure 9.13 that the export subsidy, by leading to a *fall* in prices, has *reduced* home welfare. The reason is that any increase in home profits—measured inclusive of the subsidy—is more than offset by the revenue cost of the subsidy, so that profits measured without the subsidy—which equals welfare—has declined.

In order for the home country to gain from an export policy, it must impose an export *tax* rather than a subsidy. This has the effect of moving the equilibrium above

B , upwards along R^*R^* in figure 9.13, and raising home welfare due to the increase in prices. These results are summarized with the theorem that follows.

THEOREM (BRANDER AND SPENCER 1985; EATON AND GROSSMAN 1986)

- (a) Under Cournot duopoly, a subsidy to exports raises home welfare;
- (b) Under Bertrand duopoly, a tax on exports raises home welfare.

Part (b), due to Eaton and Grossman (1986), shows that any “strategic” use of export subsidies is extremely sensitive to the form of competition. Since we do not expect the government to be able to distinguish the market conduct, it becomes impossible to implement these policies in a welfare-improving manner. Furthermore, even if the government knew that a market was operating under Bertrand competition, the home firm would generally oppose this policy, and in practice export taxes are rarely used.³⁰

Eaton and Grossman have extended this theorem to discuss other forms of market conduct, which we do not cover. Regardless of the form of competition, however, there are a number of other reasons why export subsidies should not be expected to generate a welfare improvement. First, as we found with a tariff, the possible gains for the home country can be dissipated by free entry into the industry, so that profits are reduced to zero. This was demonstrated by Horstmann and Markusen (1986).

Second, we have assumed above that the home exporter takes the subsidy as exogenous, that is, the government sets the export subsidy in the first stage of the game, with the home and foreign firm competing in the second. But Carmichael (1987) argues that in the case of financing for export credits (such as provided by the U.S. Export-Import Bank to purchasers of U.S.-produced aircraft), these subsidies should instead be viewed as endogenous: the government essentially guarantees a low interest rate to the buyers of aircraft in the second stage of the game, with firms optimizing over price in the first stage. In that case, Carmichael shows that even under Cournot competition the export subsidies no longer improve welfare. This result is similar in spirit to what we found in the previous chapter for antidumping duties, which fail to be welfare improving for the importing country when the foreign exporter treats them as endogenous.

Third, it may well be the case that exporters differ in their productivities, which are not observable to the government. Brainard and Martimort (1997) have analyzed this possibility in the context of Cournot competition between home and foreign exporters. With multiple home exporters, the home government needs to design a subsidy policy that induces them to truthfully reveal their productivities. It turns out that only the most efficient firms receive subsidies in this case, whereas the least efficient may very well be taxed.

Finally, even if the above concerns do not apply, the foreign country could very well respond by applying subsidies of its own. In that way, the two exporters end up in a “subsidy war,” which lowers welfare for both of them but raises welfare for the importing country.³¹ Of course, this concern also arises with tariffs, where a tariff applied by one country can invite retaliation by the foreign country and both countries

³⁰An interesting exception is the tax applied to U.S. exports of cotton in the antebellum period, as discussed by Irwin (2003).

³¹See problem 9.5.

can loose from this tariff war. This will be discussed in more detail in the next chapter. Interestingly, Janeba (1998) shows that when firms can choose their country of production, the incentive to use *nondiscriminatory* subsidies is greatly diminished, since these subsidies apply equally well to home firms and foreign firms located in the home market. In that case, Janeba finds that *zero subsidies* provide the only equilibrium in the subsidy game between the countries, so that insisting on nondiscrimination is a simple device to assure this outcome.

Conversely, there are several reasons why export subsidies may be welfare improving regardless of the form of market conduct. Subsidies are often given to R&D rather than exports. Since R&D is a fixed cost that a firm must incur, these subsidies will influence the decision of a firm to enter the market or not. Dixit and Kyle (1985) investigate export subsidies in a model where these may deter entry by foreigners and encourage domestic entry, and find that they may be beneficial. Related to this issue is Maggi's (1996) investigation of subsidies to capacity rather than exports, in a model that is general enough to include both Bertrand and Cournot competition. He finds that these subsidies are weakly welfare improving; that is, they will raise or leave unchanged the welfare of the exporting country. Finally, export subsidies may be used in markets where learning is important: either producers learning about costs, or consumers learning about product quality. Bagwell (1991) has analyzed the case of consumers learning about quality and finds a first-order gain due to export subsidies. The case of falling costs is important in the manufacture of semiconductors, for example, and also aircraft.³² Then in a way similar to the infant-industry argument for a tariff, subsidies may possibly be welfare improving.

SUBSIDIES TO COMMERCIAL AIRCRAFT

A favorite industry for studying the impact of export subsidies is aircraft. In the market for large passenger aircraft, there have been just three competitors in recent decades: Boeing and McDonnell-Douglas in the United States and Airbus in Europe. The former two companies merged on August 1, 1997, and so the industry effectively became a duopoly. Various types of subsidies have been used in both the United States and Europe to support their respective firms, so this is an ideal case to evaluate the theoretical results described above.

The idea that two countries can be in a “subsidy war” appears to apply in this industry. Recognizing this possibility, the United States and the Europe Community reached an agreement in 1992, which limited subsidies to 33 percent of development costs (see Tyson 1992; Pavcnik 2002a). In addition, the agreement prohibits production subsidies and limits the ability of government agencies to subsidize the interest rate on purchases of aircraft. Irwin and Pavcnik (2004) find that as a result of this accord, prices for commercial aircraft rose by something between 3.1 and 8.8 percent. They estimate that the reduction in subsidies granted is somewhat higher than this amount—between 7.5 and 12.5 percent.

There are claims that the terms of the 1992 agreement were violated by a new aircraft developed by Airbus. As described by Irwin and Pavcnik (2004), commercial aircraft are made up of two distinct product types: narrow-body aircraft with a single

³²A simulation model of subsidies to aircraft is presented by Baldwin and Krugman (1988b), and the dynamics of production and sales in this industry are estimated by Benkard (2000, 2004).

aisle, including the Boeing 737, 757, and Airbus A-320; and wide-body aircraft with two aisles, including the Boeing 747, 777, and Airbus A-300. Since 2007, Airbus has produced another wide-body aircraft, the A-380, which is even larger than the Boeing 747 and competes directly with the 747 in long flights. The expenditures to develop the A-380 were estimated to be \$12 billion, of which the governments of France, Germany, the Netherlands, Belgium, Spain, Finland, and the United Kingdom covered up to one-third. This subsidy led to a trade dispute between the United States and Europe over whether financing of the A-380 violates the 1992 agreement, as described in Pavcnik 2002.

The goal of Irwin and Pavcnik (2004) is to estimate demand and markups for wide-body aircraft, and then simulate the effect of A-380's entry on the market equilibrium. Airbus has stated that it needs to produce at least 250 planes to cover the development costs, but that it expects to sell 1,500 aircraft over the next twenty years. It has received initial orders for sixty planes (as of early 2001), but these purchasers appear to have received a discount of at least 10 percent, and perhaps as high as 35 percent, off the list price of \$235 million per plane. Boeing, on the other hand, states that market demand for such a superjumbo will not exceed seven hundred aircraft over the next twenty years. We can expect the estimation of market demand and prices, and the simulation of A-380's entry, to inform us as to which projection (700 planes or 1,500 planes) is more accurate.

Demand for a differentiated product depends on the quality-adjusted prices, which we wrote earlier as $q_i = p_i/z_i$. Suppose that we instead measure quality-adjusted prices q_i by the difference between nominal prices p_i and quality. In addition, we assume that quality is a linear function of characteristics, so the quality-adjusted prices become $q_j = \alpha p_j - \beta' z_j - \xi_j$. In this notation, ξ_j reflects unmeasured characteristics of product j , which we treat as a random variable. Then the logit specification of demand is

$$s_i = \frac{e^{-q_i}}{\left[\sum_{j=1}^N e^{-q_j} \right]} = \frac{e^{\beta' z_i - \alpha p_i + \xi_i}}{\left[\sum_{j=1}^N e^{\beta' z_j - \alpha p_j + \xi_j} \right]}, \quad (9.33)$$

where s_i denotes the *market share* of product i , measured by quantity demanded of product i relative to total quantity of all products $j = 1, \dots, N$. As shown in appendix B, the logit specification in (9.33) arises quite naturally from a discrete choice model and can also be viewed as the demand from a representative consumer using McFadden's aggregation theorem.

To estimate the parameters of (9.33), we take logs and difference with an “outside” good indexed by 0, to obtain

$$\ln s_i - \ln s_0 = \beta' z_i - \alpha p_i + \xi_i \quad (9.34)$$

where the quality-adjusted price q_0 of the “outside” good is normalized to zero. Thus, in (9.34) we regress the difference in market shares on the prices and characteristics of product i , to estimate the parameters (α, β) . The approach can be contrasted with the hedonic regression in (9.17), where we are regressing *prices* on characteristics, while in (9.34) we are regressing *market shares* on characteristics and price. In both cases, we interpret β as the consumer valuation of characteristics. We can think of (9.34) as an alternative procedure to the hedonic regression, as proposed by Berry (1994).

Irwin and Pavcnik treat wide-body aircraft as differentiated into two market segments: those for medium-range and those for long-range trips. Denoting the products

in each of these two groups by J_g , for $g = 1, 2$, Berry (see appendix B) shows that the quantity-share of demand for each type of aircraft is rewritten as

$$\ln s_i - \ln s_0 = \beta' z_i - \alpha p_i + \rho \ln s_{ilg} + \xi_p, \quad i \in J_g, \quad (9.35)$$

where s_i is the share of demand for wide-body aircraft i relative to all aircraft purchased annually; s_0 is the share of demand for narrow-body aircraft (treated as the “outside” good); and s_{ilg} is the share of wide-body aircraft i *within* the medium range or long-range group $g = 1, 2$. We expect that the parameter ρ is positive but less than unity, and Irwin and Pavcnik estimate ρ as 0.41 (standard error of 0.17) in their preferred specification.³³ This means that long-range wide-body aircraft are closer substitutes for each other than they are for medium-range aircraft. For example, the A-380 will compete more with the Boeing 747 than with the other medium-range aircraft sold by Airbus.

We can also use the demand equation (9.35) to obtain the optimal prices for multiproduct firms. Irwin and Pavcnik (2004) consider both Bertrand and Cournot competition, but we focus here on the former case. Suppose that the total demand for aircraft (wide-body plus narrow-body) is given by H units per year. The demand for each model is then $s_i H$. If the firm in question sells products $i = 1, \dots, M$, then its profit maximization problem is

$$\max_{p_i} \sum_{i=1}^M [p_i - g_i(z_i)] s_i H, \quad (9.36)$$

where $g_i(z_i)$ denotes marginal costs. Maximizing this over the choice of prices p_i , we obtain the first-order conditions

$$s_i + \sum_{j=1}^M [p_j - g_j(z_j)] \frac{\partial s_j}{\partial p_i} = 0, \quad i = 1, \dots, M. \quad (9.37)$$

To simplify this expression, the derivatives of the market shares can be calculated from (9.35) and substituted into (9.37). In addition, we suppose that marginal costs are linear in characteristics, so $g_i(z_i) = \gamma' z_i + \omega_i$, where ω_i is a random error.

Irwin and Pavcnik (2004) estimate the demand parameters from (9.35) and substitute these into (9.37) in order to compute the markups. They find that these markups are high: current prices are about twice as high as marginal costs (under either Bertrand or Cournot competition), and even higher in past years. Having obtained estimates of demand and markups, they then simulate the effect of A-380's entry into the market. On the demand side, we know that this plane will be a close substitute for other long-range, wide-body aircraft such as the Boeing 747. But on the pricing side, having multiproduct firms means that Airbus will be *less* concerned than Boeing about this new product stealing demand away from its existing products. That is, Boeing will respond more aggressively than Airbus in lowering its own prices to prevent market share from falling (since Airbus is happy to have demand shifted toward its own new product). These predictions are borne out in the simulations. Relative to their previous sales, the percentage decline in demand is smaller for Boeing than for

³³When estimating (9.35) we need to use instruments to correct for the endogeneity of both s_{ilg} and price on the right side; the estimate of 0.41 comes when using the widest set of instruments.

Airbus, because Boeing responds more aggressively in cutting its prices. Nevertheless, including the new sales of A-380, the overall market share of Airbus increases.

How many of the new A-380s will Airbus sell? This depends, of course, on the price charged. Irwin and Pavcnik experiment with discounts from the list price of 0–30 percent. With the largest discount, they calculate that Airbus will sell thirty-four planes annually, or 680 over twenty years, yielding \$110 billion in revenues. But annual sales of the A-380 are only one-half (one-quarter) as much when discounts of 20 percent (10 percent) are used instead, indicating that demand for the A-380 is highly elastic. High markups in this market are sustained because firms are selling multiple products, but at the full list price for the A-380 its demand is extremely low—only four or five planes annually, which would not cover development costs. Lack of detailed cost data means that we cannot estimate the profit-maximizing price for the A-380, but it is evident that high discounts are needed to achieve significant sales. Even with the 30 percent discount, however, demand falls well below the market projection of 1,500 planes made by Airbus. Rather, sales at that discount are closer to the seven hundred planes expected by Boeing, whose company representatives state that demand at that level will not allow the project to be profitable.

CONCLUSIONS

In the previous chapter, we identified three effects of a tariff on welfare: (a) a dead-weight loss; (b) a terms of trade effect; (c) a reduction in the monopoly distortion if the output of home firms increases (without leading to inefficient entry). While again we have not derived the welfare criterion for an import quota, a similar decomposition would apply. However, any potential terms of trade gain becomes a loss if the importing country gives up the quota rents to the exporter, as under a VER. In that case, the only possible source of gain for the importing country would be under the third effect, if the import quota led to a significant increase in home output and offset a monopoly distortion. But we have found that the reverse case is more likely to hold: with either a home monopoly or Bertrand duopoly, the quota (say, at the free-trade level of imports) will lead to an increase in price and reduction in home quantity, so that the monopoly distortion is worsened rather than offset.

For export subsidies under perfect competition, the welfare criterion becomes: (a) a deadweight loss; (b) a terms of trade effect. Whereas the terms of trade will improve due to a tariff for a large country, it will instead *worsen* due to the export subsidy in a two-good model. We have noted, however, that with more goods there is the possibility that a subsidy on some goods will raise the terms of trade for others. The Ricardian model with a continuum of goods provides a useful illustration and allows for welfare-improving export subsidies when they are targeted on a narrow range of goods. Determining whether this has occurred in any actual cases would be an interesting area for further research.

Under imperfect competition, the welfare criterion is simplified by considering sales to a third market. In this case the home firm earns profits from its export sales, and welfare is simply the difference between these and the revenue cost of the subsidy. This provides the clearest example of a potential role for “strategic” trade policy in shifting profits toward the home firm. As we have seen, however, subsidies may or may not be desirable: under Cournot competition export subsidies raise welfare, but

under Bertrand competition they do not. Since it is very difficult to know the type of competition being used by firms, it becomes impossible for the government to implement this policy in a welfare-improving manner.

Empirically, the analysis of import quotas and export subsidies are linked because the industries involved are often producing discrete products, with multiproduct firms. This chapter has served as an introduction to the empirical techniques that can be used on such industries, including hedonic regressions and the estimation of demand and prices as in Berry 1994. The reader is referred to appendix B for more details on discrete choice methods, which should be useful for many trade policy issues.

PROBLEMS

- 9.1 Suppose that the home and foreign firms are engaged in *Cournot* competition in the home market. Illustrate the free-trade equilibrium, and answer the following:
 - (a) Suppose that a VER is negotiated with the foreign firm so that it agrees to *not sell more* than its free-trade quantity. What is the effect on the equilibrium?
 - (b) Suppose that the VER is negotiated at *below* the free-trade quantity. Show the effect on the equilibrium.
- 9.2 Suppose that the home and foreign firms are initially engaged in *Bertrand* competition in the home market, as discussed in the text. Suppose that the VER is negotiated at *below* the free-trade quantity. Show the effect on the equilibrium and on profits.
- 9.3 Building on problem 5.3 in chapter 5, consider the CES utility function written as in (9.2),

$$U = \sum_{i=1}^N a_i c_i^\theta, \quad \theta = (\sigma - 1)/\sigma,$$

where $a_i = z_i^\theta$ is a transformation of the quality of each product, and can be treated as a parameter, $i = 1, \dots, N$.

- (a) Maximize this subject to the budget constraint, $\sum_{i=1}^N p_i c_i \leq I$, to obtain an expression for the demands c_i as a function of prices and income.
- (b) Use these expressions for demand to show that $\partial c_i / \partial p_j = \partial c_j / \partial p_i, i \neq j$.
- (c) Also show that equation (9.11) holds, which we rewrite as

$$-\left(\sum_{j=1}^M \frac{\partial \ln c_j}{\partial \ln p_j}\right) = \sigma + (1 - \sigma)\left(\sum_{j=1}^M \frac{p_j c_j}{I}\right).$$

- 9.4 Consider the problem of maximizing profits subject to the VER constraint. Using the same steps as were used under free trade, but ignoring any specific transport costs T_i , show that the solution for prices is $p_i = \mu[g_i(z_i) + \lambda]$.
- 9.5 Consider a Cournot duopoly exporting to a third market, and suppose that *both* countries apply export subsidies. Illustrate where this “subsidy war” can end up; that is, illustrate the Nash equilibrium in subsidies for both governments.

EMPIRICAL EXERCISE

These exercises to reproduce regression results Feenstra (1988a). To complete the exercises, the files “car_7990.dta, truck_7990.dta” should be stored in: c:\Empirical _Exercise\Chapter_9\. Then do the following:

- 9.1 Run the programs “pindex_c.do, pindex_t.do, unit_value.do” to reproduce the price indexes and unit-values for cars and trucks in table 9.3. What formula is used for the price indexes, and how do these differ from the unit-values?
- 9.2 Run the program “car_reg.do” to reproduce the column (1) in table 9.4, and the program “truck_reg.do” to reproduce column (2). What weights are being used in the regression, and how does this affect the results?
- 9.3 Pooling car and truck data, run “system_7985.do” to reproduce columns (3) and (4) in table 9.4 with the constraints specified in equation (9.22). How are these constraints built into the program for the nonlinear regression?
- 9.4 Run the programs “unit_quality.do, qindex_c.do, qindex_t.do” to reproduce the quality indexes and unit-qualities for cars and trucks in Table 9.3. What formula is used for the quality of each model and the quality indexes, and how do these differ from the unit-qualities?
- 9.5 Pooling car and truck data, run “system_nocom.do” and “system_wcon.do” to reproduce columns (1)–(4) in table 9.5, with constraints specified in equation (9.22).

Political Economy of Trade Policy

In the previous two chapters we have covered the basic analysis of trade policies, including models of imperfect competition. The idea that noncompetitive markets give rise to opportunities for governments to gain from trade policies has been an important line of research. It has ultimately been concluded, however, that such opportunities for “strategic” use of trade policy are very limited. This raises the obvious question of why trade policies are used so often. One answer is that such policies are politically motivated: tariffs are granted in response to demands by special interest groups, such as industries and unions. The research issue is to understand how such demands are mediated through the political process. In this chapter we outline research on the political economy of protection, including the median voter model of Mayer (1984) and the “protection for sale” model of Grossman and Helpman (1994).

The “protection for sale” model has been extended in a number of directions. Mitra (1999) has shown how to introduce *endogenous* lobbies into that framework, as we shall discuss. Grossman and Helpman (1995a) further use this framework to analyze a tariff war between two countries, and to consider the potential benefits from international agreements. Along the same lines, Bagwell and Staiger (1999, 2002) examine the economic rationale for the trade rules embodied in the General Agreement on Trade and Tariffs (GATT) and the World Trade Organization (WTO). These rules include GATT’s principle of “reciprocity,” whereby each country agrees to reduce its trade barriers in return for a reciprocal reduction by another. Bagwell and Staiger’s framework is general enough to include both the median voter model and the “protection for sale” model, and we use it to show the outcome of a trade war, along with the benefits from reciprocal tariff reductions using GATT rules.

Another important principle of GATT is the “most favored nation” (MFN) provision, which states that all member countries of GATT should be granted the same tariffs. This provision is violated when countries join into regional trade agreements, granting zero tariffs to countries *within* but not outside the agreement. Our third topic is to examine the incentives for countries to join regional trade agreements versus multilateral agreements. A key question here, raised by Bhagwati (1993; 2002, lecture 3), is whether joining a regional trade agreement *helps or hinders* the ultimate goal of multilateral free trade. Models indicating that regional agreements may hinder multilateral free trade include Krishna (1998) and McLaren

(2002), and views on the other side include Baldwin (1995) and Ethier (1998). We will examine this issue using the median voter model, as in Levy (1997). In the basic version of this model a regional agreement cannot hinder a multilateral agreement, but in an extended version of the model (including product variety) this can occur.

We conclude the chapter with an application of political economy to a nondemocratic setting: the People's Republic of China. In this case, the special interest groups do not take the form of well-funded lobbying organizations as in the United States, but rather include the large state-owned industries and the smaller but growing private industries, as well as foreign firms. The distribution of these firms is very uneven across provinces. China's entry into the WTO in 2001 came only after regional concerns in China about the impact of import competition were overruled. We argue that the regional variation toward openness (as evidenced by the inflow of foreign firms), combined with the variation in location of state-owned firms, can be used to identify the political weights given to the various interest groups. We present the results of Branstetter and Feenstra (2002), who show how the Grossman-Helpman framework can be applied in this setting.

MEDIAN VOTER MODEL

Used in many applications, the median voter model presumes that policies are established by majority vote. Provided that preferences are "single peaked" over the policy being voted upon (i.e., each person has a unique maximum), then it follows that the policy adopted will maximize the utility of the median voter. In the application to trade policy we assume that the policy is an import tariff or subsidy. The optimal tariff for the median voter will depend on the production structure in the economy, and for simplicity in this section we assume the two-by-two Heckscher-Ohlin (HO) model (Mayer (1984) also considers other structures). Then we will show that if the median voter owns a lower capital/labor ratio than overall for the economy, and if imports are labor intensive, then the economy will have a positive import tariff.

As in chapter 8, we will suppose that each individual has a quasi-linear utility function given by $c_0^h + U(c^h)$, where c_0^h is consumption of a numeraire export good, and c^h is the consumption of the import good for consumer $h=1, \dots, L$. (We use L rather than H to denote the number of consumers, since L also equals the number of workers). Consumers all have the same increasing and strictly concave utility function U , so they have the same optimal consumption $c^h = d(p)$, $d'(p) < 0$, with remaining income spent on the numeraire good, $c_0^h = I^h - p'd(p)$. Then individual utility is

$$V(p, I^h) \equiv I^h - p'd(p) + U[d(p)]. \quad (10.1)$$

Both the export and import goods are produced using labor and capital. The total endowments of labor and capital are L and K , respectively. The fixed world price of the import is denoted by p^* , and this good has a specific tariff of t , so the domestic price is $p = p^* + t$. We let $y(p)$ denote the supply of the import-competing good, with $y'(p) > 0$. Imports are then $m(p) = d(p)L - y(p)$, and tariff revenue collected is $T = tm(p)$, which is redistributed with a poll subsidy. We will suppose that individual h has one unit of labor and K^h units of capital, $h=1, \dots, L$, so that individual income is $I^h = w + rK^h + (T/L)$. We can reexpress this as

$$I^h = \frac{1}{L}(wL + rK^h L + T) = \frac{1}{L}(wL + \rho^h rK + T), \quad (10.2)$$

where $\rho^h = K^h/(K/L)$ is the capital/labor ratio for the individual in question relative to the *overall* capital/labor ratio in the economy. Total GDP in the economy is $G = y_0(p) + py(p) = wL + rK$. It follows that we can rewrite individual income in (10.2) as

$$I^h = \frac{1}{L}[wL + rK + (\rho^h - 1)rK + T] = \frac{1}{L}[(\rho^h - 1)rK + y_0(p) + py(p) + T]. \quad (10.2')$$

Differentiating individual utility in (10.1) with respect to the tariff, we obtain

$$\begin{aligned} \frac{dV^h}{dt} &= -d(p) + \frac{dI^h}{dt} \\ &= (\rho^h - 1)\frac{dr}{dp}\frac{K}{L} + \left[\frac{y(p)}{L} - d(p) \right] + \frac{1}{L}\frac{dT}{dp} \\ &= (\rho^h - 1)\frac{dr}{dp}\frac{K}{L} + \frac{t}{L}m'(p), \end{aligned} \quad (10.3)$$

where the first line follows from Roy's Identity, the second line using (10.2') and the envelope theorem, and the third line using tariff revenue of $T = t[d(p)L - y(p)] = tm(p)$.

If the tariff is determined by majority vote, then the tariff prevailing will be that which maximizes the utility of the median voter. Denoting the median voter by m , with utility $V^m = V(p, I^m)$, this tariff will satisfy $dV^m/dt = 0$ and $d^2V^m/dt^2 < 0$. Setting (10.3) equal to zero, this tariff is

$$t^m = (1 - \rho^m)\frac{dr}{dp}\frac{K}{m'(p)}, \quad (10.4)$$

where ρ^m is the capital/labor ratio for the *median* individual h relative to the overall capital/labor endowment for the economy. This ratio is less than one for all countries (Alesina and Rodrik 1994), so that $\rho^m < 1$. Then, since $m'(p) < 0$, we see that the tariff t^m is *positive* when the import good is labor intensive, so that $dr/dp < 0$, but *negative* when the import good is capital intensive, so that $dr/dp > 0$. In other words, import tariffs should be used in capital-abundant industrialized countries, but import subsidies in labor-abundant developing countries.

In practice, import subsidies are rarely observed, despite this prediction from the median voter model. There are many reasons for this, some of which we will investigate later in this section. Setting aside this most obvious limitation of the median voter model, Dutt and Mitra (2002) ask whether there is some *other* prediction that might accord better with real-world evidence. In particular, suppose that we compare countries with varying degrees of inequality, which we measure by $(1 - \rho^m)$, that is, with lower values of the median voter's capital/labor endowment ρ^m , corresponding to higher inequality. Then differentiating the first-order condition $\frac{dV^m}{dt} = 0$, we obtain $\frac{d^2V^m}{dt^2} + \frac{d^2V^m}{d(1-\rho^m)dt} d(1 - \rho^m) = 0$, so that from (10.3)

$$\frac{dt^m}{d(1 - \rho^m)} = \frac{dr}{dp}\frac{K}{L} \Big/ \frac{d^2V^m}{dt^2}, \quad (10.5)$$

where $d^2V^m/dt^2 < 0$ from the second-order condition.

It follows that for capital-abundant countries importing the labor-intensive good, so that $dr/dp < 0$, the median voter model predicts that *increased* inequality (a rise in $1 - \rho^m$) will lead to a *higher* tariff. Conversely, for labor-abundant countries importing the capital-intensive good, so that $dr/dp > 0$, the median voter model predicts that *increased* inequality (a rise in $1 - \rho^m$) will lead to a *reduced* tariff or increased subsidy (a fall in t^m). Stated less formally, we expect increased inequality to be associated with more restrictive trade policies in industrialized countries, but more open trade policies in developing countries.

Dutt and Mitra test this prediction by running the regression

$$TR^i = \alpha_0 + \alpha_1 INEQ^i + \alpha_2 INEQ^i(K/L)^i + \alpha_3(K/L)^i + X^i\beta + \varepsilon_i, \quad (10.6)$$

where TR^i is a measure of trade restrictions in country i , $INEQ^i$ is an index of income inequality in that country, $(K/L)^i$ is the capital/labor ratio in country i , and X^i is a matrix of other control variables. Taking the partial derivative of TR^i with respect to $INEQ^i$, we obtain

$$\frac{\partial TR^i}{\partial INEQ^i} = \alpha_1 + \alpha_2(K/L)^i. \quad (10.6')$$

The prediction from the median voter is that this derivative should be negative for low levels of the capital/labor ratio $(K/L)^i$, but positive for higher levels of the capital/labor ratio. This will occur if $\alpha_1 < 0$ and $\alpha_2 > 0$, with the turning point between the negative and positive derivatives occurring where (10.6') equals zero, or at the capital/labor ratio $(K/L)^i = -\alpha_1/\alpha_2 > 0$.

In their estimates, Dutt and Mitra confirm these signs of α_1 and α_2 for several different measures of tariffs used for TR^i . The turning point $(K/L)^i = -\alpha_1/\alpha_2$ turns out to be quite close to the median capital/labor ratio in the sample (South Korea). For developing countries with lower capital/labor ratios, greater inequality leads to lower tariffs. Conversely, for industrialized countries with higher capital/labor ratios, greater inequality leads to higher tariffs. This provides striking support for the median voter framework in the context of the Heckscher-Ohlin model. In addition, Dutt and Mitra find that this relationship holds better in democracies than in dictatorships.

This confirmation of the median voter model seems at odds with its first prediction, that tariffs should be positive in advanced countries importing labor-intensive goods, but negative in developing economies importing capital-intensive goods. As we have mentioned, import subsidies are rarely observed. Fernandez and Rodrik (1991) provide one explanation for the “antitrade” bias of nearly all countries. Specifically, they argue that even if policies are determined by majority vote, when individuals do not know whether they will be included among the gainers or losers, there is a tendency for voters to prefer the status quo. This occurs even in a model where everyone is perfectly informed about the aggregate gains and losses in each industry, but cannot predict their individual returns. Therefore, there is a tendency to apply tariffs to offset import competition and preserve the status-quo income distribution.¹ This can help to

¹ The idea that trade policy is applied to preserve the status quo income distribution has also been proposed by Corden (1974), and is known as the *conservative social welfare function*; see also Deardorff (1987). In Fernandez and Rodrik (1991), this is the outcome of majority vote rather than a criterion imposed by the government.

reconcile the positive tariffs observed in most countries with the median voter model, and the logic of Fernandez and Rodrik also holds equally well in other models.

PROTECTION FOR SALE

The median voter model assumes that policies are determined by majority vote. This is an overly simplified description of representative democracies where the electorate votes for legislators, who then determine the policies. In such settings the policies chosen will be jointly influenced by votes, voice, and dollars from the campaign contribution of lobbying groups. The second model we consider, due to Grossman and Helpman (1994), proposes an elegant solution to the problem of how the government simultaneously considers the contributions of numerous lobbies, as well as consumer welfare, in determining trade policy.²

We assume that there are N goods plus the numeraire commodity, all of which are additively separable in utility. Consumer utility functions are $c_0^h + \sum_{i=1}^N u_i(c_i^h)$, where c_0^h is the numeraire export good, and c_i^h is the consumption of good $i = 1, \dots, N$. Maximizing utility subject to the budget constraint gives the per capita consumption $d_i(p_i)$ of each good, $i = 1, \dots, N$, with remaining income spent on the numeraire good, $c_0^h = I^h - p'd(p)$. We let $d(p) = [d_1(p_1), \dots, d_N(p_N)]$ denote the vector of per capita consumptions, depending on prices $p = (p_1, \dots, p_N)$. Then individual utility is similar to that in (10.1):

$$V(p, I^h) \equiv I^h - p'd(p) + \sum_{i=1}^N u_i[d_i(p_i)]. \quad (10.7)$$

Notice that the last two terms on the right of (10.7) give per capita consumer surplus, or $S(p) \equiv \sum_{i=1}^N u_i[d_i(p_i)] - p'd(p)$, so that consumer welfare can be written as

$$V^h(p, I^h) \equiv I^h + S(p), \quad (10.7')$$

with $\partial S(p)/\partial p_i = -d_i(p_i)$, by Roy's Identity.

On the production side, each of the N industries has the production function $y_i = f_i(L_i, K_i)$, where capital K_i is specific to each sector. The numeraire commodity is produced with one unit of labor, so wages are fixed at unity. Given the product price p_i in each sector, the return to the specific factor in that sector is

$$\pi_i(p_i) = \max_{L_i} [p_i f_i(L_i, K_i) - L_i]. \quad (10.8)$$

From (10.8) we can determine the optimal outputs $\pi'_i(p_i) = y_i(p_i)$ in each industry. The international prices of the goods are fixed at p_i^* , and each industry may receive a specific trade policy of t_i , $i = 1, \dots, N$, where $t_i > (<) 0$ indicates a tariff (subsidy) in an import industry, and a subsidy (tariff) in an export industry. Imports of each good are

² Earlier political economy models, such as Findlay and Wellisz (1982), simply assume a functional relationship between lobbying contributions and tariffs. Hillman (1989) allows for a government objective function with a general trade-off between the benefits to a special interest group and costs to consumers of tariff protection. The advantage of Grossman and Helpman (1994) is that these functional relationships are endogenously determined rather than assumed.

then $m_i(p_i) = d_i(p_i)L - y_i(p_i)$ (which are negative for exports), and tariff revenue collected is $T(p) = \sum_{i=1}^N (p_i - p_i^*)m_i(p_i)$. We assume that this revenue is redistributed by a poll subsidy of (T/L) per person.

The specific factor in each industry i is owned by H_i members of the population, so that $H = \sum_{i=1}^N H_i$ is the total number of persons owning some capital. For simplicity, we suppose that every individual also owns one unit of labor. The total population is L , so there are an additional $(L - H) \geq 0$ persons who own one unit of labor but no capital. The owners of specific capital in each industry earn the return $\pi_i(p_i)$ from (10.8), and obtain their wages of unity plus consumer surplus, along with redistributed tariff revenue. Summing these various terms, the owners of specific capital in industry i earn

$$W_i(p) = \pi_i(p_i) + H_i[1 + S(p)] + (H_i/L)T(p), \quad i = 1, \dots, N. \quad (10.9a)$$

The remaining $(L - H)$ persons obtain their wages plus consumer surplus and redistributed tariff revenue, so their welfare is

$$W_0(p) = (L - H)[1 + S(p)] + [(L - H)/L]T(p). \quad (10.9b)$$

Summing (10.9) over all workers and industries, we obtain total welfare:

$$W(p) = \sum_{i=0}^N W_i(p) = \sum_{i=1}^N \pi_i(p_i) + L[1 + S(p)] + T(p). \quad (10.10)$$

We suppose that a subset of the industries $j \in J_o$ is organized into lobbies, while the complementary set $j \in J_u$ is unorganized industries, with $J_o \cup J_u = \{1, \dots, N\}$. The purpose of each lobby is to provide contributions to the government in return for influencing the tariff/subsidy schedule. Specifically, they announce a *campaign contribution* schedule $R_j(p)$ that they are willing to pay, depending on the vector of prices $p_i = p_i^* + t_i$ prevailing across the industries, $i = 1, \dots, N$. The government values campaign contributions, but also weighs these against the consumer welfare of all individuals. Giving social welfare the weight of $\alpha > 0$, Grossman and Helpman assume that the government chooses tariffs and subsidies t_i to maximize

$$G(p) = \sum_{j \in J_o} R_j(p) + \alpha W(p). \quad (10.11)$$

The key question is how the lobbies in industries $j \in J_o$ determine their campaign contributions. The answer to this comes from the work of Bernheim and Whinston (1986). They argue that in the Nash equilibrium of the game we have described—with each lobby optimally choosing its contribution schedule $R_j(p)$, taking as given the schedules of the other groups, and knowing that the tariffs will be chosen to maximize (10.11)—then the lobbies can do no better than to select a contribution schedule of the form

$$R_j(p) = \max \{0, W_j(p) - B_j\}, \quad j \in J_o \quad (10.12)$$

where B_j is a constant.³ Bernheim and Whinston refer to this as a *truthful contribution schedule*, since it reflects the true welfare levels $W_j(p)$ obtained by the lobby for

³ Notice that by using the truthful contribution schedules in (10.12), the welfare of each lobby *net* of the contributions becomes $W_j(p) - R_j(p) = \min \{W_j(p), B_j\}$, so B_j is an upper bound on *net* welfare. Grossman and Helpman (1994, 843–47) discuss how B_j might be determined, as does Mitra (1999).

various tariffs. They argue that a *truthful Nash equilibrium*, where each lobby uses a schedule like (10.12), is included among the equilibria of the game.

Accepting this result of Bernheim and Whinston (1986), we can substitute (10.12) into (10.11) to obtain

$$G(p) = \sum_{j \in J_o} [(1+\alpha) W_j(p) - B_j] + \sum_{j \notin J_o} \alpha W_j(p), \quad (10.11')$$

where the summation over $j \notin J_o$ also includes the welfare W_0 of workers with no capital. Notice that this statement of the government's objective function indicates that it gives *differential weights* to the welfare of organized and unorganized industries: the organized lobbies have the weight $(1+\alpha)$, whereas other industries plus workers have the weight α .

Choosing the tariffs t_j to maximize (10.11') is equivalent to choosing the domestic prices p_j , $j \in J_o$ to maximize it. Before computing the first-order condition, we can differentiate welfare for an organized industry, unorganized industry, and workers as

$$\frac{\partial W_j}{\partial p_j} = y_j - H_j d_j(p) + \left(\frac{H_j}{L} \right) \left[m_j + (p_j - p_j^*) \frac{dm_j}{dp_j} \right], \quad \text{for } j \in J_o, \quad (10.13a)$$

$$\frac{\partial W_i}{\partial p_j} = -H_i d_j(p) + \left(\frac{H_i}{L} \right) \left[m_j + (p_j - p_j^*) \frac{dm_j}{dp_j} \right], \quad \text{for } i \in J_u, \quad (10.13b)$$

$$\frac{\partial W_0}{\partial p_j} = -(L - H) d_j(p) + \left(\frac{L - H}{L} \right) \left[m_j + (p_j - p_j^*) \frac{dm_j}{dp_j} \right], \quad (10.13c)$$

where $m_j = d_j(p_j)L - y_j(p_j)$ is the imports of good j (which is negative for exports).

Then multiply (10.13a) by $(1+\alpha)$, and (10.13b) and (10.13c) by α , and sum these over all organized and unorganized industries to obtain

$$\frac{\partial G}{\partial p_j} = (1+\alpha)y_j - \left(\sum_{i \in J_o} H_i \right) d_j(p) - \alpha L d_j(p) + (\lambda_o + \alpha) \left[m_j + (p_j - p_j^*) \frac{dm_j}{dp_j} \right], \quad j \in J_o, \quad (10.14)$$

where $\lambda_o \equiv \sum_{j \in J_o} (H_j/L)$ denotes the fraction of the population owning a specific factor in an organized industry, and this first-order condition holds for an *organized* industry j . For an industry that is *not organized* into a lobby, the term $(1+\alpha)y_j$ that appears first on the right side of (10.14) would be replaced by αy_j , since the unorganized industry receives the weight α rather than $(1+\alpha)$ in the government's objective function.

We can make use of the definition of imports, $m_j = d_j(p_j)L - y_j$, to simplify (10.14) as

$$\frac{dG}{dp_j} = y_j(1 - \lambda_o) + (\alpha + \lambda_o)(p_j - p_j^*) \frac{dm_j}{dp_j}, \quad j \in J_o \quad (10.14')$$

while for an industry *without* a lobby, the term $y_j(1 - \lambda_o)$ on the right is replaced by $-y_j \lambda_o$. Setting (10.14) and the modified condition for the unorganized industry equal to zero, we solve for the equilibrium tariffs $t_j = (p_j - p_j^*)$, $j = 1, \dots, N$, as

$$\frac{t_j}{p_j} = - \left(\frac{\delta_j - \lambda_o}{\alpha + \lambda_o} \right) \left(\frac{y_j}{m_j} \right) \left(\frac{\partial m_j}{\partial p_j} \frac{p_j}{m_j} \right)^{-1}, \quad \text{where } \delta_j = \begin{cases} 1 & \text{for } j \in J_o \\ 0 & \text{otherwise} \end{cases}. \quad (10.15)$$

This simple equation linking the tariffs/subsidies to underlying determinants is the key prediction of the “protection for sale” model of Grossman and Helpman (1994). To interpret it, note that the import elasticity appearing in (10.15) is negative, while δ_j is an indicator variable that equals unity for organized industries j , and zero otherwise. Recall that $\lambda_o \equiv \sum_{j \in J_o} (H_j/L)$ equals the fraction of the population owning a specific factor in an organized industry. Then by inspection, for $0 < \lambda_o < 1$ the tariffs in (10.15) are negative (i.e., import subsidies or export taxes) for unorganized industries, but are positive (i.e., import tariffs or export subsidies) for industries organized into a lobby. On the other hand, if either $\lambda_o = 0$ or $\lambda_o = 1$ (i.e., no one or everyone belongs to a lobby), then the tariffs in (10.15) are zero, so free trade is the political optimum.

The result that unorganized industries receive import subsidies or export taxes serves as a way to lower their domestic prices and therefore benefit consumers. The fact that these instruments are seldom observed in reality may reflect political opposition to them, or it stems from some other reason. We do not take this to be a refutation of the “protection for sale” model, any more than we interpret the fact that import subsidies are not observed in capital-scarce countries as a refutation of the median voter model. Instead, to test the “protection for sale” model we look for other correlations implied by (10.15). Notice that the *magnitude* of the tariffs or subsidies depends on the ratio of production to imports (y_j/m_j), and also on the inverse of the import demand elasticity. Having higher domestic production relative to imports will lead to *higher* import tariffs or export subsidies for organized industries (since then $\delta_j - \lambda_o = 1 - \lambda_o > 0$), but to *lower* import tariffs or export subsidies for unorganized industries (where $\delta_j - \lambda_o = -\lambda_o < 0$). These are the key predictions that will be tested.

Goldberg and Maggi (1999) denote the import demand elasticity (measured as a positive number) by e_j , and multiply (10.15) by this magnitude to obtain

$$e_j \left(\frac{t_j}{p_j} \right) = \beta_0 + \beta_1 \left(\frac{y_j}{m_j} \right) + \beta_2 \delta_j \left(\frac{y_j}{m_j} \right) + \varepsilon_j, \quad (10.16)$$

where $\beta_1 = -[\lambda_o/(\alpha + \lambda_o)]$, $\beta_2 = 1/(\alpha + \lambda)$, and β_0 is a constant term.⁴ Notice that $\beta_1 < 0$ indicates how the trade barriers vary with the output/import ratio in *any* industry, while $\beta_2 > 0$ reflects the *additional* impact of having the industry organized as a lobby. So a change in the output/import ratio in an unorganized industry affects the trade barrier by $\beta_1 < 0$, and in an organized industry affects the trade barrier by $(\beta_1 + \beta_2) > 0$.

We multiply through by the import demand elasticity because that variable is measured with error, which is therefore incorporated into the error term ε_j of (10.16). Data on e_j is taken from the compendium of import elasticities in Shiells, Stern, and Deardorff (1986). For the trade barriers (t_j/p_j), Goldberg and Maggi use *nontariff* barriers in the United States. The reason for using nontariff barriers rather than tariffs is that the latter have been reduced by international agreements, so we might expect their level to be lower than that predicted by (10.16). Nontariff barriers are measured by the “coverage ratio” in each industry, that is, the fraction of disaggregate products in each industry covered by quotas or some other restriction on trade. The independent variables in (10.16) are the indicator variable δ_j for whether an industry is organized or not, and the output/import ratio. Goldberg and Maggi measure the former using

⁴ Notice that a constant term in the tariff equation is not implied by the theory in (10.15), but it is generally a good idea to include it in empirical work. It turns out that β_0 is insignificantly different from zero in the estimates of Goldberg and Maggi (1999).

a threshold of contributions by industries to the 1981–82 congressional elections in the United States, using data from Gawande (1995). The output/import ratio is treated as endogenous (since changes in trade barriers affect both output and imports), and estimation is performed using the instruments from Trefler (1993b). With this data, (10.16) is estimated over a cross-section of 107 U.S. industries.

In their results, Goldberg and Maggi find estimates of β_1 and β_2 of -0.0093 (0.0040) and 0.0106 (0.0053), respectively (with standard errors in parentheses). Both of these coefficients have the expected sign, with $(\beta_1 + \beta_2) = 0.0013 > 0$ as predicted, though the latter estimate is not significantly different from zero. Recalling that $\beta_1 = -[\lambda_o / (\alpha + \lambda_o)]$, and $\beta_2 = 1 / (\alpha + \lambda_o)$, we can use these estimates to recover $\alpha = 93$ and $\lambda_o = 0.88$. In alternative estimates, they obtain $\alpha = 53$ and $\lambda_o = 0.83$. Regardless of which we use, it is evident that the weight α on consumer welfare in the government's objective function is very high: between fifty and one hundred times higher than the weight given to political contributions.

Gawande and Bandyopadhyay (2000) extend the Grossman-Helpman model to allow for trade in intermediate inputs. Final-goods industries that use the intermediate input suffer from a tariff on it, and so the predicted tariffs in those final-goods industries are higher. In other words, controls are added to the right side of (10.16) to reflect the tariffs on intermediate inputs, but otherwise the estimating equation is similar, with $\beta_1 < 0$, $\beta_2 > 0$ and $(\beta_1 + \beta_2) > 0$ expected. Gawande and Bandyopadhyay use an expanded dataset of 242 U.S. industries, and again use nontariff barriers as the dependent variable. In their results, they estimate β_1 and β_2 of -0.000309 (0.00015) and 0.000315 (0.00016), respectively. Both of the coefficients have the expected sign, and again we find that $(\beta_1 + \beta_2) = 0.000006 > 0$, though this coefficient is not significant. Using these estimates, we recover the values of $\alpha = 3,175$ and $\lambda_o = 0.98$, so Gawande and Bandyopadhyay find even a higher weight on consumer welfare than do Goldberg and Maggi! In alternative estimates, they obtain $\alpha = 1,750$ and $\lambda_o = 0.95$, which is still a remarkably high weight on consumer welfare.

There have been a number of other empirical applications of the Grossman-Helpman model, including McCalman (2004), who applies it to Australia, Grether, de Melo, and Olarreaga (2002), who apply it to Mexico, and Mitra, Thomakos, and Ulubasoglu (2002), who apply it to Turkey.⁵ Rather than describe these applications, we shall explore other theoretical extensions of the “protection for sale” model.

ENDOGENOUS LOBBIES

In the description of Grossman and Helpman's model above, we treated the existence of the lobbies as exogenous. This is a limitation, of course, and in reality we would expect lobbies to form when their potential returns are sufficiently high or the costs of organizing are sufficiently low. By allowing for the endogenous formation of lobby groups, following Mitra (1999), we will obtain some important additional insights from the “protection for sale” model.

Recall that $\lambda_o \equiv \sum_{j \in J_o} (H_j / L)$ denotes the fraction of the total population belonging to organized industries. This can be rewritten as $\lambda_o = (H/L)[\sum_{j \in J_o} (H_j / H)]$, and decomposed into two terms: (H/L) is the fraction of the population that owns some specific

⁵ See also the survey by Gawande and Krishna (2003).

factor, and $[\sum_{j \in J_o} (H_j/H)]$ is the fraction of specific-factor owners who are organized. Let us impose some symmetry on the model, so that H_j is the same across all industries j , and also capital K_j and the production functions f_j are the same across all $j = 1, \dots, N$. Then denoting the number of organized industries by $N_o \leq N$, it is immediate that $[\sum_{j \in J_o} (H_j/H)] = (N_o/N)$, which measures the fraction of industries that are organized. So we can rewrite the term λ_o as $\lambda_o = \lambda_k n_o$, where $\lambda_k \equiv (H/L)$ is the fraction of the population that owns some specific factor (i.e., capital), and $n_o \equiv (N_o/N) = [\sum_{j \in J_o} (H_j/H)]$ is the fraction of industries that are organized.

Substituting $\lambda_o = \lambda_k n_o$ into (10.16), the predicted trade policies are

$$\frac{t_j}{p_j} = -\left(\frac{\delta_j - \lambda_k n_o}{\alpha + \lambda_k n_o}\right)\left(\frac{y_j}{m_j}\right)\left(\frac{\partial m_j}{\partial p_j} \frac{p_j}{m_j}\right)^{-1}, \quad \text{where } \delta_j = \begin{cases} 1 & \text{for } j \in J_o \\ 0 & \text{otherwise} \end{cases}. \quad (10.17)$$

In the Grossman-Helpman model, a *rise in either λ_k or n_o will lower (t_j/p_j)* across all industries in (10.17). That is, holding the import demand elasticity and output/import ratio constant, we have

$$\frac{d(t_j/p_j)}{d(\lambda_k n_o)} = \frac{(\alpha + \delta_j)}{(\alpha + \lambda_k n_o)^2} \left(\frac{y_j}{m_j}\right)\left(\frac{\partial m_j}{\partial p_j} \frac{p_j}{m_j}\right)^{-1} < 0, \quad (10.18)$$

where the sign of (10.18) follows because the import demand elasticity is negative (or for exports, this supply elasticity is positive but (y_j/m_j) is negative). Continuing to hold this elasticity constant, if we also allow the output/import ratio to respond, then this would *reinforce* the negative sign on (10.18): the rise in either λ_k or n_o will lower (t_j/p_j) as in (10.18), which lowers the domestic prices $p_j = p_j^* + t_j$, and further lowers the output/import ratio (y_j/m_j) .

Now let us work this thought experiment in reverse. A *decline* in λ_k corresponds to fewer people owning the stock of capital, which is a more unequal income distribution. This would *raise* (t_j/p_j) across all industries as in (10.18) and also raise all domestic prices, meaning that import tariffs and export subsidies rise in organized industries, while import subsidies and export taxes are reduced in unorganized import industries. It can be argued that this has a *greater* beneficial impact on the organized industries receiving protection than on the unorganized industries.⁶ That would create an incentive for new groups to enter, so if n_o is treated as endogenous, then it would rise. But that would *offset* the initial decline in λ_k , since the increase in n_o would lower (t_j/p_j) in (10.17). So with the number of lobbies treated as endogenous, it is no longer clear whether a more unequal income distribution leads to more or less protection. That is the issue that Mitra (1999) aims to resolve.

To model the entry of groups, we impose a high degree of symmetry on the model. We already assumed K_i and H_i do not vary across industries, and further assume that demand and production functions are symmetric across products, as are international prices. Then the tariff/subsidies computed from (10.16) will be t_o/p_o for all organized industries, and t_u/p_u for all unorganized industries. These policies depend on the *product* $n_o \lambda_k$ from (10.17), reflecting on the fraction of organized industries (which will be endogenous), and the fraction of population that owns capital (which is exogenous). These policies fully determine the welfare $W_o(n_o \lambda_k)$ to capital owners in organized

⁶ See problems 10.1–10.3.

industries, and welfare $W_u(n_o\lambda_k)$ to capital owners in unorganized industries, as in (10.9a). The *gross benefits* to forming a lobby are then

$$\Delta W(n_o\lambda_k) \equiv W_o(n_o\lambda_k) - W_u(n_o\lambda_k), \quad (10.19)$$

while the *net benefits* are obtained by subtracting the cost of political contributions:

$$NB(n_o\lambda_k) \equiv \Delta W(n_o\lambda_k) - R_o(n_o\lambda_k). \quad (10.20)$$

Following Mitra (1999), some properties of these schedules can be derived. Given the symmetry we have assumed, the gross benefits in (10.19) are simply the difference between the returns to the fixed factors in organized and unorganized industries: $\Delta W(n_o\lambda_k) = \pi(p_o) - \pi(p_u) > 0$. We argued above that as λ_k declines, leading to a rise in the tariff/subsidies as in (10.18), a greater benefit on the organized than unorganized industries is created, so that $\pi(p_o) - \pi(p_u)$ increases and gross benefits rise.⁷ Thus, $\Delta W'(n_o\lambda_k) < 0$.

For net benefits, Mitra investigates the determinants of the truthful contribution schedules $R_o(n_o\lambda_k)$, including solving for the lower-bounds B_j to organized-industry welfare in (10.12). He is able to show that even when the impact of the number of lobbies n_o to contributions $R_o(n_o\lambda_k)$ is taken into account, it is still the case that net benefits rise if the number of lobbying groups falls. Thus, $NB'(n_o\lambda_k) < 0$. The declining net benefits as a function of the number of organized industries is shown in figures 10.1 and 10.2.

What about the costs of forming a lobby? In addition to the political contributions, there may be some significant costs of grass-roots organizing and communication among members so as to overcome the free-rider problem. Let us denote these costs (measured in terms of the numeraire good) by $C(n_o)$, where we assume that $C'(n_o) \geq 0$, so the cost of creating a new lobby is nondecreasing in the number of lobbies already there. This cost schedule is also illustrated in figures 10.1 and 10.2. Then we suppose that lobbies form up to the point where net benefits just equal the costs of creating a new lobby, which is illustrated by the lobbies n_o in both figures.

Now consider the effect of a fall in λ_k , which is a worsening of the income distribution. The impact effect from (10.17) (holding n_o constant) would be an increase in (t_j/p_j) and higher domestic prices for all industries. This creates additional gains to organizing an industry, and in figures 10.1 and 10.2, the net benefits schedule rises to NB' . Equivalently, net benefits remain the same if along with the decline in λ_k there is an equiproportional increase in n_o , so that the net benefit curves shifting rightward by the amount $-d\lambda_k/\lambda_k$.

In figure 10.1, where the cost of organizing lobbies is increasing, the induced increase in the number of lobbies from n_o to n_o' is clearly less than the rightward shift of the NB schedule. Therefore, $dn_o/n_o < -d\lambda_k/\lambda_k$, and it follows that $dn_o/n_o + d\lambda_k/\lambda_k < 0$, so that $n_o\lambda_k$ falls in equilibrium. This leads to an increase in the trade policies (t_j/p_j) from (10.17), so that organized industries receive higher import tariffs or export subsidies, and unorganized industries have reduced import subsidies or export taxes. Overall, there is a rise in domestic prices, and in this sense the net level of protection has increased due to the more concentrated income distribution and increased lobbying.

⁷ See problems 10.1–10.3.

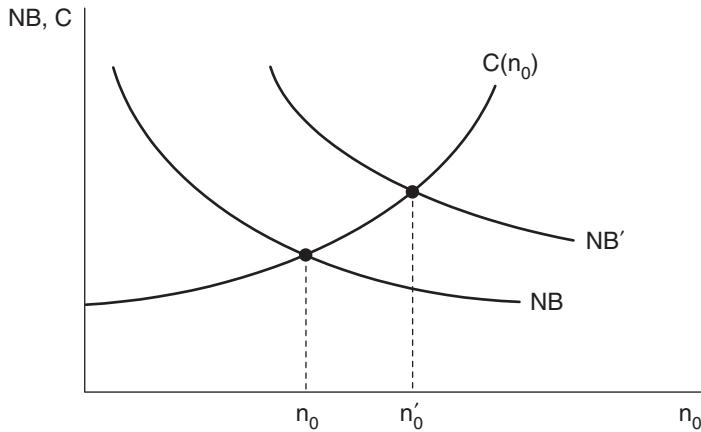


Figure 10.1

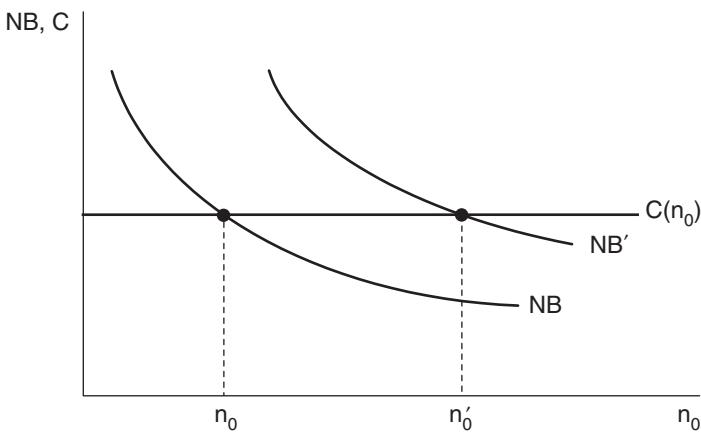


Figure 10.2

In figure 10.2, by contrast, the cost of organizing lobbies is constant. In that case the increase in the number of lobbies from n_o to n'_o is just equal to the rightward shift of the NB schedule, so that $dn_o/n_o = -d\lambda_k/\lambda_k$ and $n_o\lambda_k$ is *unchanged* in equilibrium. Therefore the trade policies from (10.17) are also unchanged. We see that the overall impact of income inequality on protection is very sensitive to the structure of lobbying costs $C(n_o)$, which could reflect a wide range of legal and political features in a country. It would be difficult, then, to predict the effects of changes in income distribution on protection either within or across countries. This finding makes it all the more remarkable that in the median voter model, discussed earlier, Dutt and Mitra (2002) find a systematic (though non-monotonic) relationship between trade barriers and inequality. Obtaining such time-series or cross-country empirical results from the “protection for sale” model would be more difficult. That model is ideally suited, however, to explaining *cross-industry* trade protection, as we have already discussed.

TWO-COUNTRY MODEL

In the median voter model and the “protection for sale” model of Grossman and Helpman (1994), the government treats the international price p^* as fixed. But what if instead the country is large, so that its tariffs affect the terms of trade? Obtaining a terms of trade gain creates an additional reason to use tariffs. In a two-country model, both countries would have this incentive, and we could conjecture that they would both end up with tariffs higher than those we solved for above. Bagwell and Staiger (1999, 2002) argue that this creates an important role for international institutions such as the GATT/WTO: to offset or eliminate the incentive to manipulate the terms of trade.

To formalize this idea, we return to the median voter or the “protection for sale” model. For simplicity, suppose that there are only two goods: the additively separable numeraire good that is exported from the home country, and a second good with demand $d(p)$ at home and supply $y(p)$. Home imports are then $m(p) = d(p) - y(p)$. Likewise, the foreign country has demand $d^*(p^*)$ for this good, and supply $y^*(p^*)$, depending on its own price p^* . Foreign exports are denoted by $x^*(p^*) = y^*(p^*) - d^*(p^*)$.

Home and foreign prices differ due to tariffs in both countries. Suppose that the home country applies an ad valorem tariff, and let τ equal *one plus the ad valorem tariff*. Then the home prices are $p = p^w\tau$, where p^w are the world prices. The foreign country applies an ad valorem tariff on its own imports, and let τ^* equal *one plus the foreign ad valorem tariff*. Since p^w is the world relative price of the home import and foreign export, then $1/p^w$ is the world relative price of the foreign import. Applying the tariff of τ^* means that the foreign price of its import good is τ^*/p^w , so the relative price of the foreign export is p^w/τ^* .

Using these prices, market clearing means that home imports equal foreign exports:

$$m(p^w\tau) = x^*(p^w/\tau^*) \Rightarrow p^w(\tau, \tau^*). \quad (10.21)$$

This equation determines the world equilibrium price $p^w(\tau, \tau^*)$, depending on the tariff in each country. Under the standard assumptions on the import demand and export supply curves, it is readily confirmed that

$$\frac{dp^w}{d\tau} < 0 < \frac{dp}{d\tau}, \text{ and } \frac{dp^w}{d\tau^*} > 0 > \frac{dp^*}{d\tau^*}. \quad (10.22)$$

In other words, the home tariff *lowers* the world price of imports for the home country and raises the domestic price, while the foreign tariff *raises* the world price of its exports and lowers its local price of the export good.⁸

Let us now consider the government objective function in each country. In the median voter model the objective function was welfare of the median voter $V(p, I^m)$, where from (10.2') income of the median voter was $I^m = [(\rho^m - 1)r(p)K + y_0(p) + py(p) + T]^9$. Using the ad valorem tariffs defined above, tariff revenue in this expression is $T = (\tau - 1)p^w[d(p) - y(p)]$. Because domestic prices are $p = p^w\tau$, we can substitute this into tariff revenue and income and obtain an expression for income that depends on the world price and domestic tariff, $I^m(p^w, \tau)$. It follows that the objective of the home government is

⁸ You are asked to demonstrate (10.22) in problem 10.4.

⁹ We now normalize the size of the population at $L = 1$ to simplify notation.

$$G(p^w, \tau) = V[p^w\tau, I^m(p^w, \tau)]. \quad (10.23)$$

Thus, the objective function of the home government can be written as a function of the world prices and the tariff. Bagwell and Staiger argue that this formulation of the objective function is general enough to encompass a number of models, including the median voter model and the “protection for sale” model of Grossman and Helpman.¹⁰

Furthermore, it is readily verified that the tariffs in the median voter model (equation 10.4) or in the “protection for sale” model (equation 10.15), satisfy the first-order condition

$$G_\tau(p^w, \tau) = 0. \quad (10.24a)$$

That is, these tariffs are obtained by maximizing the government’s objective while treating the world price p^w as *fixed*. For the foreign country there will be an analogous objective function $G^*(p^w, \tau^*)$, with first-order condition for the tariff

$$G_{\tau^*}(p^w, \tau^*) = 0. \quad (10.24b)$$

Bagwell and Staiger refer to the tariffs satisfying (10.24) as *politically optimal*.

However, the tariffs satisfying (10.24) will generally *not be* chosen by a government that recognizes its ability to influence the terms of trade. Instead, the tariffs would be chosen in the two countries to achieve

$$\frac{d}{d\tau} G[p^w(\tau, \tau^*), \tau] = G_p \frac{\partial p^w}{\partial \tau} + G_\tau = 0, \quad (10.25a)$$

and in the foreign country,

$$\frac{d}{d\tau} G^*[p^w(\tau, \tau^*), \tau^*] = G_p^* \frac{\partial p^w}{\partial \tau^*} + G_{\tau^*}^* = 0. \quad (10.25b)$$

Notice the obvious difference between the politically optimal tariffs in (10.24), and the first-order conditions (10.25) that incorporate the terms of trade effects. Bagwell and Staiger refer to the tariffs satisfying (10.25) as the *Nash equilibrium* tariffs, since they are the best response by each country given the tariff choice of the other.

The question then arises as to how the politically optimal tariffs (which ignore the terms of trade effects) compare with the Nash equilibrium tariffs (which incorporate these effects). To make this comparison, it is convenient to define the *reduced form* government objective functions as

$$\tilde{G}(\tau, \tau^*) \equiv G[p^w(\tau, \tau^*), \tau], \quad (10.26a)$$

and in the foreign country,

$$\tilde{G}^*(\tau, \tau^*) \equiv G^*[p^w(\tau, \tau^*), \tau^*]. \quad (10.26b)$$

¹⁰ Notice that G in (10.11) is a function of domestic prices p , but also depends on international prices p^* which we did not make explicit. Then $G(p, p^*)$ can be rewritten as $G(\tau p^*, p^*)$, so that G again depends on p^* and τ .

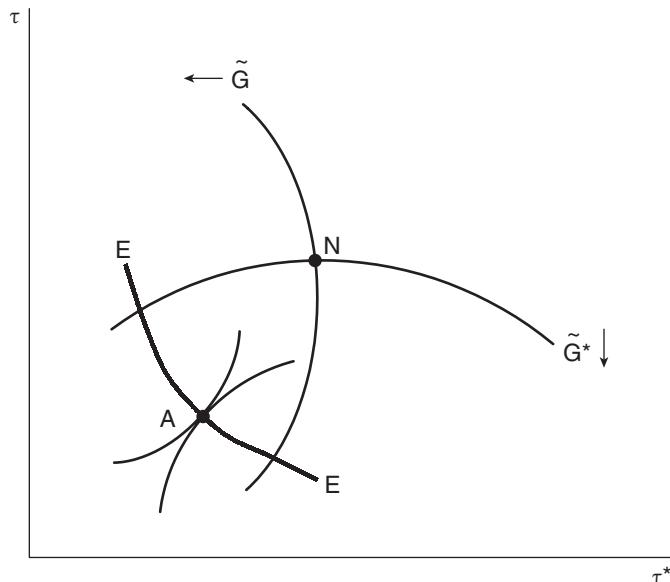


Figure 10.3

These give the objectives as functions of the tariffs in the two countries. Notice that the Nash equilibrium in (10.25a) can be equivalently written as $\tilde{G}_\tau = 0$, which defines the best home response $\tau = r(\tau^*)$ to the foreign tariff. Similarly, the Nash equilibrium in (10.25b) can be written as, $\tilde{G}_{\tau^*}^* = 0$, which defines the best foreign response $\tau^* = r^*(\tau)$ to the home tariff.

The Nash equilibrium is defined by the intersection of these reaction curves $\tau = r(\tau^*)$ and $\tau^* = r^*(\tau)$. Assuming that this intersection is unique, it is illustrated by point N in figure 10.3. Since this is the maximum of $\tilde{G}(\tau, \tau^*)$ subject to a given τ^* , the iso-curve of $\tilde{G}(\tau, \tau^*)$ has slope infinity at point N , and is increasing in the leftward direction, as illustrated. Similarly, the iso-curve of $\tilde{G}^*(\tau, \tau^*)$ has a slope of zero at point N , and is increasing in the downward direction, as illustrated. It is evident that there is a region *below and to the left* of point N , whose boundaries are defined by the iso-curves of $\tilde{G}(\tau, \tau^*)$ and $\tilde{G}^*(\tau, \tau^*)$, where *both countries are better off than at the Nash equilibrium*. Through this region, there lies the efficiency locus EE on which the iso-curves of $\tilde{G}(\tau, \tau^*)$ and $\tilde{G}^*(\tau, \tau^*)$ are tangent.

We have shown, therefore, that the Nash equilibrium does not lie on the efficiency frontier. This is the first half of the result given in the following theorem.

THEOREM (BAGWELL AND STAIGER 1999; GROSSMAN AND HELPMAN 1995A)

- (a) The Nash equilibrium is not efficient.
- (b) The political optimum is efficient.

Grossman and Helpman (1995a) demonstrate part (a) by explicitly calculating the optimal tariffs for a *large* country in the “protection for sale” model. Rather than

equation (10.15), which corresponds to the politically optimal tariffs $G_\tau(p^w, \tau) = 0$, they instead find that the tariffs are as in (10.15) *plus* the inverse elasticity of foreign export supply. Thus, the optimal tariffs are increased by exactly the same terms of trade effect that we discussed in chapter 8, and that is included in (10.25) by the terms \tilde{G}_τ and \tilde{G}^*_τ .

To demonstrate part (b) of this theorem, let us calculate the condition for tangency of the iso-curves of $\tilde{G}(\tau, \tau^*)$ and $\tilde{G}^*(\tau, \tau^*)$. This is $\tilde{G}_\tau/\tilde{G}_{\tau^*} = \tilde{G}^*_\tau/\tilde{G}^*_{\tau^*}$. Using the definition of these reduced-form objective functions in (10.26), we can rewrite this tangency condition as

$$\frac{G_\tau + G_p \partial p^w / \partial \tau}{G_p \partial p^w / \partial \tau^*} = \frac{G_p^* \partial p^w / \partial \tau}{G_{\tau^*}^* + G_p^* \partial p^w / \partial \tau^*}. \quad (10.27)$$

Now substitute in the condition for the politically optimal tariffs, $G_\tau = G_{\tau^*}^* = 0$. Then (10.27) simply becomes $(\partial p^w / \partial \tau) / (\partial p^w / \partial \tau^*) = (\partial p^w / \partial \tau) / (\partial p^w / \partial \tau^*)$, which is obviously satisfied. Therefore, the political optimum lies on the efficiency locus, as illustrated by point *A* in figure 10.3. It follows that the Nash equilibrium *N* is worse for both countries than the political optimum at point *A*.¹¹

This raises the question of whether it is possible to improve on the Nash equilibrium by some simple rules agreed upon by both countries. Bagwell and Staiger (1999, 2002) argue that this creates an important role for the GATT, and in particular, justifies the GATT principles of reciprocity and nondiscrimination. As they describe it: “the principle of reciprocity is a GATT norm under which one country agrees to reduce its level of protection in return for a reciprocal ‘concession’ from its trading partner. . . . The principle of nondiscrimination is a separate norm, under which a member government agrees that any tariff on a given product applied to the import of one trading partner applies equally to all other trading partners” (1999, 216–17). We will examine the first of these principles.

How is reciprocity applied in practice? Bagwell and Staiger suggest the following definition in our two-country model: the tariffs τ and τ^* are reduced so that *import levels in both countries rise by the same amount*. Since these imports occur in different goods, they must be compared at some prices, so let us use the *initial* world prices p^{w0} . Then the tariff reductions are reciprocal if $p^{w0}\Delta m = \Delta m^*$, where m^* is foreign imports of the numeraire good. But from trade balance, the value of imports equals the value of exports abroad, so that $\Delta m^* = m^{*1} - m^{*0} = p^{w1}x^{*1} - p^{w0}x^{*0}$. But we also know that foreign exports equal home imports, so that $x^{*1} = m^1$ and $x^{*0} = m^0$. Therefore, Bagwell and Staiger’s definition of reciprocity implies

$$\underbrace{p^{w0}\Delta m}_{\Delta \text{Home imports}} = \underbrace{p^{w0}m^1 - p^{w0}m^0}_{\Delta \text{Foreign imports}} = \underbrace{\Delta m^*}_{p^{w1}m^1 - p^{w0}m^0} = p^{w1}m^1 - p^{w0}m^0.$$

By inspection, this equality is satisfied if and only if world prices are constant, $p^{w1} = p^{w0}$. So the implication of reciprocity in the two-country model is that *mutual tariff reductions should leave the world price unchanged*.

¹¹ We have not quite shown this last result, because it is possible that the countries are so asymmetric in size that the political optimum at point *A* lies on the efficiency locus *EE*, but *outside* the region bounded by the iso-curves of \tilde{G} and \tilde{G}^* in figure 10.3. We assume that this asymmetric case does not apply.

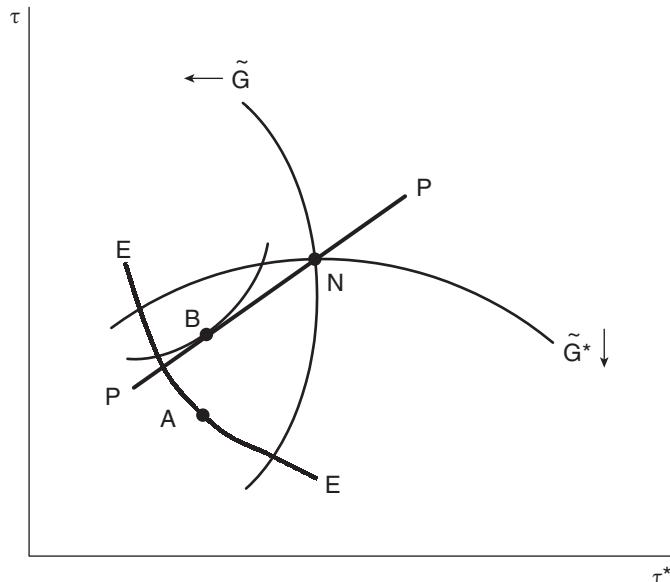


Figure 10.4

Totally differentiating the world price $p^w(\tau, \tau^*)$, it is unchanged when

$$\frac{dp^w}{d\tau} d\tau + \frac{dp^w}{d\tau^*} d\tau^* = 0 \Rightarrow \frac{d\tau}{d\tau^*} = -\frac{dp^w}{dp^w} / \frac{dp^w}{d\tau} > 0,$$

where the positive sign follows from the inequalities in (10.22). The locus of tariffs along which the world price $p^w(\tau, \tau^*)$ is held fixed at its Nash-equilibrium level is illustrated by PP in figure 10.4. Since this line has a positive slope, and the iso-curves of $\tilde{G}(\tau, \tau^*)$ and $\tilde{G}^*(\tau, \tau^*)$ have slopes of infinity and zero at point N , the PP locus clearly falls *inside* the region bounded by those iso-curve. Therefore, *reciprocal reductions in tariffs from the Nash equilibrium raise the government objective function for both countries*. This result provides strong justification for the GATT principle of reciprocity.

How far should the tariff reductions proceed? Bagwell and Staiger argue that if the two countries are symmetric in size, then the line PP will intersect the efficiency locus at precisely the political optimum at point A . That is, tariff reduction should proceed until both countries are applying tariffs as in the median voter or “protection for sale” model, and therefore avoiding the mutual losses associated with exploiting the terms of trade.

When countries are not symmetric in size, however, the situation is more complicated. As we move down the iso-price locus PP from point N in figure 10.4, the government objective function of both countries initially rises. It might be the case, though, that the objective of one country is maximized on PP before we get to the efficiency locus EE . This is illustrated in figure 10.4 by point B , where the government objective of the *home* country is maximized. At point B , both governments attain higher objectives than at the Nash equilibrium N , but further reciprocal reductions in tariffs

would lower the objectives for the home country. So in this case, the political optimum at point *A* is not reached. Bagwell and Staiger, as well as Grossman and Helpman (1995a), discuss various bargaining mechanisms that might enable the two countries to still agree on the political optimum at point *A*, or some other point on the efficiency locus. Bagwell and Staiger (1999, 2002) further discuss how many other principles of GATT—including nondiscrimination—can be understood as being efficient. For further details the reader is to refer to their work, which provides a very general economic justification for GATT rules.¹²

Home Market Effect

Ossa (2011) has developed an alternative framework to rationalize the principles of reciprocity and nondiscrimination in GATT. Instead of relying on terms of trade effects, as in the models of Bagwell and Staiger (1999, 2002), Ossa uses the *home market effect* that we studied in chapter 5. In that chapter, we saw how a larger home market (due to country size) would attract disproportionately more firms to locate in the differentiated sector at home and also lead to a fall in the home CES (constant elasticity of substitution) price index. That is, there would be a gain from locally produced product variety in larger markets. Ossa (2011) argues that such benefits give governments an incentive to protect the home market, which is a different argument for protection than the “terms of trade” effect, though it creates the same rationale for cooperation through GATT.

To demonstrate the incentives for protection and for cooperation, Ossa (2011) introduces *ad valorem* tariffs of τ^{ij} between countries *i* and *j*, in addition to iceberg trade costs of T^{ij} . Then recall from chapter 5 that with homogeneous firms and monopolistic competition, the zero-profit output of every firm is equal to $\bar{y} = (\sigma - 1)\alpha/\beta$, the home price of every firm is equal to $p \equiv p^{ij} = \beta\sigma/(\sigma - 1)$ and export price is equal to $p^{ij} = T^{ij}(1 + \tau^{ij})\beta\sigma/(\sigma - 1)$. From (5.25) a two-country version of CES price index in country *j* is

$$P^j = \frac{\sigma\beta}{(\sigma - 1)} \left(N^j + N^i ((1 + \tau^{ij}) T^{ij})^{1-\sigma} \right)^{1/(1-\sigma)}, i \neq j.$$

There is homogenous good sector in each country, which produces one unit of the good using one unit of labor. After normalizing the price of the homogeneous good at unity, the wage also becomes unity in both countries. It follows that welfare in country *j* is equal to the inverse of the CES price index, or

$$\frac{1}{P^j} = \frac{(\sigma - 1)}{\sigma\beta} \frac{1}{(N^j + N^i ((1 + \tau^{ij}) T^{ij})^{1-\sigma})^{1/(1-\sigma)}}, i \neq j. \quad (10.28)$$

With wages fixed at unity in each country, and with CES preferences so that markups are constant, there is no term of trade effect due to a change in the tariff or trade costs. But welfare is still affected by a change in the number of firms due to changing tariffs or trade costs.

¹² See also McCalman (2002) and Maggi (1999) on the role of the GATT, and the recent work of Antràs and Staiger (2012), Bagwell and Staiger (2011), and Maggi and Staiger (2011).

The revenue of every firm is equal to $r = p\bar{y}$ and is constant. Then the number of firms in every country can be obtained as expenditure on home goods divided by revenue of every firm. Total expenditures on the differentiated good in country j are equal to $\gamma L^j + X^{ji}$, where γ is the share of the differentiated good in consumption and X^{ji} equals the purchases by the foreign country.¹³ But consumers spend X^{ij} on imported goods, which must be subtracted from $\gamma L^j + X^{ji}$ to obtain purchases on the differentiated good produced *at home*. It follows that the number of goods produced in each country equals

$$N^j = \frac{\gamma L^j + TB^j}{p\bar{y}}, \quad (10.29)$$

where $TB^j \equiv X^{ji} - X^{ij}$ is trade balance in the differentiated goods sector in country j .

Similar to Bagwell and Staiger, Ossa defines reciprocal tariff change as a change in the tariffs that satisfy $dTB^j = dTB^i = 0$. It follows immediately from (10.29) that changes in the tariffs leave the number of firms unchanged in both countries if and only if they are reciprocal. It further follows from (10.28) that the reciprocal reductions in the *ad valorem* tariffs of τ^{ij} between countries will therefore raise welfare in both countries.

THEOREM (OSSA 2011)

Reciprocal trade liberalization monotonically increases welfare in both countries.

Ossa (2011) illustrates the magnitude of this effect through computations in an extended model, and argues that the model yields reasonable levels of noncooperative tariffs, as well as moderate gains from reciprocal tariff reductions, as under the GATT/WTO.

REGIONAL TRADE AGREEMENTS

Another foundation of GATT is the MFN principle of nondiscrimination. Recall that this principle is embodied in Article I (see table 7.1). This principle is violated, however, by Article XXIV, which allows for customs unions and free-trade areas in some circumstances. That raises the question of whether allowing for such regional trade agreements conflicts with the overall GATT goal of multilateral tariff reductions. Bhagwati (1993) has suggested that this question can be broken into two distinct issues: the “static impact effect” of regionalism, which is the impact of successive regional agreements (possibly leading toward global free trade) on world welfare; and the “dynamic time path,” which is the issue of whether having the option of a regional trade agreement will impact countries’ willingness to enter into a multilateral agreement. The first issue has been touched on at the end of chapter 7, drawing on the work

¹³ Notice that this expression for expenditure on the differentiated product *ignores* redistributed tariff revenue. Ossa (2011) does not consider such redistributed revenue in the theoretical exposition of the home market effect, as we present here, but does incorporate such revenue into the computations based on his model.

of Krugman (1991a,b) and Frankel (1997). We will examine here the second issue in the context of the median voter model, following Levy (1997).¹⁴

Levy proposes the following framework for thinking about sequential regional and multilateral agreements. In a multicountry world where the median voter in each country determines policy, suppose that a vote is taken initially on whether two countries should enter into bilateral free trade. For convenience, we treat the initial tariffs as prohibitive, and so these countries are voting on whether to move from autarky to free trade between them, while retaining prohibitive tariffs with other countries. In the next period, a vote is then taken in each country as to whether to join into free trade with a larger group of countries, in a multilateral agreement. At this second stage, either country in the initial bilateral agreement has *veto power* over whether that agreement can be extended to incorporate new countries.

We shall assume that multilateral free trade brings benefits to both countries as compared to their autarky positions. That is, the median voters in both countries would approve a movement from autarky to multilateral free trade. The question is whether having bilateral free trade (with the assumed veto power of either country) acts as a “stepping stone” or a “stumbling block” toward multilateral free trade, to use the terminology of Bhagwati. To answer this, we need to consider four conceivable voting paths:

- (1) The median voter in at least one country rejects the bilateral agreement, and then both countries agree to the multilateral agreement.
- (2) The median voters in both countries agree to the bilateral agreement and then agree to the multilateral agreement.
- (3) The median voters in both countries agree to the bilateral agreement and then both reject the multilateral agreement.
- (4) The median voters in both countries agree to the bilateral agreement, but then one country vetoes the multilateral deal, whereas the other country would approve multilateral free trade.

Paths (1) and (2) are both possibilities, but in those cases the bilateral arrangement does not act as a hindrance to multilateral free trade. The multilateral deal is rejected in paths (3) and (4), and so we need to examine whether these paths can actually occur. Case (4) can be *ruled out* by the assumption that the time spent in bilateral free trade is very short as compared to the time spent in multilateral free trade (if approved). This means that the median voter in the country that benefits from bilateral free trade, and then *also* benefits from multilateral trade, will anticipate that the *other* country will block the multilateral deal. Therefore, this median voter will refuse to approve the bilateral arrangement initially, correctly anticipating that multilateral free trade will still be achieved (since by assumption, this is better for both countries than autarky). By this argument, path (4) never occurs.

We are left, then, with only having to consider whether or not path (3) can occur: is it possible that the median voters in both countries are worse off going from bilateral to multilateral free trade, whereas they are better off going from autarky to multilateral free trade? We will demonstrate that this is *impossible* with a Heckscher-Ohlin (HO) production structure. After showing this, we discuss an extended model that incorporates monopolistic competition and product variety, under which path (3) can occur.

¹⁴ See also Grossman and Helpman (1995b) for a treatment of regional agreements in the “protection for sale” model, as well as the papers cited in the introduction to this chapter.

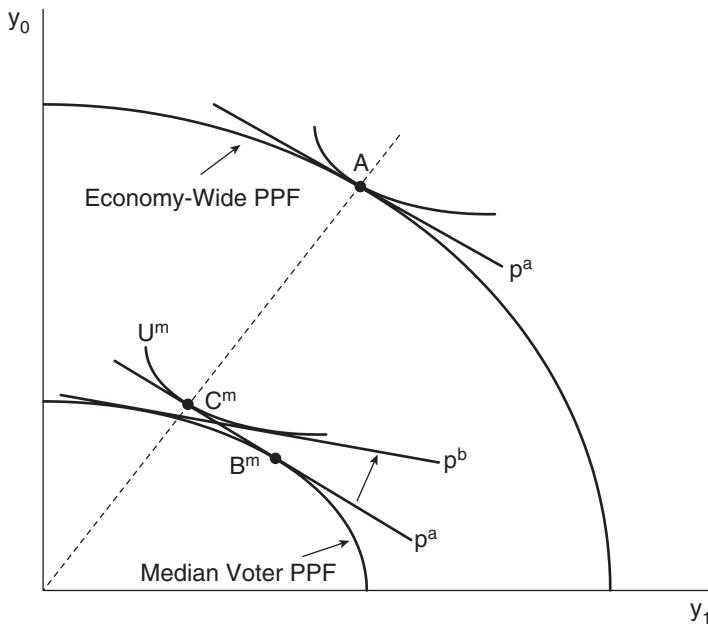


Figure 10.5

To demonstrate the impossibility of path (3) in the HO model, we use a simple but powerful graphical technique introduced by Levy. In figure 10.5, a country's production possibilities frontier (PPF) is illustrated, with the autarky equilibrium at point A. Treating all consumers as having identical homothetic tastes, there is an indifference curve tangent to the PPF at point A, with the autarky relative price of good 1 given by the slope of the line p^a . Also shown is a *smaller* PPF, which corresponds to just the labor and capital endowments owned by the median voter. Treating good 1 as labor intensive, and supposing that the median voter has a capital/labor ratio that is lower than the overall ratio for the economy, this PPF is skewed towards good 1. Thus, at the autarky price ratio p^a , the median voter would have "individual production" at point B^m , and consumption at point C^m , with utility of U^m . It is evident that there are gains from the opportunity to trade with others in the economy, since consumption at point C^m is above the median voter's "individual PPF."

Now suppose that the economy enters into trade with another country (or group of countries). We assume that the countries are similar enough in their factor endowments for factor price equalization to occur, and let k^w denote the "world" capital/labor ratio in this integrated equilibrium (this would be the overall capital/labor ratio of the two countries with bilateral free trade, or all countries with multilateral free trade). With identical homothetic tastes across countries, the equilibrium relative price of good 1 can be written as a function $p(k^w)$, with $p'(k^w) > 0$ since good 1 is labor intensive.

If the economy shown in figure 10.5 enters into trade with another country that has a *lower* capital/labor endowment than its own, this will *lower* the relative price of good 1, which then becomes the import good. A slight fall in p , from p^a in the direction of p^b , will clearly *lower* the welfare of the median voter. If this import price falls by a substantial amount, however, then the median voter can instead gain: at the price $p^b < p^a$

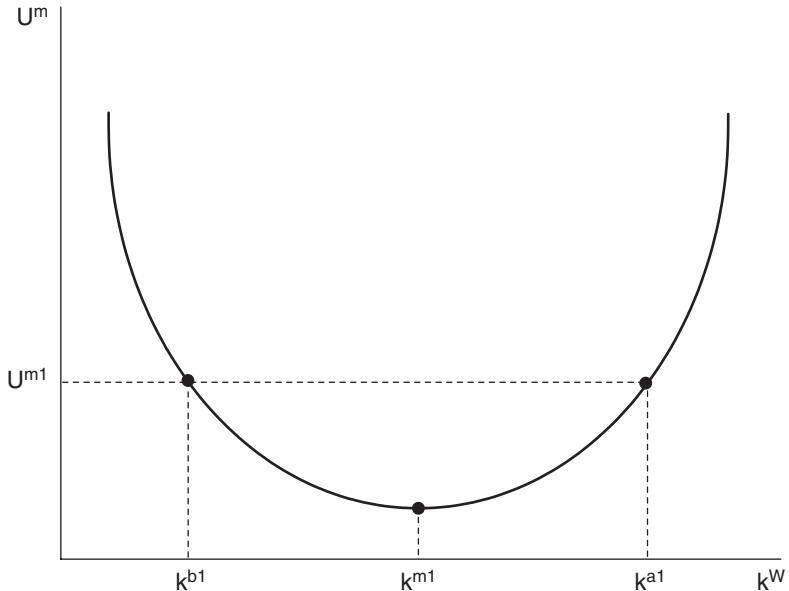


Figure 10.6

shown in figure 10.5, the median voter has the *same* utility level U^m , so for $p < p^b$ the median voter will gain. Conversely, starting at the autarky equilibrium, if the economy enters into trade with another country that had a *higher* capital/labor endowment than its own, then the relative price of good 1 will rise as that good is exported, and for $p > p^a$ in figure 10.5 the median voter again gains.

Summarizing these observations, in figure 10.6 we graph the welfare of the median voter in country 1 against the capital/labor endowment k^w of the integrated world equilibrium. The endowment of the country is k^{a1} , so in autarky the relative price of good 1 is $p^{a1} = p(k^{a1})$, and the median voter obtains utility U^{m1} . If the country now trades in an integrated equilibrium with $p(k^w)$ *slightly lower* than p^{a1} , meaning that $k^w < k^{a1}$, we have argued above that the median voter is worse off. This is shown by utility declining in figure 10.6 for k^w slightly below k^{a1} . But if the country trades in an integrated equilibrium with k^w *much lower* than k^{a1} , than we can obtain the price p^{b1} at which the median voter is back at his or her autarky utility level U^{m1} .

Let us denote the integrated capital/labor ratio at which the median voter achieves the same utility as in autarky by k^{b1} , with $p^{b1} = p(k^{b1})$. Then we see from figure 10.6 that the median voter in country 1 will *reject free trade* with other countries when the integrated equilibrium has a capital/labor endowment $k^w \in (k^{b1}, k^{a1})$, but otherwise will accept free trade for $k^w \notin (k^{b1}, k^{a1})$. Notice that the “rejection set” (k^{b1}, k^{a1}) is convex, which reflects the fact that utility of the median voter, $U^m(k^w)$, is a quasi-convex function of the capital/labor endowment in the integrated equilibrium.

Now consider the situation of countries 1 and 2 deciding whether to engage in bilateral free trade. In figure 10.7 we graph the utilities of the median voters in each country. The “rejection sets” are (k^{b1}, k^{a1}) and (k^{b2}, k^{a2}) , and in the case we have illustrated, country 2 is capital abundant, since its capital/labor endowment exceeds that of country 1, $k^{a1} < k^{a2}$. Then with bilateral free trade, the integrated world equilibrium has a capital/labor endowment denoted by $k^{12} = (K^1 + K^2)/(L^1 + L^2)$, and this lies in

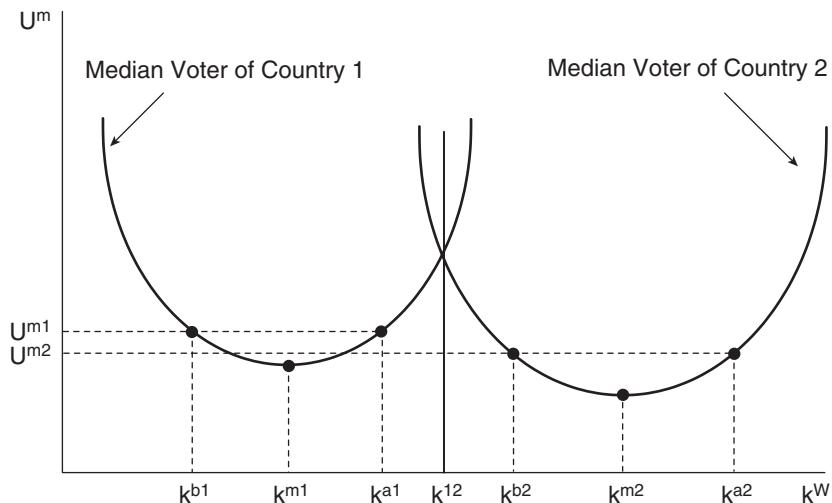


Figure 10.7

between the capital/labor endowments of the two countries, $k^{a1} \leq k^{12} \leq k^{a2}$. In order for the median voter *in both* countries to benefit from bilateral free trade, it must be the case that k^{12} lies *outside* both the sets (k^{b1}, k^{a1}) and (k^{b2}, k^{a2}) , as illustrated. Since we also have that $k^{a1} \leq k^{12} \leq k^{a2}$, it then follows that bilateral free trade can benefit both countries only if the two “rejection sets” are *disjoint*, as illustrated.

Now let us ask whether countries 1 and 2, which are both better off from *bilateral* free trade, would be better or worse off from *multilateral* free trade. Let k^w denote the world capital/labor endowment in that integrated equilibrium. If $k^w = k^{12}$, then the two countries are obviously indifferent between bilateral and multilateral free trade (since the relative price is not affected). If k^w is *slightly above* k^{12} , then we can see from figure 10.7 that country 1 would *gain* moving from bilateral to multilateral, but country 2 would *lose*. On the other hand, if k^w is *much above* k^{12} , so that $k^w > k^{a2}$ in figure 10.7, then it is possible that both countries gain. Checking the other cases, if k^w is *slightly below* k^{12} , then country 2 would *gain* moving from bilateral to multilateral, but country 1 would *lose*; while if k^w is *much below* k^{12} , so that $k^w < k^{b1}$, then it is possible that both countries gain. In all cases, we see that *countries 1 and 2 cannot both lose in the move from bilateral to multilateral free trade*. Therefore, we have proved part (a) of the following theorem.

THEOREM (LEVY 1997)

- (a) Under the two-by-two HO production structure, if the median voters in both countries gain from bilateral free trade, then at least one country must gain from multilateral free trade;
- (b) Allowing for product variety under monopolistic competition, it is possible that the median voters in both countries gain going from autarky to bilateral free trade, or from autarky to multilateral free trade, but lose going from bilateral to multilateral free trade.

Part (a) rules out path (3) and establishes that bilateral free-trade agreement cannot act as a hindrance to a multilateral deal. The essential logic is that the two countries will accept the bilateral arrangement only if their “rejection sets” are *disjoint*, with their combined capital/labor endowment k^{12} lying in between these sets. This immediately rules out having the endowment k^w of a multilateral equilibrium contained in both sets. It is not difficult to see that this logic extends to an initial free-trade area involving any number of countries, who cannot subsequently all lose from multilateral free trade.

Having established this result for the two-by-two HO model, what about more general models? In part (b) of the theorem, we suppose that one sector produces a differentiated product under monopolistic competition, while another sector produces a labor-intensive homogeneous good. Furthermore, suppose that countries 1 and 2 are *identical*. There will still be gains from free trade between them, just as was demonstrated in Krugman’s (1979) model at the beginning of chapter 5. So the median voters in countries 1 and 2 both prefer the free-trade area consisting of the country pair (1, 2).

Now suppose that there is country 3, with the same production structure as 1 and 2 but having a higher labor/capital endowment. Under multilateral free trade, it would produce more of the labor-intensive homogeneous good, and for convenience suppose that it is fully specialized in that good. Then the free-trade area consisting of (1, 2) obtains no further product variety gains from free trade with country 3, but there will still be a change in relative prices: since country 3 is labor abundant, free trade with it would lower the relative price of that good. This can quite possibly *lower* the welfare of the median voters in both country 1 and country 2. Therefore, it is entirely possible that the median voters in countries 1 and 2 *both lose* going from bilateral to multilateral free trade.

Finally, we need to check whether our initial assumption—that countries 1 and 2 are both better off with multilateral free trade than in autarky—is satisfied in this example. That is, we need to consider utility for the median voter in country 1 when it moves from autarky to free trade with countries (2, 3) combined. Country 1 then gains from increased product variety with the identical country 2, and possibly loses from the price change with labor-abundant country 3. Provided that the product variety gains *exceed* the losses for the median voter due to import competition, which is entirely possible, then country 1 gains overall from free trade with (2, 3). The same argument applies to country 2 when it moves from autarky to free trade with (1, 3). So despite the fact that multilateral free trade can bring gains as compared to *autarky*, we have found that it can bring losses as compared to *bilateral* free trade. In this situation, the bilateral agreement would prevent multilateral free trade from being pursued: it would indeed be a stumbling block rather than a stepping stone. This demonstrates part (b) of the theorem, and the general concern expressed by Bhagwati (1993; 2002, lecture 3) that regional trade agreements may inhibit the pursuit of multilateral free trade.¹⁵

POLITICAL ECONOMY OF FOREIGN INVESTMENT IN CHINA

All of the applications in the chapter so far have been to economies governed by direct democracy (with the median voter determining the optimal policy) or representative democracy (where voters elect representatives who then determine policies).

¹⁵ McLaren (2002) further argues that introducing *sector-specific sunk costs* can have a similar effect, under which a regional agreement can inhibit future multilateral deals.

Grossman and Helpman (2001) present a general formulation of democratic political systems where special interests exert an influence on policies. We conclude this chapter with an application of their framework to a nondemocratic system—the People's Republic of China. The motivation for this application is much the same as in Goldberg and Maggi (1999): to use the observed levels of some policy instrument or outcome to “reveal” the weights given by the government to various interest groups. In the case of China, import tariffs before its entry into the WTO in 2001 were bound at relatively high and inflexible levels, so that this policy instrument cannot be used. Instead, we examine the policies toward foreign investment, as indicated by the entry of foreign firms into various provinces, to infer the weights in the government's objective function.

The model we use relies on the choice of policies to influence foreign investment. We shall not discuss the decisions of firms to enter foreign markets until the next chapter, but introduce here the idea that they may enter a market to “jump the tariff,” that is, because they cannot export there due to tariff protection.¹⁶ Grossman and Helpman (1996) have introduced foreign investment into their “protection for sale” model, and Branstetter and Feenstra (2002) make use of that model to analyze policies toward multinationals in China, as follows.

Consumers are assumed to have preferences over an additively separable numeraire good d_0 , and a CES aggregate denoted by D :

$$U = d_0 + \left(\frac{\gamma}{\gamma - 1} \right) D^{(\gamma-1)/\gamma}, \quad \gamma > 0, \quad \gamma \neq 1, \quad (10.30a)$$

where the CES aggregate is

$$D = [N_h d_h^{(\sigma-1)/\sigma} + (N_f - M) d_f^{(\sigma-1)/\sigma} + M d_m^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)}, \quad \sigma > 1. \quad (10.30b)$$

The values d_j denotes consumption of the differentiated varieties from sources: $j = h$ (home firms), of which there are N_h in number; $j = f$ (foreign imported products), of which there are $(N_f - M)$; and $j = m$ (multinationals in China) of which there are M . As usual, σ is the elasticity of substitution across individual varieties, while γ is the elasticity of demand for the aggregate good D , and we add the restriction that $\sigma > \gamma$.

Branstetter and Feenstra (2002) suppose that the domestic industry is owned by the government, as with state-owned firms in China. The profits earned by these firms receive extra weight in the government's objective function.¹⁷ The presence of multinationals presents a potential threat to the state-owned enterprises through product market competition.¹⁸ This creates a conflict between the entry of foreign firms and the profits of state-owned firms that must be resolved through government policy toward entry of the multinationals.

¹⁶ Bhagwati (1987) has taken this argument a step further and suggested that foreign firms may enter a market to “defuse” the threat of future protection, in what he calls “quid-pro-quo” foreign investment. See also Bhagwati, et al. (1987), Dinopoulos (1989, 1992), and Wong (1989).

¹⁷ As Naughton (1996) and other authors have noted, the Chinese government still relies on remittances from state-owned enterprises for about two-thirds of its revenue.

¹⁸ Notice that foreign investment might also be complementary rather than competitive, because foreign firms were required to have local partners in China prior to its entry into the WTO in 2001, which could benefit the state-owned enterprises. But a number of studies suggest that the Chinese government, both national and local, is acutely aware of this competition and has taken steps to impede the ability of foreign firms to compete in the Chinese market (e.g., Li and Chen 1998; Rosen 1998).

In designing the government policies, we allow the various *provinces* of China to exercise autonomous control over multinational entry. This is broadly consistent with the multiple layers of approval required for foreign investment in China. Only the largest projects (initially \$30 million but later \$50 million) require approval from the central government, whereas all projects require approval from the regional and provincial governments. In addition, we suppose that the separate provinces control trade among themselves through *internal* border barriers. The level of such barriers is the subject of current debate among experts on the Chinese economy. On the one hand, Young (2000) has argued that internal border barriers are rising, whereas Naughton (2003) disagrees. For reasons of modeling convenience, we take an extreme version of the view in Young and assume there is no trade among the provinces. This means that when a multinational enters one province, it is selling only to the local consumers. The utility function in (10.30) therefore applies to *each* province, which differs in the number of firms of each type found there.

We will suppose that labor is the only factor of production, and one unit of the numeraire is produced with one unit of labor, so wages are unity. We shall assume, however, that the multinational firms pay a *wage premium* of $(w - 1) > 0$. The wage premium is meant to proxy for a wide array of possible benefits that multinationals bring that are not captured in our model. Costs for the locally produced products are c_j , $j = h, m$. Costs for products produced abroad are c_f , but they also face a specific tariff of t , so that marginal costs become $c_f + t$. Prices are a standard markup over these marginal costs, using the elasticity σ . We make the key assumption that $c_m < c_f + t$.

This assumption ensures that the price charged by multinationals, p_m , is less than that for imports, p_f . It follows that as more multinationals enter (M increases) and these products are sold for $p_m < p_f$, the demand for products of the state-owned firms will decline due to substitution away from these products. This is how product market competition between the multinationals and domestic firms is captured in the model.

Each foreign firm faces the decision of whether to supply through imports, or through setting up a local plant that requires a fixed cost of $F > 0$. We suppose that the government also charges the multinational a profit tax of $\tau \geq 0$. This instrument is supposed to reflect the vast range of actual policies used in China to extract rents from multinationals, and not just the corporate tax on multinationals.¹⁹ By modeling these policies as a tax on profits, we are abstracting from the inefficiencies caused by actual policies.

The net profits earned locally are thus $(1 - \tau)\pi_m - F$, where $\pi_m = p_m d_m / \sigma$ is profits under the standard CES markup-pricing rule. Alternately, the multinational could just export to the home country, and earn $\pi_f = p_f d_f / \sigma$. Thus, entry will occur if and only if $(1 - \tau)\pi_m - F \geq \pi_f$. This condition is written as

$$(1 - \tau) \frac{p_m d_m}{\sigma} - F \geq \frac{p_f d_f}{\sigma}. \quad (10.31)$$

¹⁹ In fact, foreign firms are often taxed at zero or reduced rates for the first years of operation. Despite this, there are many ways that local and national agencies extract rents from the multinationals. For example, the fact that most multinationals have had to use local partners reflects an implicit tax on their profits, which are shared with the partner; similarly, the land-use fees that are commonly charged reduce the multinationals' profits, as do conditions of technology transfer and export requirements. Of course, bribes paid to allow multinationals to enter are another example of the profit tax. Wei (2000) argues that corruption in China, which includes the need for "questionable payments," acts as a significant deterrent to foreign direct investment.

Entry is influenced through the profits tax τ and the specific tariff t , so the number of multinational firms is written as a function $M(\tau, t)$. Multinationals react in the expected manner to changes in these policies: when (10.31) holds as an equality, then $dM/d\tau < 0$ and $dM/dt > 0$.

We now define the government's objective function over the various interest groups, beginning with consumer/worker utility U , which receives a weight of α . The profits of the home state-owned firms are $N_h \pi_h = N_h p_h d_h / \sigma$, which accrue to the regional and national government. We give revenue from state-owned firms a weight β in the objective function. Finally, the government extracts rents $M\tau\pi_m$ from the multinationals, and also collects tariff revenue of $(N_f - M)td_f$. These two sources of revenue are each given weights of unity. The objective function for each province is then defined by

$$G(M, \tau, t) \equiv \alpha U + \beta N_h \pi_h + M\tau\pi_m + (N_f - M)td_f, \quad (10.32)$$

where the multinationals entering each province are endogenously determined from (10.31).

We suppose that the central and the provincial governments jointly determine the rents appropriated from the multinationals. The central government also chooses the tariff rate. Denoting provinces by the subscript i , we let $G_i[M_i(\tau_p, t), \tau_p, t]$ denote the objective function (10.32) for each. Then the profit tax and tariff are chosen to solve

$$\max_{\tau_p, t} \sum_i G_i[M_i(\tau_p, t), \tau_p, t]. \quad (10.33)$$

We solve (10.33) over the choice of the profits tax rate τ_i , while a discussion of the solution for tariffs t is in Branstetter and Feenstra (2002). Differentiating (10.33), the first-order condition is that $\partial G_i / \partial M_i = -M_i \pi_{mi} / (\partial M_i / \partial \tau_i)$. This means that the gain from attracting one more multinational ($\partial G_i / \partial M_i$) is just balanced against the fall in the revenue $\tau_i M_i \pi_{mi}$ when the tax rate is lowered to attract another multinational. That is, the regions are acting as monopsonists in attracting foreign capital.

This first-order condition for the choice of τ_i can be written as²⁰

$$s_{mi} = -\beta s_{hi} + \eta \left(\frac{w_i - 1}{w_i} \right) + \alpha(\sigma - 1) \left(\frac{w_i - 1}{w_i} \right) (s_{hi} - s_{fi}) - \sigma \left(\frac{t}{p_f} \right) s_{fi} + \left(\frac{M_i}{N_f - M_i} \right) s_{fi} + \varepsilon_p \quad (10.34)$$

where s_j , $j = h, f, m$ denotes the *share of provincial consumption* purchased from home state-owned firms, imported products, or multinational products, respectively, and ε_p is a term involving unobserved profit levels that is treated as a random error.

The first term on the right side of (10.34) is the share of provincial consumption on state-owned firms, which enters with the coefficient $-\beta$. Thus, the weight on state-owned firms in the provincial objective function is simply obtained as the coefficient on their share in the regression (10.34). A high weight on the state-owned firms indicates that in provinces where these firms are more prevalent, the share of multinational firms will be correspondingly reduced. The next term on the right side of (10.34) is the wage premium paid by multinationals, which has the coefficient $\eta \equiv \alpha(\sigma - 1)(\gamma - 1)/(\sigma - \gamma)$,

²⁰ This first-order condition is derived in Branstetter and Feenstra (2002).

which is of ambiguous sign. Following this is the wage premium times the share of spending on state-owned firms plus imports. When the wage premium is higher, we expect that regions would be more willing to accept multinationals, and this is confirmed by having a positive coefficient $\alpha(\sigma - 1)$ on that variable.

The estimate of σ itself comes from the next term, which is the tariff rate times the share of spending on imports. This term reflects the loss in tariff revenue as multinationals enter and the coefficient is $-\sigma$.²¹ Thus, combining it with the former coefficient, we can recover an estimate of α . The final term reflects the number of multinationals times the share of imports. For simplicity we measure the number of multinationals M_i by their capital stock in each province, and treat the number of foreign firms wanting to export or invest in China, N_f , as constant over provinces and time, and estimate it as a coefficient.

For estimation, a time subscript is added to most variables in (10.34). There is an obvious endogeneity problem in that the share of state-owned sales, on the right, will be correlated with the share of multinational sales (since all the shares sum to unity). To offset this, Branstetter and Feenstra estimate (10.34) in levels rather than shares (i.e., multiplying all variables by provincial apparent consumption) and also use instruments and weights.²²

The results from estimating (10.34) using a panel of data for Chinese provinces over 1984–95, are shown in table 10.1. Each of the three columns of results uses a different estimation technique, beginning with OLS, and moving to weighted TSLS (standard errors are in parentheses). OLS estimates are largely uninformative, as we are unable to obtain reasonable estimates of σ . However, our estimates improve substantially when we move to two-stage least squares, and further weight the observations by provincial GDP. Each regression includes a full set of provincial and year fixed effects.

The estimate of the parameter β , the weight of state-owned enterprise output in the government's objective function, is taken from the regression coefficient on state-owned enterprise output, and ranges from 1.3 to 1.66 (ignoring the OLS estimate).²³ Estimates of α , the weight on consumer welfare, can be derived from the regression coefficients shown in the second and third rows of table 10.1. For our preferred specification, which is the weighted TSLS estimate in the third regression, we obtain $\hat{\alpha} = 0.24$ as reported in table 10.2. The difference between the state-owned and consumer weight is $\hat{\beta} - \hat{\alpha} = 1.42$, and the standard error of this difference is computed as 0.46. Thus, we find that the weight given to consumer welfare is *significantly lower* than that applied to the output of state-owned enterprises (at the 5 percent level), with their ratio being about *one-seventh* in our preferred estimates. Without using provincial GDP as weights, the TSLS estimate in the second regression gives $\hat{\alpha} = 0.34$, which is still *one-quarter* of the weight given to the state-owned enterprises. Turning to other

²¹ See our discussion of Brecher and Diaz Alejandro (1977) in problem 12.1.

²² The instruments are provincial electricity production (used as an instrument for expenditure on state-owned firms); indexes of urban, state, and overall wages (used as instruments for the wage premium); provincial GDP, population, average rural and urban income (used as instruments for apparent consumption); provincial processing imports (used as an instrument for ordinary imports); and various interactions between these terms. The weighted regressions use provincial GDP as a weight.

²³ We note that the sign and magnitude of this estimate is contingent on including total provincial consumption as a control variable in the regression: controlling for total provincial spending, a decline in the spending on state-owned firms is associated with a rise in the spending on multinationals; but without this control variable, the sign of $\hat{\beta}$ is reversed. See the empirical exercise to reproduce tables 10.1 and 10.2.

TABLE 10.1: Provincial Spending on Output of Multinational Enterprises as Dependent Variable

	Estimation Method:	OLS	TSLS	Weighted TSLS
Independent Variable:	Coefficient			
Provincial Spending on State-owned production	$-\beta$	-0.41 (0.12)	-1.30 (0.40)	-1.66 (0.48)
Wage premium \times (state-owned + imports)	$\alpha(\sigma - 1)$	0.42 (0.23)	2.19 (0.87)	3.62 (1.16)
Tariff \times imports	σ	-0.15 (2.7)	-7.4 (3.3)	-15.9 (5.1)
Wage premium	$\frac{\alpha(\sigma - 1)(\gamma - 1)}{(\sigma - \gamma)}$	-118 (64)	-967 (435)	-3333 (1085)
FDI stock \times Imports	$1/N_f$	0.037 (0.022)	0.037 (0.017)	0.052 (0.013)
Provincial Apparent Consumption		0.11 (0.02)	0.15 (0.03)	0.15 (0.02)
R^2, N		0.85, 297	0.74, 280	0.76, 280

Notes: The sample consists of twenty-nine provinces (excluding Tibet) over 1984–95, using fixed effects for provinces and for time; only provinces with positive multinational output are included. All regressions except the first are estimated with TSLS. The weighted regressions use provincial GDP as a weight. White standard errors are reported in parentheses.

TABLE 10.2: Coefficient Estimates, by Time Period

	Coefficient:			
	N	α	β	σ
Sample:				
1984–95	280	0.24 (0.07)	1.66 (0.48)	15.9 (5.1)
1988–95	210	0.24 (0.13)	1.32 (0.50)	11.5 (6.7)
1990–95	132	0.20 (0.15)	1.04 (0.49)	10.0 (7.6)

Notes: Computed from the third regression reported in table 10.1, but run over different samples. The estimation method is weighted TSLS. White standard errors are in parentheses.

parameters, we obtain high estimates of the elasticity of substitution σ (from 7 to 16), though these are somewhat imprecise.

A breakdown of the structural coefficients of interest is provided in table 10.2 for later subsamples of the data, the 1988–95 period and the 1990–95 period. As we confine our view to the later subsamples, we lose observations and consequently precision in some of our estimates; subperiods smaller than 1990–95 do not yield many significant coefficients. In the later subperiods, the estimated magnitude of the weight on state-owned enterprises falls. This is consistent with the historical trend toward liberalization, of course. In the relatively liberal 1990–95 subperiod, for example, we

find that the weight on state-owned enterprises is *unity*, which is still considerably higher than the weight on consumer welfare (though the difference between them is no longer significantly different from zero).

These estimates provide a stark contrast to the results of Goldberg and Maggi (1999) for the United States, where consumer welfare had a weight that was fifty to one hundred times more than campaign contributions. For China, we find that state-owned enterprises have a weight that is between four and seven times greater than that given to consumers. The evidence of a political premium on state-owned industries diminishes over time, but the point estimates still indicate these firms are favored. Branstetter and Feenstra (2002) further investigate the impact on provincial objective functions of the changes in tariff structure that China has promised under WTO accession. They find that these changes could potentially lower welfare in some provinces, due to the exit of multinationals.

CONCLUSIONS

In this chapter we have relied on two major models: the median voter model (first applied to trade policy by Mayer [1984]), and “protection for sale” of Grossman and Helpman (1994, 1995a,b, 1996). There are many other models of the political economy of trade policy, and the reader is referred to Hillman (1989) and Rodrik (1995,a) for surveys. The median voter model relies on a voting mechanism under direct democracy, and despite this rather simplified framework, receives surprising empirical support from Dutt and Mitra (2002). In comparison, the “protection for sale” model allows for the government to be influenced by many industry lobbies. This has received strong empirical support for the United States from Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000), along with numerous empirical applications to other countries, as surveyed by Gawande and Krishna (2003).

In their original formulations, the median voter and “protection for sale” models both treat international prices as fixed, so that tariffs shift income between interest groups, but cannot raise the welfare of all groups. In contrast, Bagwell and Staiger (1999, 2002) explicitly introduce a large-country model, under which tariffs can improve the terms of trade and raise a country’s welfare. The government objective function used by them encompasses the median voter and “protection for sale” model, while Grossman and Helpman (1995a) also investigate terms of trade effects in the latter model. A country that uses an import tariff to improve its terms of trade does so at the expense of the other country (plus an added deadweight loss), and so from the viewpoint of world efficiency, it would be better for countries to agree to avoid such actions. This creates an important role for the GATT, and the principle of *reciprocity* in tariff reductions provides a way to eliminate the incentive to manipulate the terms of trade. Notice that even under this principle, tariffs are still used in Bagwell and Staiger (1999, 2002) and Grossman and Helpman (1995a) for political economy reasons.

We have also reviewed the role of regional trade agreements, and whether they help or hinder the movement toward multilateral free trade. Levy (1997) employs a median voter model to address this, while Grossman and Helpman (1995b) extend the “protection for sale” model in that direction. Empirically, it would be difficult to infer the impact of regional agreements on the pursuit of multilateral free trade, since the multilateral agreements of GATT are themselves so infrequent. But there are a

number of related empirical questions that can be addressed. Krishna (2003) provides empirical evidence on the distance between countries and the potential welfare gains from regional agreements. He finds that close trading partners do not appear to have a “natural” advantage in created welfare-improving free-trade areas. Baier and Bergstrand (2003) and Magee (2003) introduce the endogeneity of free-trade agreements into a gravity equation, of the type discussed in chapter 5. These authors find that the impact of FTAs is significantly affected by treating these agreements as endogenous. The variables they include to explain free-trade areas also provide a test of the determinants of these regions, and these papers are recommended for further reading.

PROBLEMS

In the model of Mitra (1999), organized and unorganized industries are treated as symmetric, having the same world prices, capital stocks, production and demand functions and differ only in their lobbying ability. Therefore, the return to the fixed factor in any industry can be written as $\pi(p_i)$ instead of $\pi_i(p_i)$, depending on the price but not on any other industry characteristic. Normalizing the world prices at unity, denote the price in organized industries by $p_o = 1 + t_o$, and in unorganized industries by $p_u = 1 + t_u$. The following questions derive additional results for that model.

- 10.1 Consider the return to the specific-factor in sector i , from (10.8):

$$\pi(p_i) = \max_{L_i} [p_i f(L_i, K) - L_i],$$

where we assume that $f_{LL} < 0$. By differentiating the first-order condition for this problem and using $\pi'(p_i) = y(p_i)$, show that profits are a convex function of the industry price.

- 10.2 Replace $\lambda_k n_o$ with λ_o in (10.17), and rewrite the optimal tariff $t_o > 0$ and subsidy $t_u < 0$ as

$$t_j = \left(\frac{\delta_j - \lambda_o}{\alpha + \lambda_o} \right) y(p_j) \left| \frac{\partial m}{\partial p_j} \right|^{-1}, \quad \text{where } \delta_o = 1 \text{ and } \delta_u = 0.$$

Output depends on the industry price, but for convenience, we treat the import derivative as fixed and equal across industries. Then compute $\frac{dt_o}{d\lambda_o}$ and $\frac{dt_u}{d\lambda_o}$, and show that $\left| \frac{dt_o}{d\lambda_o} \right| > \left| \frac{dt_u}{d\lambda_o} \right|$.

- 10.3 Continuing to denote $\lambda_k n_o$ by λ_o , consider the return to the specific factor in an organized rather than an unorganized industry, $\Delta W(\lambda_o) = \pi(p_o) - \pi(p_u)$. Using the results from problems 10.1 and 10.2, show that $\Delta W'(\lambda_o) = \pi'(p_o)(dt_o/d\lambda_o) - \pi'(p_u)(dt_u/d\lambda_o) < 0$.
- 10.4 The equilibrium prices $p^w(\tau, \tau^*)$ in (10.21) are defined by $m(p^w\tau) = x^*(p^w/\tau^*)$. Assuming that $m'(p^w\tau) < 0$ and $x''(p^w/\tau^*) > 0$, show that the inequalities in (10.22) hold.
- 10.5 Expression (10.29) shows that if the tariff cuts are reciprocal, then the number of firms is constant. In this question you are asked to show how a *unilateral* increase in the home tariff affects the number of home and foreign firms.

(a) Use (5.12), (5.40), and (5.41) to show that

$$P^1 = \left[\frac{\bar{y}p^\sigma}{\gamma L^1} \frac{1 - ((1 + \tau^{12}) T^{12})^{1-\sigma}}{1 - ((1 + \tau^{12}) T^{12} (1 + \tau^{21}) T^{21})^{1-\sigma}} \right]^{1/(\sigma-1)} \quad \text{and}$$

$$P^2 = \left[\frac{\bar{y}p^\sigma}{\gamma L^2} \frac{1 - ((1 + \tau^{21}) T^{21})^{1-\sigma}}{1 - ((1 + \tau^{21}) T^{21} (1 + \tau^{12}) T^{12})^{1-\sigma}} \right]^{1/(\sigma-1)}.$$

(b) Use the expression for price indexes from (a) and the definition of price index in (5.25) to show that the number of firms is equal to

$$N^1 = \frac{\gamma}{\bar{y}p} \left[\frac{L^1}{1 - ((1 + \tau^{12}) T^{12})^{1-\sigma}} - \frac{L^2 ((1 + \tau^{21}) T^{21})^{1-\sigma}}{1 - ((1 + \tau^{21}) T^{21})^{1-\sigma}} \right] \quad \text{and}$$

$$N^2 = \frac{\gamma}{\bar{y}p} \left[\frac{L^2}{1 - ((1 + \tau^{21}) T^{21})^{1-\sigma}} - \frac{L^1 ((1 + \tau^{12}) T^{12})^{1-\sigma}}{1 - ((1 + \tau^{12}) T^{12})^{1-\sigma}} \right].$$

EMPIRICAL EXERCISE

This exercise is to reproduce regression results from Branstetter and Feenstra (2002). To complete the exercise, the file “china_fdi.dta” should be stored in: c:\Empirical_Exercise\Chapter_10\.

- 10.1 Run the program “share_reg.do” to reproduce the regression results in table 10.1. Notice that apparent consumption (“appcon3”) is included as a control variables in these regressions. What happens if this variable is dropped?
- 10.2 Open the excel file “standard_errors.xls” to see how the results in table 10.2 are computed. What formula is used for α , and for its variance and standard error? Justify these formulas.

Hint: Consider the easier case of taking the difference between α and β . Then

$$\text{var}(\alpha - \beta) = E[(\alpha - \beta) - (\bar{\alpha} - \bar{\beta})]^2 = Ef(\alpha, \beta),$$

where $f(\alpha, \beta)$ is the quadratic function indicated. To compute this expected value, we take a second-order Taylor series expansion:

$$f(\alpha, \beta) \approx f(\bar{\alpha}, \bar{\beta}) + f_{\alpha}(\bar{\alpha}, \bar{\beta})(\alpha - \bar{\alpha}) + f_{\beta}(\bar{\alpha}, \bar{\beta})(\beta - \bar{\beta}) + \frac{1}{2} f_{\alpha\alpha}(\bar{\alpha}, \bar{\beta})(\alpha - \bar{\alpha})^2 + \frac{1}{2} f_{\beta\beta}(\bar{\alpha}, \bar{\beta})(\beta - \bar{\beta})^2 + f_{\alpha\beta}(\bar{\alpha}, \bar{\beta})(\alpha - \bar{\alpha})(\beta - \bar{\beta}).$$

It follows that

$$Ef(\alpha, \beta) \approx \frac{1}{2} f_{\alpha\alpha}(\bar{\alpha}, \bar{\beta}) \sigma_{\alpha}^2 + \frac{1}{2} f_{\beta\beta}(\bar{\alpha}, \bar{\beta}) \sigma_{\beta}^2 + f_{\alpha\beta}(\bar{\alpha}, \bar{\beta}) \text{cov}(\alpha, \beta).$$

So computing the derivatives of the quadratic function $f(\alpha, \beta)$, we obtain

$$\text{var}(\alpha - \beta) = E[(\alpha - \beta) - (\bar{\alpha} - \bar{\beta})]^2 \approx \sigma_{\alpha}^2 + \sigma_{\beta}^2 - 2 \text{cov}(\alpha, \beta).$$

Now given the formula for α , use the same approach to justify its variance.

Trade and Endogenous Growth

The link between trade and growth has long been a question of interest in both theory and policy. David Ricardo developed a dynamic model of corn and velvet production, with corn produced by land and labor and velvet produced with labor alone.¹ The need to pay labor from a “wage fund” established prior to production prevents all labor from being employed initially (i.e., there is a liquidity constraint on firms). Ricardo showed that in autarky the gradual expansion of the wage fund and growth of velvet production would lower its relative price, until a long-run equilibrium was reached. Opening trade, however, would allow the relative price of velvet to be maintained at world levels, thereby benefiting capitalists at the expense of landowners. This remarkable model demonstrates many of the issues that are of interest in modern discussions of trade and growth: the possibility that growth will be associated with continual changes in prices, and conversely, the impact of trade on prices and growth rates themselves.

In the first of these issues we will be primarily concerned with the effects of growth on the *terms of trade*, that is, the price of exports relative to imports. This brings us to the famous case of “immiserizing growth,” due to Bhagwati (1958), which is growth that can actually lower a country’s welfare due to a fall in the terms of trade. The idea that developing countries might be subject to a *decline* in their terms of trade, particularly for primary commodities, is associated with Latin American economist Raul Prebisch (1950). While there is little evidence to support that hypothesis in general, it is still the case that declines in the terms of trade due to growth have been observed (Acemoğlu and Ventura 2002; Debaere and Lee 2003).

In the second part of the chapter, we model the underlying determinants of growth through research and development. This class of models goes by the name “endogenous growth,” and their application to trade has been most thoroughly studied by Grossman and Helpman (1991). We will describe just one of the models described by those authors, which is a dynamic version of the CES (constant elasticity of substitution) monopolistic competition model presented in chapter 5.²

¹See Findlay (1984, 187–91), who cites this model to Ricardo’s *Essay on the Influence of a Low Price of Corn upon the Profits of Stock* (Ricardo 1951, 4:1-42).

²This model was introduced by Judd (1985), Grossman and Helpman (1989, 1990), and Romer (1990), and also applied to trade by Rivera-Batiz and Romer (1991a,b). The alternative to the CES model is the “quality ladders” model of product improvement, as described by Segerstrom, Anant, and Dinopoulos (1990), Grossman and Helpman (1991), and Taylor (1993). There are many other writings on endogenous growth and trade, including Taylor (1994) on the gains from trade and Dinopoulos and Segerstrom (1999a,b) on the dynamic effects of tariffs.

Even in this case, we will find that trade can have a fairly wide range of effects on product innovation within a country and therefore on growth. When there is free trade in *all goods*, including *free international flows of knowledge*, trade raises growth rates and brings gains to all countries; but this is not necessarily the case with free trade in only a subset of the goods without flows of knowledge. Empirically, then, the beneficial effects of trade on growth depend, very strongly on whether there are international spillovers of knowledge, and so we shall review the empirical evidence on that and other topics.

The models in Grossman and Helpman (1991) depend on “scale effects,” the idea that large countries grow faster. That feature of the model has been challenged in closed-economy growth models by Jones (1995a,b), who introduces the term “semi-endogenous growth” to describe models where the long-run growth rate does not depend on scale, but where features of the endogenous growth models still operate in the short run. It has taken some time for the international trade literature to develop models of endogenous growth *without* scale effects, where international trade still has a long-run impact on the growth rate. This long-run result is now available in the work of Perla, Tonetti, and Waugh (2014) and Sampson (2016), both of which rely on firms that are heterogeneous in their productivities, as in the Melitz (2003) model studied in chapter 6. We shall present a simplified version of the Sampson (2016) model.

We begin the chapter with a discussion of how to measure productivity growth. This is quite easy in the case of perfect competition, but not under imperfect competition. In the latter case, it turns out that conventional measures of productivity growth are biased due to increasing returns to scale and the markup of price over marginal cost. Hall (1988) has proposed a method to correct for this problem and, at the same time, estimate the markups. In the context of trade reform it allows us to estimate the extent to which trade liberalization has led to a reduction in markups by firms facing import competition. This approach has been applied to Turkey by Levinsohn (1993) and to the Ivory Coast by Harrison (1994), as we shall discuss.

MEASUREMENT OF PRODUCTIVITY

Growth in the output of a firm or industry can occur due to growth in inputs (labor, capital, and human capital), or due to an increase in output that is *not explained* by inputs: the latter is called productivity growth. Specifically, suppose that output y_i is produced using labor L_i and capital K_i , with the production function $y_i = A_i f(L_i, K_i)$, where A_i is a measure of Hick's neutral technological progress. Totally differentiating this, we have $dy_i = dA_i f(L_i, K_i) + A_i f_{il} dL_i + A_i f_{ik} dK_i$. Dividing this by $y_i = A_i f(L_i, K_i)$, and letting $\hat{z} = dz/z$ denote the percentage change in any variable, we see that

$$\hat{y}_i = \hat{A}_i + \left[\left(\frac{f_{il} L_i}{f(L_i, K_i)} \right) \hat{L}_i + \left(\frac{f_{ik} K_i}{f(L_i, K_i)} \right) \hat{K}_i \right]. \quad (11.1)$$

In general, then, output growth can be measured by productivity growth plus a weighted average of the growth in inputs, where the weights are the *elasticities of output with respect to each factor*. The difficulty is that these weights are not directly observable. With the added assumption of perfect competition, however, the marginal products equal $A_i f_{il} = w/p_i$ and $A_i f_{ik} = r/p_i$, in which case the elasticities appearing in (11.1) are measured by $\theta_{il} = wL_i/p_i y_i$ and $\theta_{ik} = rK_i/p_i y_i$, which are the *revenue shares* of labor and capital. Under the further assumption of constant returns to scale, these

shares sum to unity, so that $\theta_{ik} = 1 - \theta_{il}$. Then we see from (11.1) that technological progress \hat{A}_i can be measured by the difference between the growth of output and a share-weighted average of the growth in inputs, which is the definition of total factor productivity (TFP) due to Solow (1957):³

$$\text{TFP}_i \equiv \hat{y}_i - [\theta_{il}\hat{L}_i + (1 - \theta_{il})\hat{K}_i]. \quad (11.1')$$

Notice that this measure of productivity growth is indeed a “residual”—the portion of output growth that is not explained by inputs—or a “measure of our ignorance,” in the memorable phrase of Abramowitz (1956).

Let us now weaken our assumptions and allow for imperfect competition and increasing returns to scale. To introduce increasing returns to scale, we assume that the production function is homogeneous of degree $\mu_i > 1$, so it follows that $\mu_i f(L_i, K_i) = (f_{il} L_i + f_{ik} K_i)$.⁴ Substituting this into (11.1), we obtain

$$\begin{aligned} \hat{y}_i &= \hat{A}_i + \mu_i \left[\left(\frac{f_{il} L_i}{f_{il} L_i + f_{ik} K_i} \right) \hat{L}_i + \left(\frac{f_{ik} K_i}{f_{il} L_i + f_{ik} K_i} \right) \hat{K}_i \right] \\ &= \hat{A}_i + \mu_i \left[\left(\frac{f_{il} L_i}{f_{il} L_i + f_{ik} K_i} \right) \hat{L}_i + \left(1 - \frac{f_{il} L_i}{f_{il} L_i + f_{ik} K_i} \right) \hat{K}_i \right] \\ &= \hat{A}_i + \left(\frac{f_{il} L_i}{f(L_i, K_i)} \right) (\hat{L}_i - \hat{K}_i) + \mu_i \hat{K}_i, \end{aligned} \quad (11.2)$$

where the last line follows from $\mu_i f(L_i, K_i) = (f_{il} L_i + f_{ik} K_i)$ and applying simple arithmetic. With imperfect competition in the product market, the equilibrium condition for the hiring of labor is

$$p_i \left(1 - \frac{1}{\eta_i} \right) A_i f_{il} = w, \quad (11.3)$$

where η_i is the (positive) elasticity of demand. It follows that the real wage equals $(w/p_i) = [(\eta_i - 1)/\eta_i] A_i f_{il}$. The share of labor in total revenue is then

$$\theta_{il} = \frac{w L_i}{p_i y_i} = \frac{(\eta_i - 1) A_i f_{il} L_i}{\eta_i y_i} = \left(\frac{\eta_i - 1}{\eta_i} \right) \left(\frac{f_{il} L_i}{f(L_i, K_i)} \right), \quad (11.4)$$

where the final equality follows from $y_i = A_i f(L_i, K_i)$. Thus, the labor share *understates* the elasticity of output with respect to labor, since $(\eta_i - 1)/\eta_i < 1$. Substituting (11.4) into (11.2), we see that output growth is related to true productivity change by⁵

³In practice, total factor productivity in (11.1') would be measured using first differences rather than infinitesimal changes, as $\text{TFP}_i \equiv \Delta \ln y_i - [\theta_{il} \Delta \ln L_i + (1 - \theta_{il}) \Delta \ln K_i]$. We have already seen this definition of TFP in our discussion of chapter 4, as well as the dual definition, $\text{TFP}_i \equiv (\theta_{il} \Delta \ln w_i + \theta_{ik} \Delta \ln r_i) - \Delta \ln p_i$, where w_i is the wage, r_i is the rental on capital, and p_i is the price in industry i . Note that with the log change in input quantities or prices measured between two points in time, the factor cost-shares θ_j should be measured as the *average* between these two periods; this method is called the Törnqvist index of total factor productivity (see appendix A).

⁴See problem 1.2.

⁵Alternatively, we can difference (11.5) and deduct $[\theta_{il} \Delta \ln L_i + (1 - \theta_{il}) \Delta \ln K_i]$ from both sides, obtaining $\text{TFP}_i = \Delta \ln A_i + [1/(\eta_i - 1)] \theta_{il} (\Delta \ln L_i - \Delta \ln K_i) + (\mu_i - 1) \Delta \ln K_i$, using the definition of TFP from note 3 of this chapter.

$$\hat{y}_i = \hat{A}_i + [\eta_i/(\eta_i - 1)] \theta_{il} (\hat{L}_i - \hat{K}_i) + \mu_i K_i. \quad (11.5)$$

Thus, we see that output growth is composed of true productivity change \hat{A}_i , plus the change in labor/capital growth times the coefficient $\eta_i/(\eta_i - 1)$, plus capital growth times the returns to scale parameters μ_i . The coefficient $\eta_i/(\eta_i - 1)$ equals the ratio of price to marginal cost, so any changes in this coefficient are interpreted as evidence of changes in the price-cost markup. Running (11.5) as a regression therefore allows us to measure the impact of trade liberalization on the markups charged, as well as estimate the returns to scale. Harrison (1994) applies this technique to firm-level data for the Ivory Coast, which had an import liberalization beginning in 1985. The estimating equation is (11.5) but written with discrete rather than infinitesimal changes:

$$\begin{aligned} \Delta \ln y_{it} &= \Delta \ln A_{it} + \beta_t \theta_{il} (\Delta \ln L_{it} - \Delta \ln K_{it}) + \mu \Delta \ln K_{it} \\ &= \alpha_i + \beta_t \theta_{il} (\Delta \ln L_{it} - \Delta \ln K_{it}) + \mu \Delta \ln K_{it} + \varepsilon_{it}, \end{aligned} \quad (11.5')$$

where in the second line we replace the true productivity change $\Delta \ln A_{it}$ by $\alpha_i + \varepsilon_{it}$, composed of a firm fixed-effect α_i and a random component ε_{it} . The estimated ratio of price to marginal cost, $\beta_t \equiv \eta_t/(\eta_t - 1)$, is treated as common across firms $i = 1, \dots, N$ within each industry but is allowed to change in the pre- and post-liberalization sample. Thus, by examining the change in β_t we can determine the effect of liberalization on markups.

Estimation is performed with both OLS and instrumental variables. Initial estimates generally provided only insignificant changes in the estimated price-cost ratios. So rather than just letting the markups β_t change over time, Harrison instead interacts this variable with industry-level estimates of import penetration. Specifically, let $\beta_t = \beta_0 + \beta_1 M_t$, where M_t is the ratio of imports to consumption in each industry. Substituting this into (11.5'), we obtain an interaction term between M_t and $\theta_{il} (\Delta \ln L_{it} - \Delta \ln K_{it})$, with the coefficient β_1 . We expect that $\beta_1 < 0$, so that higher imports lead to lower markups. Harrison obtains an estimate of -0.25 for β_1 over the entire sample (standard error of 0.05). This indicates, for example, that an increase in import penetration from 0 to 50 percent would reduce the ratio of price to marginal cost by 12 percent. She also finds that reduced tariffs are associated with lower price-cost ratios, but that result is only significant at the 15 percent level. So the connection between import liberalization and reduced markups is weakly confirmed in her study.

Notice that in our derivation of (11.5) we *did not* assume that the rental on capital is observable, but only used the labor share. If instead the rental is observable, then the capital share is $\theta_{ik} = rK_i/p_i y_i = [(\eta_i - 1)/\eta_i] [f_{ik} K_i/f(L_i, K_i)]$, analogous to (11.4). Substituting these two equations into (11.1), we obtain an alternative relationship between output growth and true productivity change⁶

$$\hat{y}_i = \hat{A}_i + \left(\frac{\eta_i}{\eta_i - 1} \right) (\theta_{il} \hat{L}_i + \theta_{ik} \hat{K}_i). \quad (11.6)$$

⁶Notice that (11.5) and (11.6) are identical if there are zero profits in the industry, so that price equals average cost. In that case, the ratio of average cost to marginal cost, which equals the returns to scale μ_i , also equals the ratio of price to marginal cost, which is $\eta_i/(\eta_i - 1)$. In addition, the labor and capital shares sum to unity, so that $\theta_{ik} = 1 - \theta_{il}$.

Levinsohn (1993) applies equation (11.6), written using discrete changes, to firm-level data for Turkey, which embarked on an ambitious liberalization program in 1980. He replaces the true productivity change $\Delta \ln A_{it}$ by $\alpha_t + \varepsilon_{it}$, composed of a year fixed-effect α_t that is common across firms and a random component ε_{it} . Once again, the estimated ratio of price to marginal cost, $\beta_t \equiv \eta_t / (\eta_t - 1)$, is treated as common across firms $i = 1, \dots, N$ within each industry but is allowed to vary over time, as liberalization proceeds. In his results, Levinsohn finds a strong connection between the industries that experience liberalization and those where the estimated price-cost ratios are reduced: in three industries experiencing significant liberalization, the estimated ratios fell; whereas in two other industries experiencing increased protection, the ratios rose; and in the remaining five industries where the ratios were insignificantly different from unity (i.e., the industries are perfectly competitive), there was no clear pattern to the changes in the point estimates of β_t . So for both the Ivory Coast and Turkey, the evidence supports the hypothesis that import liberalization reduces markups.

We noted above that estimation of (11.5') requires instrumental variables because the error ε_{it} is likely correlated with factor inputs on the right: a firm that experiences a boost to productivity will hire more inputs. Appropriate instruments should be correlated with factor usage by the firm, but not with productivity, and for this purpose Harrison uses the nominal exchange rate, the price of energy, sectoral real wages, and the firms' reported debt. The endogeneity of inputs and the appropriate methods to correct for it are addressed by Hall (1988), Olley and Pakes (1996), and Levinsohn and Petrin (2003). More recently, De Loecker (2009) has shown how the Olley-Pakes technique can be *combined* with the Hall technique to obtain estimates of the markups of firms from production data. De Loecker (2011), De Loecker and Warzynski (2012), and De Loecker, Goldberg, Khandelwal, and Pavcnik (2014) have applied these techniques to isolate the impact of international trade in the markups charged by firms. The latter authors find that tariff reductions in India led to reduced markups for import-competing firms, but also led to *increased* markups for those firms relying on imported intermediate inputs who therefore pay lower prices for those inputs.

IMMISERIZING GROWTH

We now turn to the link between growth and the terms of trade, and the case of “immiserizing growth” due to Bhagwati (1958). This case is illustrated in figure 11.1, where good 1 is the exportable and good 2 is the importable. With the initial production possibility frontier (PPF), the economy produces at point B and consumes at point C . Due to growth, the PPF shifts out. If the terms of trade were unchanged, then consumption would change to a point like C' and the representative consumer would be better off. With a fall in the relative price of the export good, however, it is possible that consumption could instead be at point C'' with production at B'' . This illustrates the borderline case where utility is unchanged due to growth, and with any further fall in the terms of trade, the representative consumer would be worse off.

We are interested in solving for the conditions under which immiserizing growth can occur. To this end, it will be convenient to first solve for the conditions such that utility is *constant*, as illustrated in figure 11.1. For simplicity we suppose that there are just two goods and choose the importable y_2 as the numeraire, so its price is unity. Let $G(p, \alpha) = py_1 + y_2$ denote the GDP function for the economy, where α is a scalar that represents a shift parameter for the PPF. For example, α can represent a factor

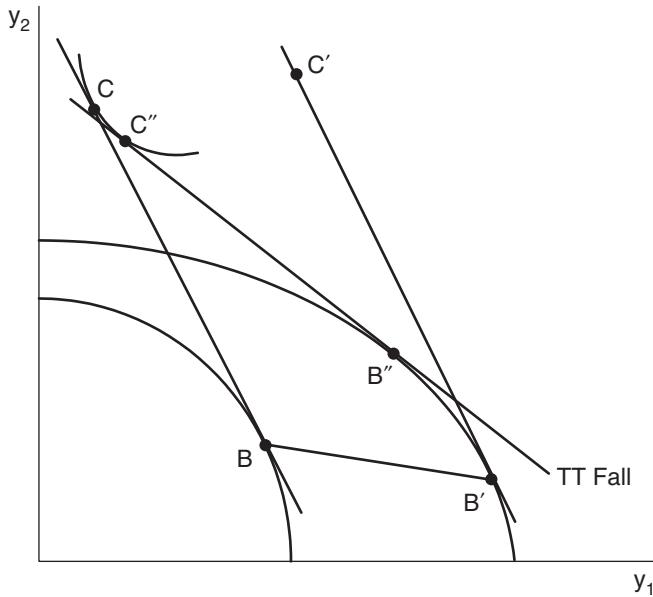


Figure 11.1

endowment or technological progress in some industry. The total change in GDP is measured by

$$dG = \frac{\partial G}{\partial p} dp + \frac{\partial G}{\partial \alpha} d\alpha = y_1 dp + \left(p \frac{dy_1}{d\alpha} + \frac{dy_2}{d\alpha} \right) d\alpha. \quad (11.7)$$

We presume that $\frac{\partial G}{\partial \alpha} = (p \frac{dy_1}{d\alpha} + \frac{dy_2}{d\alpha}) > 0$ and $\frac{dy_1}{d\alpha} > 0$, meaning that growth at constant prices increases both GDP and production of the exportable.

As in chapter 8, we will allow for different consumers but suppose that the numeraire good is additively separable in consumption, so that consumer utilities can be summed. Total social welfare is then $W[p, G(p, \alpha)]$, which serves as an indirect utility function for the economy, where $\partial W/\partial p = -c_1 = -d_1(p)$ is the (negative of) consumption of good 1 and $\partial W/\partial G \equiv 1$.⁷ Then social welfare is constant due to growth if and only if

$$dW = \frac{\partial W}{\partial p} dp + \left(\frac{\partial G}{\partial p} dp + \frac{\partial G}{\partial \alpha} d\alpha \right) = (y_1 - c_1) dp + \frac{\partial G}{\partial \alpha} d\alpha = 0. \quad (11.8)$$

Thus, the drop in the export price that will just keep welfare constant is

$$dp = \frac{\partial G}{\partial \alpha} d\alpha / (c_1 - y_1). \quad (11.9)$$

⁷See problem 8.1 to derive the demand functions for an individual with quasi-linear utility.

Next, we solve for the equilibrium change in the relative price of exports and compare that to (11.9). Equilibrium in the export market means that $(y_1 - c_1) = m_1^*$, where m_1^* is import demand from the rest of the world. Totally differentiating this, we obtain

$$\frac{dy_1}{d\alpha} d\alpha + \left(\frac{\partial y_1}{\partial p} - \frac{\partial c_1}{\partial p} \right) dp = \frac{dm_1^*}{dp} dp.$$

Thus, the equilibrium change in the price of exports is

$$dp = \frac{\frac{dy_1}{d\alpha} d\alpha}{\left[\frac{dm_1^*}{dp} - \left(\frac{\partial y_1}{\partial p} - \frac{\partial c_1}{\partial p} \right) \right]}. \quad (11.10)$$

The denominator of (11.10) is negative while the numerator is positive, so this expression indicates the drop in the export price due to growth. Welfare is constant if (11.9) = (11.10), and welfare falls if (11.10) < (11.9). This fall will occur if and only if

$$\frac{dy_1}{d\alpha} (y_1 - c_1) / \frac{\partial G}{\partial \alpha} > \left[\left(\frac{\partial y_1}{\partial p} - \frac{\partial c_1}{\partial p} \right) - \frac{dm_1^*}{dp} \right]. \quad (11.11)$$

Dividing this equation through by $(y_1 - c_1) = m_1^*$, and making use of $\frac{\partial G}{\partial \alpha} = (p \frac{dy_1}{d\alpha} + \frac{dy_2}{d\alpha})$, we can express this necessary and sufficient condition for immiserizing growth as

$$p \frac{dy_1}{d\alpha} / \left(p \frac{dy_1}{d\alpha} + \frac{dy_2}{d\alpha} \right) > \left[\left(\frac{\partial y_1}{\partial p} - \frac{\partial c_1}{\partial p} \right) \frac{p}{m_1^*} - \frac{dm_1^*}{dp} \frac{p}{m_1^*} \right]. \quad (11.12)$$

To interpret this expression, note that the first term on the right-hand side is the elasticity of export supply and is positive. For immiserizing growth to occur, it is therefore necessary that the term on the left-hand side *exceeds* the amount $-\left(\frac{dm_1^*}{dp} \frac{p}{m_1^*}\right)$, which is the elasticity of foreign demand for imports. If this elasticity is less than unity, that is, foreign demand is inelastic, then immiserizing growth can occur even when the expression on the left side of (11.12) is also less than unity. However, if foreign demand for imports is elastic, then for (11.12) to be satisfied it must also be the case that the expression on the left side exceeds unity. By inspection, this occurs if and only if $dy_2/d\alpha < 0$, that is, when growth reduces the output of good 2 (at constant prices). Summarizing these results, we have the following theorem.

THEOREM (BHAGWATI 1958)

Immiserizing growth occurs if and only if (11.12) holds. Necessary conditions for this are that either (a) the foreign demand for imports is inelastic, or (b) growth reduces the output of the importable good (at constant prices).

Condition (a) is not surprising, since having inelastic demand for a product is the same condition that would permit farmers to be worse off after a bumper crop, as in the example given in introductory economics textbooks. What is more surprising is

that immiserizing growth can occur even when foreign demand is *not* inelastic, but this requires that condition (b) hold, that is, that growth reduces the output of the importable good. We know from the Rybczynski theorem that this condition would hold in the two-by-two model if the factor used intensively in exports had grown. So this is one scenario when immiserizing growth can occur. Are there others?

Findlay and Grubert (1959) answer this question in the affirmative. They considered Hicks neutral technological progress in good 1. To demonstrate the effects of this, we work with the GDP function, defined as

$$G(pA, V) \equiv \max_{v_i \geq 0} pA_1 f_1(v_1) + f_2(v_2) \text{ subject to } v_1 + v_2 = V. \quad (11.13)$$

That is, the parameter α that appeared in the earlier GDP function now equals A_1 , which is the Hicks neutral productivity parameter on good 1. Notice that pA_1 enters in a multiplicative form as an argument of the GDP function, since that is how it appears in the objective function.

We have demonstrated in chapter 1 that the derivative of the GDP function with respect to price equals the output of that good, or $\partial G / \partial p = y_1$. Letting $G_1(pA_1, V)$ denote the partial derivative of G with respect to its first argument, we therefore have

$$y_1 = \frac{\partial G}{\partial p} = \frac{\partial G}{\partial(pA_1)} \frac{d(pA_1)}{dp} = G_1(pA_1, V) A_1. \quad (11.14)$$

Totally differentiating this expression, we obtain $dy_1 = G_1 dA_1 + A_1 G_{11}(pdA_1 + A_1 dp)$. Dividing through by $y_1 = G_1 A_1$, it follows that if *prices are constant*, then

$$\hat{y}_1 = \hat{A}_1 + \left(\frac{pA_1 G_{11}}{G_1} \right) \hat{A}_1 \geq \hat{A}_1, \quad (11.15)$$

where the inequality is obtained because the GDP function is convex in prices, so that $G_{11} \geq 0$.

If we instead consider the cases where A_1 is constant and p changes, then it is readily shown that $(pA_1 G_{11}/G_1) = (\partial y_1 / \partial p)(p/y_1) \geq 0$, which is the elasticity of supply in industry 1. When this inequality is strict, we obtain $\hat{y}_1 > \hat{A}_1$, so that the output of industry 1 increases by *more than* technological progress in the industry. The intuition for these results is that technological progress raises factor prices, so that industry 1 attracts resources from the rest of the economy.

This result is illustrated in figure 11.2, where with *unchanged* inputs industry 1 would expand by \hat{A}_1 due to the technological progress, but after attracting resources from the rest of the economy, it expands by $\hat{y}_1 > \hat{A}_1$. Moreover, as illustrated in figure 11.2, this will imply that *output in industry 2 must contract*, because it is losing factor inputs. Thus, there is a “Rybczynski-like” effect at work even when the source of growth is neutral technological progress in one industry, leading to a contraction of the other industry (at constant prices).⁸ This type of growth therefore satisfies condition (b) of the above theorem, making immiserizing growth possible.

The result in (11.13) that the GDP function should have price multiplied by the Hicks neutral productivity parameter in each industry was already used in chapter

⁸In problem 11.1 you are asked to provide two other proofs of this result, using Jones algebra and also graphically.

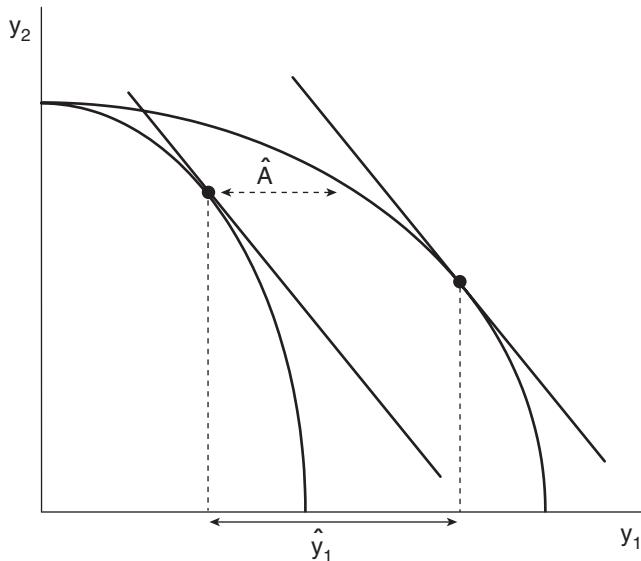


Figure 11.2

3 (see equation 3.24), where we reviewed the work of Harrigan (1997). He estimated a translog GDP function over the OECD countries while allowing for technology to differ across countries. Industry prices were treated as equal across countries (due to free trade), so it was *only* the technology parameters and factor endowments that appeared in the GDP function. The result in (11.15) that output expands by more than productivity in an industry then becomes a testable hypothesis: it will hold empirically provided that certain parameters of the translog GDP function satisfy appropriate conditions.⁹ Thus, by estimating the GDP function we are able to test this result of Findlay and Grubert (1959). For example, Kee (2001) finds that the industry principally benefiting from productivity growth in Singapore is electronics, where TFP growth is 6.4 percent per year on average over 1970–99, whereas output growth is 15 percent per year. The difference between these is accounted for by labor and capital drawn into that sector, including foreign capital.

ENDOGENOUS GROWTH

Our discussion in the previous section did not model growth in any meaningful way, but simply treated it as a comparative statics increase in productivity or a factor endowment.

Obviously, this is unsatisfactory. The celebrated model of Solow (1956) specifies that capital accumulation should depend on investment, which equals savings in the autarky one-sector economy, so that capital would gradually rise to its steady-state level. But that model still assumes that technological progress is *exogenous*, and a positive rate of technological progress or growth in the labor force is needed to obtain

⁹See problem 11.2.

steady-state growth. The same assumptions are used in two-sector, open economy versions of this model, such as Smith (1977). Ultimately, interest in the profession has turned to models that instead allow for an *endogenous* rate of technological innovation, and we shall outline one of these models, taken from Grossman and Helpman (1990; 1991, chapters 3, 8, and 9).

Many of the “endogenous growth” models build upon the monopolistic competition framework we introduced in chapter 5, but rather than thinking of differentiated final products, we instead consider *differentiated intermediate inputs*. The idea is that an increase in the variety (N) of differentiated inputs will allow for an increase in output, much like an increase in variety of final goods allowed for higher consumer utility in chapter 5.¹⁰ To make this precise, suppose that there is a single final good, with output y , produced with the CES production function

$$y = \left[\sum_{i=1}^N x_i^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \quad (11.16)$$

where x_i is the quantity of input variety $i = 1, \dots, N$. As usual, we assume $\sigma > 1$, so that it is meaningful to think of changes in the number of inputs.¹¹

If inputs are all priced equally in equilibrium, then their quantities are also equal, $x_i = x$, and so (11.16) can be rewritten as

$$y = N^{\sigma/(\sigma-1)} x = N^{1/(1-\sigma)} X, \quad (11.17)$$

with $X \equiv Nx$. We can think of X as the “aggregate” amount of the intermediate input, and holding this magnitude fixed, (11.17) shows that *increases in N raise output y* (since $\sigma > 1$).

This amounts to a productivity gain in the final-good industry, due to the proliferation of differentiated intermediate inputs. Indeed, we could relabel $N^{1/(\sigma-1)}$ on the right side of (11.17) as A to obtain the simple production function $y = AX$, where technological progress A will depend on the number of new inputs.¹² By carefully modeling the dynamics of how N evolves due to the entry of new firms, we therefore have a story of endogenous technological progress, or endogenous growth.

Autarky Equilibrium

We begin by describing the autarky equilibrium. The price of the final good at each point in time is $P(t)$. If the consumer spends $E(t)$ on this good, then the quantity $y(t) = E(t)/P(t)$ will be purchased, and we assume that this provides instantaneous utility of $\ln y(t) = \ln E(t) - \ln P(t)$. The consumer’s problem is then

$$\max_{E(\tau)} \int_t^\infty e^{-\rho(\tau-t)} [\ln E(\tau) - \ln P(\tau)] d\tau, \quad (11.18)$$

¹⁰See problem 5.1. The use of differentiated inputs to generate economy-wide returns to scale is often attributed to Ethier (1979).

¹¹If $\sigma < 1$, then output is zero whenever any input is zero. Thus, all inputs are essential to production, and we cannot think of introducing new inputs. In contrast, when $\sigma > 1$, output will increase by a finite amount when new inputs are introduced, as shown in (11.17).

¹²This is like the so-called “ $y = AK$ ” production function used in some macroeconomic models.

subject to the budget constraint

$$\int_t^\infty e^{-[R(\tau)-R(t)]} E(\tau) d\tau \leq \int_t^\infty e^{-[R(\tau)-R(t)]} w(\tau) L d\tau + B(t), \quad (11.19)$$

where $R(t)$ is the cumulative interest rate from time 0 to time t , so that $\dot{R}(t) \equiv dR/dt$ is the instantaneous interest rate. This budget constraint states that the discounted value of labor income $w(t)L$, plus initial assets $B(t)$, cannot exceed the discounted value of expenditure. The solution to this optimization problem is¹³

$$\frac{\dot{E}}{E} = \dot{R} - \rho, \quad (11.20)$$

where we omit the time index on variables so this will not cause confusion.

Condition (11.20) states that when the nominal interest rate \dot{R} exceeds the consumer's discount rate, then expenditure is rising. This is because with a high interest rate it is optimal to put a large amount of assets into savings initially, and gradually increase consumption thereafter. Generally, the nominal interest $\dot{R}(t)$ will depend on the normalization made on prices. Grossman and Helpman (1990, 1991) choose the normalization $E(t) \equiv 1$, which implies from (11.20) that $\dot{R}(t) \equiv \rho$, as we shall also use.

On the production side, the final good y is manufactured under perfect competition with the production function (11.17), where x_i is the quantity of each of N intermediate inputs (all these variables depend on t). These inputs are produced under monopolistic competition, and each unit of x_i requires one unit of labor, which is the only input. It follows that the price of each input is

$$p_i \left(1 - \frac{1}{\sigma}\right) = w, \quad \text{or} \quad p_i = w \left(\frac{\sigma}{\sigma-1}\right), \quad (11.21)$$

where w is the wage (which depends on t).

Input-producing firms have a fixed cost only when they begin production. The key question is how should this fixed cost be modeled? If we presume that the fixed cost is a fixed amount of labor α , as in the static monopolistic competition model, then there will be a limit to the number of products that could be profitably developed. In other words, growth of inputs $N(t)$ will cease at some point in time. The first dynamic monopolistic competition models had this feature (Judd 1985; Grossman and Helpman 1989), while the work of Romer (1990) and Grossman and Helpman (1990, 1991) avoided this outcome by presuming that fixed costs are *inversely proportional* to the number of products already developed, so that fixed labor costs are $\alpha/N(t)$. The idea is that the creation of products contributes to a "public stock of knowledge," which makes the invention of new products even easier. Notice that in a two-country model, this will allow for the potential spillover of public knowledge across borders, which will turn out to be very important.

With fixed labor costs of $\alpha/N(t)$, the nominal fixed costs are $\alpha w(t)/N(t)$. These are financed by consumers purchasing equity in the firms, which provides both dividends and capital gains. The instantaneous profits of the firms are

¹³As solved for in problem 11.3.

$$\pi_i = (p_i - w)x_i = \left(\frac{1}{\sigma-1}\right)wx, \quad (11.22)$$

using (11.21) and symmetry of the equilibrium, so that $x_i = x$. The zero-profit conditions for the firms are that the present discounted value of instantaneous profits in (11.22) must equal the fixed costs $\alpha w(t)/N(t)$,

$$V(t) \equiv \int_t^{\infty} e^{-\rho(\tau-t)} \left(\frac{1}{\sigma-1}\right) w(\tau) x(\tau) d\tau = \frac{\alpha w(t)}{N(t)}, \quad \text{for all } t. \quad (11.23)$$

Since (11.23) holds for all t , we can differentiate it with respect to t , obtaining

$$\dot{V} = -\left(\frac{1}{\sigma-1}\right) w(t) x(t) + \rho \int_t^{\infty} e^{-\rho(\tau-t)} \left(\frac{1}{\sigma-1}\right) w(\tau) x(\tau) d\tau = \frac{\alpha \dot{w}}{N} - \frac{\alpha w \dot{N}}{N^2}. \quad (11.24)$$

To simplify (11.24), we can divide by the fixed costs $\alpha w/N$, obtaining

$$\underbrace{\left(\frac{1}{\sigma-1}\right) \frac{Nx}{\alpha}}_{\pi/\text{Fixed cost}} + \underbrace{\left(\frac{\dot{w}}{w} - \frac{\dot{N}}{N}\right)}_{\dot{V}/V} = \underbrace{\rho}_{\text{interest rate}}. \quad (11.25)$$

This is the key arbitrage condition of the endogenous growth model with expanding input variety, as derived by Grossman and Helpman (1990, 1991). The left-hand side of (11.25) is the sum of dividends plus capital gains, which should equal the discount rate on the right-hand side, which we interpret as the real interest rate.

With (11.25) in hand we can complete the model quite easily by using the full-employment condition for the economy. This can be written as

$$L = Nx + \left(\frac{\alpha}{N}\right) \dot{N}. \quad (11.26)$$

The left-hand side of (11.26) is the fixed endowment of labor, and the right-hand side equals the sum of labor used in production, $Nx = X$, and labor used in R&D, which is the fixed cost (α/N) times the growth of new products. Let us define the *growth rate* of inputs as $g \equiv \dot{N}/N$. Then we can rewrite (11.26) as $L = X + \alpha g$, or $X = L - \alpha g$. Substituting this into (11.25) we obtain

$$\left(\frac{1}{\sigma-1}\right) \left(\frac{L-\alpha g}{\alpha}\right) + \left(\frac{\dot{w}}{w} - g\right) = \rho. \quad (11.27)$$

Now consider a steady-state solution with $\dot{w} = 0$. Substituting this into (11.27), we can readily solve for the autarky growth rate as

$$g^a = \left(\frac{1}{\sigma}\right) \left[\frac{L}{\alpha} - (\sigma-1)\rho \right]. \quad (11.28)$$

Thus, the economy achieves steady-state growth even with a fixed labor supply, provided that $(L/\alpha) > (\sigma-1)\rho$. Notice that with the growth rate fixed at g^a , the labor devoted to production is also fixed at $X = L - \alpha g^a$. Then from (11.17) we can solve for

the growth rate of GDP as $\dot{y}/y = g^a/(\sigma - 1)$. So along with the continual growth of new products, there is also continual growth of GDP and utility.¹⁴

Trade Equilibrium with Knowledge Spillovers

Let us compare the autarky growth rate in (11.28) with that achieved under free trade between the two countries. Suppose that the foreign country has the same production function as at home and differs only in its labor endowment L^* . We will suppose that the final goods produced in the two countries are imperfect substitutes, with the instantaneous utility function in each country:

$$U(y, y^*) = \ln[y^{(\eta-1)/\eta} + (y^*)^{(\eta-1)/\eta}]^{\eta/(\eta-1)}. \quad (11.29)$$

This is the log of a CES function with elasticity of substitution $\eta > 1$. We suppose that the final goods from home and abroad are freely traded, at the prices p_y and p_y^* , respectively.

Given expenditure E in the home country, it follows that the demand for the final goods of each country is¹⁵

$$y = (p_y/P)^{-\eta}(E/P), \quad \text{and} \quad y^* = (p_y^*/P)^{-\eta}(E/P), \quad (11.30)$$

where P refers to the overall price index of the final goods, defined as

$$P(p_y, p_y^*) = [p_y^{1-\eta} + (p_y^*)^{1-\eta}]^{1/(1-\eta)}. \quad (11.31)$$

Demand for the final goods from the foreign county is similar, and is obtained by just replacing home expenditure E by foreign expenditure E^* in (11.30).

Using (11.30), it is readily confirmed that utility in (11.29) equals $U(y, y^*) = \ln(E/P)$, or the log of real expenditures. We will suppose that the capital markets are fully integrated, so there is a single, cumulative world interest rate $R(t)$. Maximizing the presented discounted sum of home utility gives us the same objective function as in (11.18), so the first-order condition (11.20) still applies. If we again use the normalization $E(t) \equiv 1$, the real interest rate equals ρ , which is the discount rate in both countries.

With this setup, we want to contrast the effects of trade in two scenarios. In the first, we will suppose that in addition to free trade in the final goods and integrated capital markets, there is *also* free trade in the intermediate inputs and complete transfer of knowledge across countries. The latter assumption means that the fixed costs of creating a new product in either country are $\alpha/[N(t) + N^*(t)]$. Under this set of assumptions, the integrated world equilibrium with the two countries is simply a “blown up” version of either country in autarky. The effect of this increase in size is to *raise* the growth rate from that shown in (11.28) to

$$g^w = \left(\frac{1}{\sigma}\right) \left[\frac{(L + L^*)}{\alpha} - (\sigma - 1)\rho \right]. \quad (11.32)$$

¹⁴While the economy achieves a continual growth of utility, it is not actually the case that welfare is maximized. That would require a choice of the growth rate different from g^a in (11.28). See problem 11.4.

¹⁵See (5.24) and (5.25), and also problem 5.2.

The result that the growth rate increases in proportion to the size of the world economy is referred to as a “scale effect,” and is the dynamic analogue to the static gains from trade that we saw in Krugman’s model in the beginning of chapter 5. This result has come under some criticism in the growth literature, not because of the result that trade increases growth, but rather, because of another result, that subsidies to R&D permanently raise the growth rate. Both of these propositions depend very strongly on our specification that fixed costs are *inverse proportional* to the number of products already developed. If this strict inverse proportionality does not hold, for example, if we specified fixed costs as α/N^β , with $0 < \beta < 1$, then new product development in the absence of population growth will eventually stop. Jones (1995a,b) has referred to models of this type as “semi-endogenous” growth, and they imply that free trade or subsidies to R&D have only temporary effects on the growth rate. Empirical work questions which class of models is most realistic, as we shall discuss later.

Trade Equilibrium without Knowledge Spillovers

Let us now contrast the results from a fully integrated world economy to the case where there are *zero* spillovers of knowledge across borders, as in Grossman and Helpman (1991, chapter 8) and Feenstra (1996). This means that the fixed costs in each country are $\alpha/N(t)$ and $\alpha/N^*(t)$, respectively. Initially, we suppose that there is no trade in intermediate inputs, so that only trade in final goods and financial capital is permitted; later in the section, we will reintroduce free trade in intermediate inputs. We shall demonstrate that in the absence of knowledge spillovers across borders, free trade *will not* increase the growth rate for both countries, but rather will slow down product development in the smaller country.

With the countries growing at different rates, we cannot simultaneously use the normalizations $E(t) \equiv 1$ and $E^*(t) \equiv 1$, so it is no longer the case that $\dot{R}(t) \equiv \rho$. Accordingly, the profits of input-producing firms in (11.23) are discounted using the cumulative interest rate $R(t)$, and the arbitrage condition in each country is rewritten from (11.25) as

$$\left(\frac{1}{\sigma-1}\right)\frac{Nx}{\alpha} + \left(\frac{\dot{w}}{w} - \frac{\dot{N}}{N}\right) = \dot{R}, \quad (11.33)$$

and,

$$\left(\frac{1}{\sigma-1}\right)\frac{N^*x^*}{\alpha} + \left(\frac{\dot{w}^*}{w^*} - \frac{\dot{N}^*}{N^*}\right) = \dot{R}, \quad (11.34)$$

As before, we will use the notation $X \equiv Nx$ and $X^* \equiv N^*x^*$ to denote the amount of labor devoted to production of intermediate inputs in each country. In addition, let $g \equiv (\dot{N}/N)$ and $g^* \equiv (\dot{N}^*/N^*)$ denote the home and foreign growth rates of new inputs. The full-employment condition in each country implies that $X = L - \alpha g$ and $X^* = L^* - \alpha g^*$, and substituting these into (11.33) and (11.34), we can derive the growth rates:

$$g = \left(\frac{L}{\alpha\sigma}\right) - \left(\frac{\sigma-1}{\sigma}\right)\left[\dot{R} - \left(\frac{\dot{w}}{w}\right)\right], \quad \text{and} \quad g^* = \left(\frac{L^*}{\alpha\sigma}\right) - \left(\frac{\sigma-1}{\sigma}\right)\left[\dot{R} - \left(\frac{\dot{w}^*}{w^*}\right)\right]. \quad (11.35)$$

These equations show that the growth rates are *inversely related* to the real interest rates $[\dot{R} - (\dot{w}/w)]$ and $[\dot{R} - (\dot{w}^*/w^*)]$: having a higher nominal interest rate \dot{R} lowers the

discounted value of profits and therefore expenditure on R&D; conversely, having a rising path of wages and prices leads to rising profits and a lower real interest rate, with higher R&D. Taking the difference between the two growth rates in (11.35), we obtain

$$g - g^* = \left(\frac{L - L^*}{\alpha\sigma} \right) + \left(\frac{\sigma - 1}{\sigma} \right) \left(\frac{\dot{w}}{w} - \frac{\dot{w}^*}{w^*} \right). \quad (11.36)$$

Let us assume henceforth that the home country is larger, $L > L^*$. From (11.36) we see that this difference in size will tend to be associated with a faster growth rate at home, $g > g^*$. The growth rates are also affected, however, by the change in wages on the right-hand side. In the autarky equilibrium discussed above, we had focused on the steady-state solution with $\dot{w} = 0$. But now it is impossible to assume this for both countries: assuming $\dot{w} = \dot{w}^* = 0$ implies that $g > g^*$ from (11.36), and we will argue below that this implies rising relative wages for the home country, $\dot{w}/w > \dot{w}^*/w^*$, which is a contradiction. Thus, the assumption of a steady-state equilibrium is inconsistent with the dynamic equations for the two countries. Accordingly, we need to solve for the growth rates outside of the steady state.

We begin by solving for the change in wages, from the production function $y = N^{1/(\sigma-1)}X$ in (11.17), where X is the amount of intermediate inputs purchased at the price of $p_i = w\sigma/(\sigma - 1)$. It follows that the marginal cost of producing the final good is $p_y = p_i/N^{1/(\sigma-1)} = w\sigma/[(\sigma - 1)N^{1/(\sigma-1)}]$ at home, and $p_y^* = w^*\sigma/[(\sigma - 1)(N^*)^{1/(\sigma-1)}]$ abroad. Differentiating these and taking the difference, we obtain

$$\left(\frac{\dot{p}_y}{p_y} \right) - \left(\frac{\dot{p}_y^*}{p_y^*} \right) = \left(\frac{\dot{w}}{w} - \frac{\dot{w}^*}{w^*} \right) - \left(\frac{g - g^*}{\sigma - 1} \right). \quad (11.37)$$

This equation states that a higher growth rate at home will tend to be associated with rising relative wages, but this also depends on the changes in the prices of the final goods in each country.

To determine these prices, we make use of the CES demands in (11.30), which imply that $(p_y/p_y^*) = (y/y^*)^{-1/\eta}$. Differentiating this equation using the production functions $y = N^{1/(\sigma-1)}X$ and $y^* = (N^*)^{1/(\sigma-1)}X^*$, we obtain

$$\left(\frac{\dot{p}_y}{p_y} \right) - \left(\frac{\dot{p}_y^*}{p_y^*} \right) = -\frac{1}{\eta} \left[\left(\frac{g - g^*}{\sigma - 1} \right) + \frac{\dot{X}}{X} - \frac{\dot{X}^*}{X^*} \right]. \quad (11.38)$$

Then combining (11.36)–(11.38), we can derive the fundamental relation

$$g - g^* = \Delta \left(\frac{L - L^*}{\alpha\sigma} \right) - \Delta \left(\frac{\sigma - 1}{\eta\sigma} \right) \left(\frac{\dot{X}}{X} - \frac{\dot{X}^*}{X^*} \right), \quad (11.39)$$

where

$$\Delta \equiv \left[1 - \frac{(\eta - 1)}{\eta\sigma} \right]^{-1} > 1. \quad (11.40)$$

Equation (11.39) shows that the difference in the growth of new inputs across the countries depends on the difference in their size. There is also another term on the right-hand side of (11.39), depending on \dot{X} and \dot{X}^* , but we can safely presume that

this term approaches zero as $t \rightarrow \infty$. Notice that the difference in the *autarky* growth rates of the two countries is exactly $(L - L^*)/\alpha\sigma$, but in (11.39), the limiting growth rates of the countries will differ by this *times* the amount $\Delta > 1$. Therefore, in the absence of knowledge spillovers, the effect of trade is to *magnify* the initial difference in the growth rates of the two countries.

We have not yet determined whether trade will increase the growth rate of the large country, or slow down the growth rate of the small country. Some additional calculations show that the first case is not possible: the growth rate of the large country must approach its autarky rate of g^a as $t \rightarrow \infty$.¹⁶ Our results are then summarized by the following theorem.

THEOREM (FEENSTRA 1996)

- (a) $\lim_{t \rightarrow \infty} g = g^a$, which equals the autarky growth rate g^a of the large country.
- (b) $\lim_{t \rightarrow \infty} g^* = g^{a*} - (\Delta - 1)(L - L^*)/\alpha\sigma$, which is less than the autarky growth rate g^{a*} of the small country, $g^{a*} = [(L^*/\alpha) - (\sigma - 1)\rho]/\sigma$.
- (c) $\lim_{t \rightarrow \infty} (\dot{w}/w) - (\dot{w}^*/w^*) > 0$, and $\lim_{t \rightarrow \infty} (\dot{p}_y/p_y) - (\dot{p}_y^*/p_y^*) < 0$.

Parts (a) and (b) of this theorem have already been discussed above. The results in (c) follow from substituting the limiting growth rates into (11.37) and (11.38). Relative wages move in favor of the larger country, which reflects the rising productivity of that country due to the faster-growing number of intermediate inputs. But this same efficiency gain leads to a fall in the relative price of final goods. These different wage and price movements emphasize how difficult it is to make a general prediction of the terms of trade movement in this model: it depends on which goods/factors are used to measure the terms of trade.

The finding that the small country has its rate of product development slowed down does not necessarily translate into a welfare loss, though it might. That will depend on whether intermediate inputs are traded or not. Initially let us continue with our assumption that inputs are not traded. The welfare obtained by consumers in the small country will depend, of course, on the price index $P(p_y, p_y^*)$ for purchases of the final goods. Let us choose the foreign wage as the numeraire, $w^* \equiv 1$. Then since the number of intermediate inputs produced abroad is slowed down by trade, the price of its final good p_y^* is higher than in autarky. This will create a potential welfare loss. But the foreign consumers *also* have the final good from the home country available, which creates a potential welfare gain. Which of these effects dominate will depend on a comparison of elasticities: if $\eta > (\sigma + 1)$, then there is a potential welfare loss; but if $\eta \leq (\sigma + 1)$, then there is a likely welfare gain.¹⁷

What about when the intermediate inputs are traded? Then the argument is quite different. Despite the fact that product development has been slowed in the small

¹⁶ See problem 11.5.

¹⁷ Specifically, Feenstra (1996, Proposition 2) finds that the price index under free trade is less than that under autarky when $\eta \leq (\sigma + 1)$, and when a weak additional condition is satisfied. In contrast, the finding that welfare in the smaller country can potentially fall is related to the fact that R&D is being undertaken at less than the socially optimal rate in autarky, as demonstrated in problem 11.4. Then trade in final goods but not in intermediate inputs worsens that distortion in that country. Markusen (1989) also discusses how trade in final goods, but not in intermediate inputs, can lead to a welfare loss even in a static model.

country, it can still purchase intermediate inputs from abroad. If t is sufficiently large, it will definitely be the case that the efficiency gain from the additional imported inputs more than offsets the loss from the reduced number of local inputs, so that the price of the final good p_y^* is *lower* than in autarky. Thus, consumers in the small country benefit from a reduction in the price of their own final good as well as the imported final good, and so there are welfare gains for both reasons.

It is worth emphasizing that our finding that trade can slow down the rate of product development occurs only when there are *zero* spillovers of knowledge. Provided that there are some positive spillovers, Grossman and Helpman (1991, chapter 9) and Rivera-Batiz and Romer (1991a,b) argue that product innovation in both countries will increase. Indeed, this occurs even if knowledge spillovers are the *only* international flows, without any trade in final goods or inputs. So the assumption of zero international spillovers we have used above is an extreme case, but it serves to highlight how important these spillovers are to growth.

GROWTH WITHOUT SCALE EFFECTS

An implication of the endogenous growth models is that “size matters” for growth. This follows directly from the equation for autarky growth rates in (11.28), where a higher labor stock (measured in units of effective R&D workers, or L/α) leads to a higher growth rate. This implication is analogous to what we found in Krugman’s static model reviewed at the beginning of chapter 5: doubling a country’s size would lead to efficiency gains through economies of scale, and additional consumer gains through increased variety. In the dynamic model outlined above, increasing a country’s size leads to increased variety of intermediate inputs, which results in efficiency gains in producing the final good and also reduces the fixed costs $\alpha/N(t)$ of inventing new inputs. The latter reduction in fixed costs is what leads to a permanently higher growth rate due to larger country size.

How realistic is this result? Over very long time periods, it is perhaps reasonable to think that economies will grow in proportion to their size. Kremer (1993) considers the period from one million years ago to the present, and finds that the growth rate of population is proportional to its level. Over shorter time spans, however, the implications of the “scale effect” in the endogenous growth model do not find support. Jones (1995a,b) proposes a direct test of endogenous growth, whereby changes in policy should have permanent effects on the growth rate, and this hypothesis is decisively rejected on data for the United States and other advanced countries. These negative results led Jones to propose the “semi-endogenous” growth model that operates *without* the scale effect.¹⁸

Growth with Heterogeneous Firms

In international trade, we are particularly interested in the hypothesis that trade promotes productivity growth. It is natural then, to consider the heterogeneous firm model of Melitz (2003) that was discussed in chapter 6. The first papers incorporating growth into the heterogeneous firm models obtained ambiguous effects of trade on productivity growth: see Baldwin and Robert-Nicoud (2008) and Gustafsson and

¹⁸See Dinopoulos and Segerstrom (1999a) and Dinopoulos and Syropoulos (2001) for early applications of semi-endogenous growth to international trade.

Segerstrom (2010), as we shall explain below. But there are now papers showing how trade can enhance productivity growth, even in the long run, due to a positive selection effect on firms. This long-run result is now available from the work of Perla, Tonetti, and Waugh (2014) and Sampson (2016), and we shall present a simplified version of the model by Sampson.

In place of the CES production function shown in (11.16), it is convenient to go back to the CES utility function used in chapter 6. The output of a firm with productivity φ is denoted by $y(\varphi, t)$, where now we reserve the symbol x to denote *exports* in the trade equilibrium. As in the Melitz-Chaney model in chapter 6, we assume a Pareto distribution for firm productivities. We now introduce a lower bound of $\varphi \geq \varphi_0$ to the Pareto distribution, which is written as

$$G(\varphi | \varphi_0) = 1 - (\varphi | \varphi_0)^{-\theta}. \quad (11.41)$$

The essential assumption of the models by Perla, Tonetti, and Waugh (2014) and Sampson (2016) is that entering firms do not draw from the stationary distribution (11.41), but rather, draw from the distribution of productivities that improves over time. Sampson allows for a quite general improvement in the productivity distribution. While we rely on his model in all other respects, we follow Perla, Tonetti, and Waugh in using simply the *successful* firms at any point in time as the relevant distribution from which new firms draw their productivity.

Autarky Equilibrium

In the autarky equilibrium, let us denote the productivity of the zero-cutoff-profit (ZCP) entering firms by φ_{at} . When drawing from only the successful firms with $\varphi \geq \varphi_{at}$, then the conditional distribution of firms productivities is

$$G(\varphi | \varphi_{at}) = 1 - (\varphi | \varphi_{at})^{-\theta}. \quad (11.42)$$

We normalize the wage at unity, as in chapter 6. Then with price $p(\varphi) = \sigma / [\varphi(\sigma - 1)]$ and firm output of $y(\varphi, t)$, the instantaneous profit of a firm is $[1/\varphi(\sigma - 1)] y(\varphi, t) - f_d$.

As in chapter 6, the ZCP condition of a firm with productivity φ_{at} is

$$\left(\frac{1}{\sigma - 1} \right) \frac{y(\varphi_{at}, t)}{\varphi_{at}} = f_d. \quad (11.43)$$

Consider a firm with a productivity draw $\varphi > \varphi_{at}$ at time t . It will earn positive instantaneous profits at that instant, but as time passes the newly entering firms will become more productive (provided that φ_{at} rises over time) and its profits will fall. Denote by $T_a(\varphi, t)$ the time at which the profit of a firm entering at time t with productivity φ will be zero, so that

$$\left(\frac{1}{\sigma - 1} \right) \frac{y(\varphi, T_a(\varphi, t))}{\varphi} = f_d. \quad (11.44)$$

It is immediate by comparing (11.43) and (11.44) that $T_a(\varphi_{at}, t) = t$, that is, the firm receiving the ZCP productivity as time t will immediately exit again. In comparison, a firm entering with higher productivity $\varphi > \varphi_{at}$ at time t will remain in the market for some time $T_a(\varphi, t) > 0$.

As in our discussion of the Grossman-Helpman growth model (see equation 11.23), the present discounted value of profits for an entering firm should equal the fixed (i.e., sunk) costs of entry. With heterogeneous firms, we compute the present discounted value by integrating over all possible draws of productivity from the Pareto distribution in (11.42), and also integrating over the time period $T_a(\varphi, t)$ that a firm of each productivity will be in the market

$$V(t) \equiv \int_{\varphi_{at}}^{\infty} \int_t^{T_a(\varphi, t)} e^{-\rho(s-t)} \left[\left(\frac{1}{\sigma-1} \right) \frac{y(\varphi, s)}{\varphi} - f_d \right] ds dG(\varphi | \varphi_{at}) = f_e, \text{ for all } t. \quad (11.45)$$

For simplicity, we have used the same constant discount rate of ρ that we used in equation (11.23). Analogous to the earlier equation, (11.45) must hold for all t , so we can differentiate it with respect to t , obtaining

$$\begin{aligned} \dot{V} = & - \int_{\varphi_{at}}^{\infty} \left[\left(\frac{1}{\sigma-1} \right) \frac{y(\varphi, t)}{\varphi} - f_d \right] dG(\varphi | \varphi_{at}) + \rho V(t) \\ & + \int_{\varphi_{at}}^{\infty} e^{-\rho[T_a(\varphi, t) - t]} \left[\left(\frac{1}{\sigma-1} \right) \frac{y(\varphi, T_a(\varphi, t))}{\varphi} - f_d \right] \frac{\partial T_a(\varphi, t)}{\partial t} dG(\varphi | \varphi_{at}) \\ & - \int_t^{T_a(\varphi_{at}, t)} e^{-\rho(s-t)} \left[\left(\frac{1}{\sigma-1} \right) \frac{y(\varphi_{at}, s)}{\varphi_{at}} - f_d \right] \theta \frac{\dot{\varphi}_{at}}{\varphi_{at}} ds + \theta V(t) \frac{\dot{\varphi}_{at}}{\varphi_{at}}. \end{aligned} \quad (11.46)$$

The first line of (11.46) is obtained by differentiating (11.45) with respect to t as it appears in the lower-limit of integration, and then as it appears within the discount factor $e^{-\rho(s-t)}$. The second line of (11.46) is obtained by differentiating the upper limit of integration $T_a(\varphi, t)$ with respect to t . The final line is obtained by differentiating φ_{at} as the lower limit of integration with respect to t , and using $dG(\varphi | \varphi_{at}) = \theta(\varphi^{-\theta-1}/\varphi_a^{-\theta})d\varphi$ so that $dG(\varphi_{at} | \varphi_{at}) = \theta\varphi_{at}^{-1}d\varphi$; and then by differentiating φ_{at} as it appears within the Pareto distribution (11.42) with respect to t .

To proceed further, we need to solve for the three integrals in the above expression. Fortunately, the second integral is zero because of (11.44), and the third integral is zero because $T_a(\varphi_{at}, t) = t$, and so the upper and lower limits of integration coincide. To simplify the first integral, note that because we have assumed CES preferences and prices are $p(\varphi, t) = \sigma/[\varphi(\sigma-1)]$, then the quantity demanded for two products with productivity φ and φ_{at} are related by $y(\varphi, t)/y(\varphi_{at}, t) = (\varphi/\varphi_{at})^\sigma$. It follows that the first integral appearing in (11.46) is

$$\begin{aligned} \int_{\varphi_{at}}^{\infty} \left(\frac{1}{\sigma-1} \right) \frac{y(\varphi, t)}{\varphi} dG(\varphi | \varphi_{at}) &= \int_{\varphi_{at}}^{\infty} \left(\frac{1}{\sigma-1} \right) \frac{y(\varphi_{at}, t)}{\varphi_{at}^{\sigma-\theta}} \varphi^{\sigma-\theta-2} d\varphi \\ &= \left(\frac{1}{\sigma-1} \right) \frac{y(\varphi_{at}, t)}{\varphi_{at}^{\sigma-\theta}} \frac{1}{(\sigma-\theta-1)} \varphi^{\sigma-\theta-1} \Big|_{\varphi_{at}}^{\infty} = \left(\frac{\theta}{\sigma-1} \right) \frac{y(\varphi_{at}, t)}{\varphi_{at}} \frac{1}{(\sigma-\theta-1)}, \end{aligned} \quad (11.47a)$$

where in the final step we use the assumption that $\theta > \sigma - 1$. From (11.43) it follows that

$$\int_{\varphi_{at}}^{\infty} \left[\left(\frac{1}{\sigma-1} \right) \frac{y(\varphi, t)}{\varphi} - f_d \right] dG(\varphi | \varphi_{at}) = \left[\frac{\theta}{(\theta-\sigma+1)} - 1 \right] f_d = \frac{(\sigma-1)f_d}{(\theta-\sigma+1)}. \quad (11.47b)$$

Then setting expression (11.46) equal to 0, dividing by the fixed costs $V = f_e$, and using the above expression, we obtain

$$\rho + \theta \frac{\dot{\varphi}_{at}}{\varphi_{at}} = \frac{(\sigma-1)f_d}{(\theta-\sigma+1)f_e} \Rightarrow \frac{\dot{\varphi}_{at}}{\varphi_{at}} = \frac{1}{\theta} \left[\frac{(\sigma-1)f_d}{(\theta-\sigma+1)f_e} - \rho \right]. \quad (11.48)$$

Thus, provided that the above expression is positive, we have solved for a constant growth of the ZCP productivity in the autarky equilibrium. We can further show that this is the *only* source of productivity growth in equilibrium, and that in contrast to the Grossman-Helpman endogenous growth model, the mass of surviving firms does not change over time. To establish this result, we note that the labor market clearing condition can be represented as

$$L = N_t f_d + N_t \int_{\varphi_{at}}^{\infty} \frac{y(\varphi, t)}{\varphi} dG(\varphi | \varphi_{at}) d\varphi + \dot{N}_t f_e,$$

where N_t is the mass of surviving firms, and the first term on the right-hand side is the labor used for fixed costs; the second term represents the labor used for variable costs of production; and the third term is the labor used for sunk costs of entry. Using the same steps as in (11.47a,b), we can show that the integral in the second term is constant. It follows that the labor-market clearing condition has a steady-state solution with $\dot{N}_t = 0$ and N_t proportional to L .

Trade Equilibrium

We now consider the trade equilibrium and assume two identical countries, while continuing to normalize wages at unity. The set of successful firms is identical in both countries, namely, those firms with $\varphi \geq \varphi_{dt}$ where φ_{dt} is the ZCP productivity under international trade, using the notation from chapter 6. Then the conditional distribution is

$$G(\varphi | \varphi_{dt}) = 1 - (\varphi / \varphi_{dt})^{-\theta}.$$

Domestic prices are still $p(\varphi) = \sigma / [\varphi(\sigma-1)]$, and domestic output equals $y(\varphi, t) = Lc(\varphi, t) = Lp(\varphi)^{-\sigma} / P_{dt}^{1-\sigma}$. With iceberg transport costs of $\tau > 1$, export prices are $p_x(\varphi, t) = \tau \sigma / [\varphi(\sigma-1)]$, while $y_x(\varphi, t) = Lc_x(\varphi, t) = Lp_x(\varphi)^{-\sigma} / P_{dt}^{1-\sigma}$ denotes export quantity to the foreign market (with the same CES price index P_{dt}). Because of the symmetry across countries, we can see that domestic and export quantities are related by

$$y(\varphi, t) \tau^{-\sigma} = y_x(\varphi, t). \quad (11.49)$$

The ZCP productivity for a firm selling in the domestic market, φ_{dt} , satisfies

$$\left(\frac{1}{\sigma-1} \right) \frac{y(\varphi_{dt}, t)}{\varphi_{dt}} = f_d, \quad (11.50)$$

while ZCP productivity for a firm selling in the foreign market, φ_{xt} , satisfies

$$\left(\frac{\tau}{\sigma-1} \right) \frac{y_x(\varphi_{xt}, t)}{\varphi_{xt}} = f_x. \quad (11.51)$$

Consider a firm with a draw of productivity $\varphi > \varphi_{xt}$ at time t . It will earn positive instantaneous profits from exports at that instant, but as time passes the incoming firms will become more productive and this firm's profits will fall. Denote by $T_x(\varphi, t)$ the time at which its instantaneous profits from exports are zero, so

$$\left(\frac{\tau}{\sigma-1}\right) \frac{y_x(\varphi, T_x(\varphi, t))}{\varphi} = f_x. \quad (11.52)$$

It is immediate by comparing (11.51) and (11.52) that $T_x(\varphi_{xt}, t) = t$, that is, the firm receiving the ZCP productivity φ_{xt} at time t will immediately stop being an exporter. Over time, its domestic profits will also fall, and we denote by $T_d(\varphi, t)$ the time at which its instantaneous profits from domestic sales are zero, so that

$$\left(\frac{1}{\sigma-1}\right) \frac{y(\varphi, T_d(\varphi, t))}{\varphi} = f_d. \quad (11.53)$$

While this cutoff condition looks similar to (11.44), they differ because the CES price index used in the autarky and trade equilibrium differ, so that consumption and firm output also differ.

As noted earlier, outputs are related by $y(\varphi, t) = y(\varphi', t) (\varphi/\varphi')^\sigma$. Then, also using (11.49)–(11.51), we see that the share of exporters is

$$1 - G(\varphi_{xt} | \varphi_{dt}) = \left(\frac{\varphi_{dt}}{\varphi_{xt}}\right)^0 = \frac{1}{\tau^0} \left(\frac{f_d}{f_x}\right)^{0/(\sigma-1)}. \quad (11.54)$$

As in chapter 6, we assume that $f_d < f_x \tau^{\sigma-1}$ so that the above expression is less than unity. The fact that the ratio $\varphi_{dt}/\varphi_{xt}$ is fixed means that φ_{dt} and φ_{xt} will grow at the same rate.

Now we can write down the free-entry condition under international trade. The present discounted value of profits from domestic sales is similar to (11.45), namely,

$$V_d(t) \equiv \int_{\varphi_{dt}}^{\infty} \int_t^{T_d(\varphi, t)} e^{-\rho(s-t)} \left[\left(\frac{1}{\sigma-1}\right) \frac{y(\varphi, s)}{\varphi} - f_d \right] ds dG(\varphi | \varphi_{dt}).$$

Conditional on exporting, the present discounted value of exporting profits equals

$$V_x(t) \equiv \int_{\varphi_{xt}}^{\infty} \int_t^{T_x(\varphi, t)} e^{-\rho(s-t)} \left[\left(\frac{\tau^{1-\sigma}}{\sigma-1}\right) \frac{y(\varphi, s)}{\varphi} - f_x \right] ds dG(\varphi | \varphi_{xt}).$$

Notice that productivity distribution appears here as $dG(\varphi | \varphi_{xt})$, because we are assuming that the firm is profitable enough to export. To convert this to profits that are unconditional on exporting, we need to multiply by the probability of being an exporter. By construction, every firm that enters the market is going to sell at home, because they are drawing from a productivity distribution with lower bound φ_{dt} . Using this distribution, the probability of being a successful exporter is shown by (11.54). Then the overall present discounted value of profits equals

$$V(t) \equiv V_d(t) + \Pr_x(t) V_x(t) = f_e \Rightarrow \dot{V}(t) \equiv \dot{V}_d(t) + \Pr_x(t) \dot{V}_x(t) + \dot{\Pr}_x(t) V_x(t) = 0. \quad (11.55)$$

To determine the growth rate of the ZCP productivities with trade, we need to solve for the various terms appearing in (11.55). The first term is not different from autarky case as ZCP conditions still hold. So similarly to autarky case we obtain

$$\dot{V}_d(t) = -\frac{(\sigma-1)f_d}{\theta-\sigma+1} + \rho V_d(t) + \theta V_d(t) \frac{\dot{\varphi}_{dt}}{\varphi_{dt}}.$$

The expression for the second term in (11.55) is also very similar: replacing f_d with f_x , φ_{dt} with φ_{xt} , and also using the per-period profit $\tau^{1-\sigma} y(\varphi, t)/[(\sigma-1)\varphi]$ before deducting fixed costs, the functional form of V_d and V_x is similar, so that¹⁹

$$\dot{V}_x(t) = -\frac{(\sigma-1)f_x}{\theta-\sigma+1} + \rho V_x(t) + \theta V_x(t) \frac{\dot{\varphi}_{xt}}{\varphi_{xt}}.$$

Finally, notice that the probability of exporting does not change over time because (11.54) is fixed. Combining these various results, the free-entry condition (11.55) becomes

$$-\frac{(\sigma-1)f_d}{\theta-\sigma+1} + \rho V_d(t) + \theta V_d(t) \frac{\dot{\varphi}_{dt}}{\varphi_{dt}} + \left(\frac{\varphi_{xt}}{\varphi_{dt}} \right)^{-\theta} \left[-\frac{(\sigma-1)f_x}{\theta-\sigma+1} + \rho V_x(t) + \theta V_x(t) \frac{\dot{\varphi}_{xt}}{\varphi_{xt}} \right] = 0.$$

Using $V_d(t) + (\varphi_{xt}/\varphi_{dt})^{-\theta} V_x(t) = f_e$ and also (11.54), so that $\dot{\varphi}_{dt}/\varphi_{dt} = \dot{\varphi}_{xt}/\varphi_{xt}$, this condition is reexpressed as the following result, which is a simplified version of what is found in Sampson (2016).²⁰

THEOREM (SAMPSON 2016)

Under international trade, the growth rate of the ZCP level of productivity is

$$\frac{\dot{\varphi}_{dt}}{\varphi_{dt}} = \frac{\dot{\varphi}_{xt}}{\varphi_{xt}} = \frac{1}{\theta} \left[\frac{f_d(\sigma-1) \left(1 + \tau^{-\theta} (f_d/f_x)^{\frac{1-\sigma+1}{\sigma-1}} \right)}{(\theta-\sigma+1)f_e} - \rho \right].$$

This result shows that there is a continuous and constant growth rate in both the domestic and exporting ZCP productivities φ_{dt} and φ_{xt} . That growth will translate into continuous increases in GDP, average productivity, and in welfare.²¹ The above expression shows that the growth rate exceeds the autarky growth rate in (11.48), and is increasing as trade costs τ fall or as the fixed costs of entry or exporting fall. Surprisingly, however, the above growth rate is *increasing* in the fixed costs of production, f_d . That counterintuitive result arises in the model because as f_d rises, then the share of exporters in (11.54) rises. As a result, there are more firms that benefit from the access to

¹⁹There is no trade cost τ in this expression because it was eliminated using the ZCP condition for exporting.

²⁰Sampson (2016) describes a more general version of the model presented here, allowing for population growth, multiple countries, any elasticity of intertemporal substitution, and a wider class of distributions that new entrants face.

²¹See exercise 11.6.

the foreign market, so that opening to trade can be expected to have a greater impact on growth. This result explains why the first papers analyzing trade and growth with heterogeneous firms (Baldwin and Robert-Nicoud 2008; Gustafsson and Segerstrom 2010) found that opening to trade could increase or decrease the growth rate. As in the Grossman-Helpman model, they allowed for international spillovers of knowledge that would lower the fixed costs of production, and according to the above theorem, such a reduction will lower the growth rate of the ZCP productivity.

EMPIRICAL EVIDENCE

International Trade and Convergence

An implication of the above models is that trade should increase growth rates. There is an active debate over whether this hypothesis holds empirically. Advocates of this view include Dollar (1992), Sachs and Warner (1995), Edwards (1998), Ben-David (1993, 1998, 2001), and Frankel and Romer (1999). But these empirical results are all dismissed by Rodriguez and Rodrik (2000), and more specific criticisms on individual studies are made by Harrison (1996) and Slaughter (2001). In order to evaluate these arguments, it is useful to first relate them to another line of empirical research dealing with the *convergence* of countries to their steady-state growth rates.

The convergence literature is not motivated by the endogenous growth models at all, but rather, by the *exogenous* growth model such as Solow (1956) and extensions thereof. In those models, all countries converge to the same steady-state growth rate, but this pattern still allows for different levels of GDP per capita (depending on characteristics of the countries). So countries that are below their steady-state level of GDP per capita should grow faster, and conversely, countries that are above their steady-state GDP per capita should grow slower, so as to approach the steady state. This property follows directly from the diminishing marginal product of capital. There is both convergence in growth rates, and some degree of convergence in GDP per capita (though not necessarily to the same level across countries), and these two criteria tend to be used interchangably. The same is true for the papers dealing with trade, which examine either the impact of trade on growth rates or on GDP per capita across countries.

In the initial work of Barro (1991) and Barro and Salai-i-Martin (1991, 1992) the convergence hypothesis found strong empirical support. It was also demonstrated by Mankiw, Romer, and Weil (1992), and led them to conclude that the original Solow model, suitably extended to allow for human capital, is “good enough” at explaining cross-country differences in growth rates: the endogenous growth models are apparently not needed! These findings, however, have been questioned in subsequent works that consider a broader range of countries. For example, Easterly and Levine (2001) emphasize a *divergence* in the absolute levels of income per capita across countries: the rich have grown richer, and some of the poorest countries have become even poorer.²²

A similar approach to assessing the effects of trade is taken by the authors cited above. Sachs and Warner (1995) divide their countries into two groups, one of which has an “open” trade regime and the other of which is judged to be “closed.” Within the open group, they find evidence of convergence, with the poorer countries growing

²²Bernanke and Gürkaynak (2001) also question the conclusions of Mankiw, Romer, and Weil (1992).

faster, but this is not true for those countries with closed trade regimes. By this argument, openness leads to convergence of incomes across countries. Ben-David (1993, 1998, 2001) groups countries based on their membership in regional trade agreements, or by the strength of their bilateral trade ties. He also finds significant evidence of income convergence *within* these groups, but not necessarily across the groups. Dollar and Kraay (2002, 2004) contrast the experience of a group of developing countries that have become more open (including China and India) with other developing countries. They argue that growth in the former group is explained by openness and has particularly benefited the poor. As noted above, all these studies are subject to some empirical criticisms: among other difficulties, openness itself is fundamentally endogenous. Frankel and Romer (1999) deal with this endogeneity by first estimating a gravity equation, and then demonstrating that the *predicted* trade shares from this equation are indeed significant in explaining income per capita.

The finding that convergence occurs within groups, but not across groups, presents a challenge to the exogenous growth models. But this finding is in the spirit of the endogenous growth model without knowledge spillovers, discussed above, where the two countries converged to different growth rates. In a more general multicountry growth model that incorporates trade, Ventura (1997) also finds this result. In his model, capital *does not* have a diminishing marginal product due to Rybczynski effects: increases in capital raises the output of the capital-intensive good, and at fixed product prices, does not result in a reduced rental. Nevertheless, Ventura shows that in the global equilibrium, there is convergence to steady-state growth rates that differ across countries (as also demonstrated by Cuñat and Maffezzoli [2004]). The reason for convergence is that countries with expanding capital will have falling equilibrium prices for their capital-intensive goods, and therefore, a falling rental due to the Stolper-Samuelson theorem. So there is a general-equilibrium linkage between capital accumulation and the rental, even though there are no diminishing returns with fixed prices.

We might think of the model in Ventura (1997) as offering a third type of growth model, in which there is neither dimishing returns to capital as in the Solow model nor endogenous growth due to increased input variety, but growth *without* dimishing returns due to Rybczynski effects. Findlay (1996) argues that this framework may well be the most relevant one to analyze the growth experience of developing countries. These studies are recommended for further reading.

Spillovers of Knowledge

Another area of empirical investigation is international spillovers of knowledge. It is clear from our discussion of the Grossman-Helpman model that gains from trade are much more likely when knowledge flows across borders. Coe and Helpman (1995) test for the presence of such spillovers using data for the OECD countries.²³ Their hypothesis is that if international spillovers occur, the growth rates of countries should be correlated with both their *own* R&D expenditures and also the R&D expenditures of their trading partners. Thus, they construct measures of each country's R&D expenditures, and weight these using bilateral *import* shares to obtain an estimate of partner countries' R&D expenditures. They find that TFP growth rates at the country level are

²³Coe, Helpman, and Hoffmaister (1997) expand the sample to include developing countries.

indeed correlated with own- and partner-country R&D expenditures, lending support to the idea that research and development carried out in one country “spills over” to its trading partners.

An econometric difficulty with the Coe and Helpman estimates is that the TFP growth rates and R&D expenditures are very likely to exhibit unit-roots in their time-series behavior. In that case, regressions of TFP on R&D could lead to estimates that are significantly different from zero using conventional t-tests, even though the relationship might be spurious. To investigate this, Keller (1998) reconstructs the weighted R&D expenditures of the trading partners used by Coe and Helpman, but using *random* rather than actual import weights. He shows that this measure still has a positive correlation with TFP, suggesting that the relationship has little to do with trade, and may be spurious. The unit-roots problem can be corrected using panel co-integration techniques, as done by Funk (2001). He finds that there is no significant relationship between TFP growth and *import-weighted* R&D expenditures of partner countries. However, he does find a significant relationship between TFP and *export-weighted* R&D expenditures of partner countries, supporting the idea that exporters learn from their customers.

Keller (2002) has extended the results of Coe and Helpman using sectoral rather than aggregate country data. He again constructs a weighted average of R&D expenditures in other countries, but now uses the *distance* to each partner as the weight. Denoting the distance from country i to j by d^{ij} , Keller weights the other countries' R&D expenditures by $e^{-\delta d^{ij}}$. The parameter δ measures how quickly the impact of other countries' R&D expenditures on TFP diminishes with distance. In fact, this parameter turns out to be positive and highly significant, implying that spillovers are quite highly localized. For the two closest countries in his sample, Germany and the Netherlands, one dollar of R&D expenditure in Germany has 31 percent of the impact on Dutch productivity than has one dollar of R&D expenditure in the Netherlands. The effects of foreign R&D are lower for all other countries; for example, the relative effect of American R&D on Canada is only 4 percent of the value of Canadian R&D on its own productivity. So Keller's results provide evidence of precisely estimated but small R&D spillovers across borders.

Spillovers of knowledge have also been assessed through the use of patent data. Many studies have applied this technique to firms within a country, and Branstetter (2001) has extended it to consider firms in both the United States and Japan. When firms apply for a patent in the United States, they must classify it in one or more product areas. We can treat each product area as a component of the vector $B_i = (b_{i1}, \dots, b_{iK})$, where b_{ik} is the number of patents taken out by firm i in product area k . For two firms i and j , we can use the correlation between B_i and B_j as a measure of their *proximity* in technology space. Then we can construct a weighted average of the R&D expenditures of other firms, using the proximity with firm i as the weights. This gives us a measure of the *potential* spillovers from other firms. Branstetter constructs these measures separately for spillovers from U.S. firms and spillovers from Japanese firms.²⁴ Together with each firm's own R&D expenditures, these are regressed on a measure of each firm's performance (either number of patents taken out annually, or TFP growth), to determine the importance of the various spillovers.

²⁴Both measures are constructed using the patents that Japanese and American firms take out in the United States.

In his results, Branstetter (2001, 75) finds that spillovers are primarily *intranational* in scope. Some regressions show that Japanese firms benefit from the weighted R&D activity of American companies, but this effect does not remain significant when the domestic spillovers are included in the regression. Conversely, there is no evidence that American firms benefit from the R&D activities of the Japanese. This raises the idea that knowledge spillovers may be *asymmetric*, and presumably flow from the most technologically advanced country outward. In other work, Branstetter (2000) has explored whether Japanese firms that have foreign affiliates in the United States act as a “conduit” for knowledge flows, that is, whether there is a greater spillover between these foreign affiliates and other American firms. It appears that this is indeed the case, so that foreign investment increases the flow of knowledge spillovers both to and from the investing Japanese firms.

CONCLUSIONS

This chapter began with a discussion of productivity growth for a firm or an industry. The measure of total factor productivity (TFP) due to Solow (1957) is the growth of output minus a weighted average of the growth of inputs. The weights equal the elasticity of output with respect to the inputs. Under perfect competition and constant returns to scale, these weights are easily measured by the revenue shares of the inputs. Without these assumptions, however, the measurement of productivity growth becomes more difficult. Using the techniques of Hall (1988), we showed how “true” productivity growth can be identified, along with estimating the price-cost margins and returns to scale in an industry. This technique has been extended by De Loecker (2009) to take into account the endogeneity of input choices, and applied to measure how international trade affects markups, by De Loecker (2011), De Loecker and Warzynski (2012), and De Loecker, Goldberg, Khandelwal, and Pavcnik (2014).

Dropping the assumptions of perfect competition and constant returns to scale is also the starting point of the literature on endogenous growth. While there are several types of endogenous growth models, the one we considered in this chapter relied on expanding varieties of intermediate inputs. An essential assumption was that the fixed cost of inventing a new input is *inversely proportional* to the number of inputs already created. In the case where free trade in goods and inputs also brings international spillovers of knowledge, then trade leads to an increase in the growth rates of both countries (Grossman and Helpman 1990, 1991). But if there are no knowledge spillovers across borders, then free trade in goods has the effect of slowing down product development in the smaller country, which may also bring losses due to trade (Feenstra 1996). Recent literature on semi-endogenous growth has introduced the selection of heterogeneous firms, as in the Melitz (2003) model, into a dynamic framework. In particular, Perla, Tonetti, and Waugh (2014) and Sampson (2016) show how trade will increase the growth rate of countries.

These results lead to several important empirical questions: (a) Does free trade increase growth rates? (b) Are knowledge spillovers global or local? (c) Is there any direct evidence on the link between product variety and productivity? Empirical work exists on all these questions, but there are still some unanswered questions. In particular, it seems that the differences in per capita GDP across countries, which are as large as ever, must be due to *technology differences* across countries (Easterly and Levine

2001). This brings us right back to chapter 2, where we found that the HO model can account for trade patterns only if it is extended to incorporate such technology differences. What is the source of these differences? The endogenous growth model studied in this chapter has economy-wide increasing returns to scale, whereby larger countries create more inputs and are thereby more productive. In contrast, the growth model without scale effects from Sampson (2014) introduces a link between openness and growth due to the selection of firms, even though growth does not depend on country size. It seems to me that we have not yet related these potential explanations for country productivity differences to the implied productivity differences arising from the HO model. Could it be that large and open countries are the most productive due to their size and proximity to neighbors, leading to input variety and/or positive selection of firms, and that this productivity advantage also accounts for their trade patterns in a virtuous cycle (as suggested by Matsuyama 2013)? Addressing this question is an important area for further research.

PROBLEMS

- 11.1 In the two-good, two-factor model with *constant* prices, suppose that good 1 experiences Hicks-neutral technological progress, that is, $y_1 = A_1 f_1(L_1, K_1)$, with A_1 rising. Also assume that good 1 is labor intensive. Write down the zero-profit and full-employment conditions for the economy, and use the Jones algebra to show the following:
- $\hat{w} > \hat{A}_1 > 0 > \hat{r}$.
 - $\hat{y}_1 > \hat{A}_1 > 0 > \hat{y}_2$. Note that this is tricky because the factor prices are changing from part (a).
 - If you have read the appendix to chapter 1, then use the Lerner diagram to demonstrate part (a) graphically. Using this result, can you think of a graph that will also demonstrate part (b)?
- 11.2 Let us adopt a translog functional form for the GDP function in (11.13), extended to include many outputs:

$$\begin{aligned}\ln G = \alpha_0 + \sum_{i=1}^N \alpha_i \ln(A_i p_i) + \sum_{k=1}^M \beta_k \ln V_k + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} \ln(A_i p_i) \ln(A_j p_j) \\ + \frac{1}{2} \sum_{k=1}^M \sum_{\ell=1}^M \delta_{k\ell} \ln V_k \ln V_\ell + \sum_{i=1}^N \sum_{k=1}^M \varphi_{ik} \ln(A_i p_i) \ln V_k.\end{aligned}$$

Differentiating this with respect $\ln p_i$, we obtain the output share equations

$$s_i = \alpha_i + \sum_{j=1}^N \gamma_{ij} \ln(A_j p_j) + \sum_{k=1}^M \varphi_{ik} \ln V_k, \quad i = 1, \dots, N$$

where $s_i = p_i y_i / G$ is the share of each output in GDP.

- Write the quantity of each output as $\ln y_i = \ln(s_i G / p_i)$. Differentiate this with respect to $\ln p_i$ and obtain an expression for the output elasticity $(\partial \ln y_i / \partial \ln p_i)$. What restriction on the translog parameters must hold for this elasticity to be positive?
- Write the quantity of each output as $\ln y_i = \ln(s_i G / p_i)$. Differentiate this with respect to $\ln A_i$ and obtain an expression for the “Rybczynski-like”

elasticity ($\partial \ln y_i / \partial \ln A_i$). If the restrictions in (a) are satisfied, what can we say about the magnitude of this elasticity?

- 11.3 Consider the problem of maximizing (11.18) subject to (11.19), where for convenience we set $t = 0$. Write this as the Lagrangian

$$\begin{aligned} & \int_0^\infty e^{-\rho\tau} [\ln E(\tau) - \ln P(\tau)] d\tau + \lambda \left\{ \int_0^\infty e^{-R(\tau)} w(\tau) L d\tau + B(t) - \int_0^\infty e^{-R(\tau)} E(\tau) d\tau \right\} \\ &= \int_0^\infty \{e^{-\rho\tau} [\ln E(\tau) - \ln P(\tau)] d\tau + e^{-R(\tau)} \lambda [w(\tau)L - E(\tau)]\} d\tau + \lambda B(t), \end{aligned}$$

where in the second line we bring the Lagrange multiplier inside the integral. The expression inside the integral must be maximized at every point in time. So differentiate this expression with respect to $E(\tau)$ and then τ to obtain the first-order condition (11.20).

- 11.4 Consider a central planner who chooses the time path of output to maximize utility. This problem can be written as

$$\max_{Y(\tau)} \int_t^\infty e^{-\rho(\tau-t)} \ln Y(\tau) d\tau,$$

subject to

$$Y = N^{\sigma/(\sigma-1)} x = N^{1/(\sigma-1)} X \quad \text{and} \quad L = X + \alpha(\dot{N}/N),$$

where $X \equiv Nx$ and the constraints are the production function (11.17) along with the full-employment condition (11.26). Let us restrict our attention to steady-state solutions where $N(\tau) = N(t) e^{g(\tau-t)}$. Substituting this equation along with the constraints into the objective function, differentiate it with respect to g to compute the socially optimal growth rate g^{**} of new products. Show that $g^{**} > g^a$, and interpret this result.

- 11.5 Let us solve for the limiting values of the growth rates in (11.35):

$$g = \left(\frac{L}{\alpha\sigma} \right) - \left(\frac{\sigma-1}{\sigma} \right) \left[\dot{R} - \left(\frac{\dot{w}}{w} \right) \right], \quad \text{and}, \quad g^* = \left(\frac{L^*}{\alpha\sigma} \right) - \left(\frac{\sigma-1}{\sigma} \right) \left[\dot{R} - \left(\frac{\dot{w}^*}{w^*} \right) \right].$$

To do so, we use the share of world expenditure devoted to the products of the home country, which is $s = p_y y / (E + E^*)$, and the share devoted to products of the foreign country, which is $s^* = p_y^* y^* / (E + E^*)$, with $s + s^* = 1$. In the absence of trade in intermediate inputs, each final good is assembled entirely from inputs produced in the same country. The price of home inputs is $p_i = w\sigma/(\sigma-1)$, so it follows that $p_y y = p_i X = w\sigma X / (\sigma-1) = w\sigma(L - \alpha g) / (\sigma-1)$, where in the last equality we make use of the home full-employment condition, $L = X + \alpha g$. Therefore, the expenditure share on home products can be written as $s = w\sigma(L - \alpha g) / [(E + E^*)(\sigma-1)]$. Taking logs and differentiating, we obtain

$$\frac{\dot{s}}{s} = \frac{\dot{w}}{w} - \frac{\dot{g}}{(L - \alpha g)} - \left[\frac{\dot{E}}{E} \left(\frac{E}{E+E^*} \right) + \frac{\dot{E}^*}{E^*} \left(\frac{E^*}{E+E^*} \right) \right].$$

Using (11.20), the expression in brackets on the right-hand side equals $\dot{R} - \rho$. It follows that the home and foreign real interest rates are

$$\dot{R} - \frac{\dot{w}}{w} = \rho - \frac{\dot{s}}{s} - \frac{\dot{g}}{(L - \alpha g)} \quad \text{and} \quad \dot{R} - \frac{\dot{w}^*}{w^*} = \rho - \frac{\dot{s}^*}{s^*} - \frac{\dot{g}^*}{(L^* - \alpha g^*)}.$$

- (a) The fact that the home country is developing more products than abroad means that $\lim_{t \rightarrow \infty} s = 1$. What does this imply about the limiting value of the home real interest rate? Therefore, what is the limiting value of the home growth rate?
- (b) Since $\lim_{t \rightarrow \infty} s = 1$, then $\lim_{t \rightarrow \infty} s^* = 0$. What can we say about the limiting values of the foreign real interest rate, and the foreign growth rate?
- 11.6 (a) Using the model of growth with heterogeneous firms, use the autarky growth and the CES price index in (6.17) to derive the growth rate of the real wage.
- (b) Thus, as the growth increases due to the opening of trade, what happens to the growth rate of the real wage?

Multinationals and Organization of the Firm

Despite the fact that this book is about international trade, we have so far not introduced any role for *traders*. There has been little scope for firms, let alone economic organization more generally, to have any significant influence on trade patterns between countries. This contradicts the empirical fact that a good deal of trade occurs *internally* within firms located across countries. When a firm operates in several countries it is a *multinational enterprise*, and the investment made in the foreign country is referred to as *foreign direct investment (FDI)*.¹ The first goal of this chapter is to introduce these features into our earlier trade models.

The classic treatment of a multinational is that it has some intangible asset (such as knowledge of a production process) that it can use to its advantage in a foreign market. It must decide whether to simply *export* there, or to *invest* in that market by building a plant and selling the product, or whether to engage in a *joint venture or other contractual arrangement* with a foreign firm to produce the good. We will initially focus on the first two decisions—whether to export to the foreign market or to build a plant and sell there. The third case—whether to engage in a joint venture with a foreign firm—will then be discussed.

Notice that this classic statement of the FDI problem can work equally well in reverse, whereby the acquisition of a foreign firm can *bring with it* some knowledge of value to the purchaser that could not be obtained by simply buying the products of that foreign firm. Blonigen (1997) argues that increased inflows of Japanese acquisition FDI into the United States after 1985 were motivated by the desire to acquire the knowledge assets of U.S. firms, combined with the low value of the dollar. Notice that by this argument, FDI flows depend on the level of exchange rates: when the dollar depreciated after 1985, it became cheaper for foreign companies to purchase the intangible assets of U.S. firms.² There was, in fact, a surge in FDI into the United States during that period, particularly into the *acquisition* of U.S. companies.

We see that the decision to engage in FDI by a multinational therefore involves three interrelated aspects—*ownership* of an asset, *location* to produce, and whether

¹ We might define FDI as acquiring sufficient assets in a foreign firm to exercise some managerial control, though acquiring 10 percent or more of the assets of a foreign enterprise is the definition commonly used in practice.

² An alternative reason for an appreciation of foreign currencies to lead to increase FDI in the United States is through the wealth effect on foreign portfolios, as discussed by Froot and Stein (1991).

to keep the asset *internal* to the firm—and these constitute the so-called OLI framework of multinational activity (Dunning 1977, 1981). In the following sections, we explore models giving insight into these various aspects. We begin with the simplest example whereby a firm considers shutting down its factory in one country and moving it to another, which can be modeled by a movement of physical capital across countries. While such a movement of capital does not necessarily occur with FDI (which may simply involve a change in ownership without any movement of physical capital), it is a useful place to begin our analysis.

After briefly reviewing capital inflows into a one-sector economy, we use the classic article by Mundell (1957) to investigate capital mobility in the two-sector Heckscher-Ohlin (HO) model. In that case we find that even small tariffs have a very large impact on the movement of capital from one country to the other. This phenomenon of “tariff-jumping” FDI is a typical occurrence in import-substitution regimes, but is not restricted to the developing countries: many industrialized countries have used tariffs or quotas to induce the entry of foreign firms. This policy has some economic rationale, since the entry of foreign firms in simple models leads to a welfare gain due to increased domestic wages. The link between foreign investment and wages is therefore an important topic for empirical research. In Mundell’s analysis, domestic wages are not affected by the foreign investment (due to “factor price insensitivity” in the HO model), but it is still the case that the deadweight loss of the tariff is reduced due to the entry of foreign firms.

Entry of multinationals need not lead to capital flows between countries, however, and may instead just involve the ownership and location of firms. We consider several models of that type, distinguishing *vertical* from *horizontal* multinationals. An example of the former case is a firm that has headquarters in one country but does production in another to obtain lower factor prices there (Helpman 1984c). The latter case occurs when a firm decides to duplicate production facilities and sell locally in two or more countries, due to tariffs or other barriers between them (Markusen 1984, 2002). We identify conditions under which a particular type of multinational is most likely to arise in equilibrium and, once again, find that trade policies can have a significant impact on multinational location.

Beyond trade that is internal to multinational firms, there is another large portion of trade that occurs *between* firms that have some special relationship. An example is when a firm offshores some of its production process to a lower-wage country. In chapter 4 we modeled this as the purchase of intermediate inputs, but in fact it often involves a longer-term relationship: the firm wanting to purchase the input is committing its money, and the selling firm is committing resources to production and development of the inputs. This relationship entails risk to both parties, and in many circumstances these risks cannot be offset by legal contracts. If the costs associated with these incomplete contracts are too great, then the firm wanting to purchase the inputs will instead vertically integrate into this activity, becoming a vertical multinational. Thus, a consideration of transaction costs under incomplete contracts allows us to model the *internalization* decision of a multinational firm. The work of Antràs (2003, 2005, 2015), Antràs and Helpman (2004), Antràs and Rossi-Hansberg (2009), Grossman and Helpman (2002, 2004, 2005), and the volume by Helpman, Marin, and Verdier (2008) puts this in a general equilibrium setting and investigates the implications of internalization for trade.

Transactions costs is the first and foremost reason for firms to integrate, but there are other reasons, too. For example, joint ventures between firms in different

countries can bring with them knowledge of the supplier and distribution networks in each country, and any other intangible assets owned by either firm. Rauch and Trindade (2003) model this informational issue as a problem of “matching” firms across countries, where there is some inherent uncertainty in the ability to find a good partner. As we shall describe, the efficacy of achieving matches determines the extent to which markets are integrated, that is, the extent to which factor price equalization obtains.

CAPITAL FLOWS IN A ONE-SECTOR ECONOMY

We begin our analysis with the simplest case of a capital inflow into an economy producing just one good, as in MacDougall (1960). Suppose that the production function is given by $y = f(L, K)$, which is linearly homogeneous and concave. Normalizing the product price at unity, the rental on capital is $r = f_K(L, K)$, with $f_{KK} < 0$. Denote the domestic capital stock by K_0 and the foreign inflow of capital by K^* . The equilibrium rental in the absence of the foreign capital is shown by r_0 in figure 12.1, and in the presence of the capital inflow by r_1 . The foreign capital earns the amount $r_1 K^*$, which is taken out of the country. The increase in GDP due to the capital inflow is

$$\Delta y = \int_{K_0}^{K_0 + K^*} f_K dK = A + B. \quad (12.1)$$

Subtracting the payment to foreign capital, $r_1 K^* = B$, the net welfare gain to the home country is the area A . Thus, the inflow of foreign capital creates a welfare gain.

What is the source of this gain? It is evident from figure 12.1 that as capital enters, it depresses the rental, so that the payments to foreign capital (B) are less than the total area under the demand curve ($A + B$). The other side of this is that the capital inflow raises the marginal product of labor and wages, $w = f_L(L, K)$. Indeed, the total increase

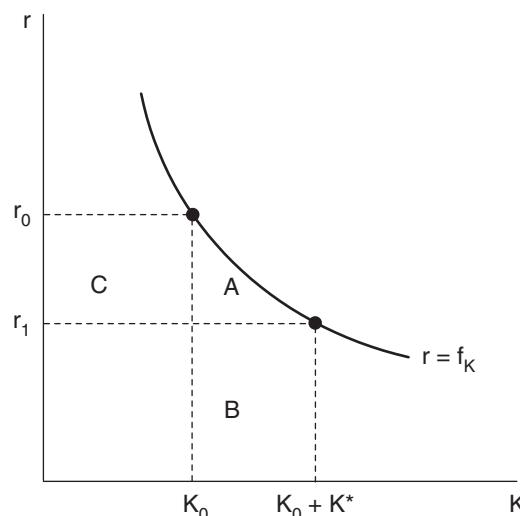


Figure 12.1

in the wage bill equals the area $A + C$ in figure 12.1.³ Area C is the redistribution of income from capital to labor, whereas area A is the net welfare gain and it accrues to labor.

These simple observations have led to a good deal of empirical work as to whether industries experiencing an inflow of foreign capital pay higher wages. For the United States, Howenstein and Zeile (1994) and Doms and Jensen (1998) find that foreign plants pay higher wages than do domestic plants. Blonigen and Tomlin (2001) further find that foreign plants typically grow faster than U.S. plants of similar size, and that they lead to a larger positive impact on local wages than does domestic investment. Combining U.S. data with information on Mexican and Venezuelan enterprises, Aitken, Harrison, and Lipsey (1996) also find a wage differential between domestic and foreign-owned enterprises, even after controlling for industry size and capital intensity.

Notice that these studies are most often comparing the wages paid by foreign and domestic plants, which is not exactly the same as the *general equilibrium* wage increase due to foreign investment predicted by figure 12.1. For example, if foreign firms simply attract the highest-quality workers, wages would increase in comparison to domestic plants, but not necessarily welfare. This same problem of self-selection occurs in empirical studies of *technology transfer* between foreign and local firms (through joint ventures or other spillovers across firms). Evidence in support of technology transfer for Mexico is found by Blomstrom (1986), Blomstrom and Persson (1983), and Kokko (1994), and for Indonesia by Blomstrom and Sjoholm (1999). These studies correlate *industry* productivity with the presence of foreign firms. However, this leaves open the question of whether the higher productivity is due to the foreign plants themselves, or due to spillovers. This has been addressed using plant-level data by Haddad and Harrison (1993) for Morocco and by Aitken and Harrison (1999) for Venezuela. They find that the foreign plants indeed have higher productivity, but that there is a weak negative correlation between their presence and productivity of the *domestic* plants in the same industry. Thus, these findings cast some doubt on the hypothesis of positive technological spillovers between foreign and domestic firms; rather, foreign firms are just more productive.

CAPITAL FLOWS IN THE TWO-SECTOR HECKSCHER-OHLIN MODEL

We turn next to a famous example of capital flows in the two-sector Heckscher-Ohlin (HO) model, due to Mundell (1957). Suppose there are two countries with identical technologies and homothetic tastes, initially engaged in free trade. With countries producing both goods and no factor intensity reversals, the rental on capital is uniquely determined by the goods prices, which we write as $r(p_1, p_2)$ for both countries. In figure 12.2 the endowment of the two countries is shown at the point V^i , which we assume is inside the factor price equalization set so the rental on capital is the same across countries.⁴ With free trade in goods, consumption of the two countries is proportional and occurs along the diagonal of figure 12.2, such as the point AD^i (where D^i is the consumption vector of each country, and multiplying by the technology matrix A converts this to the factors embodied in consumption). In the case we

³ Because the production function is linearly homogeneous, then $y_0 = w_0 L + r_0 K_0$, and $y_1 = w_1 L + r_1 (K_0 + K^*)$. It follows that $\Delta y = (A + B) = (w_1 - w_0)L + (r_1 - r_0)K_0 + r_1 K^*$, and since $B = r_1 K^*$ and $C = (r_0 - r_1)K_0$, the wage bill changes by $(w_1 - w_0)L = (A + B) - (r_1 - r_0)K_0 - r_1 K^* = A + C$.

⁴ Refer to figure 2.3 to review how this figure is constructed.

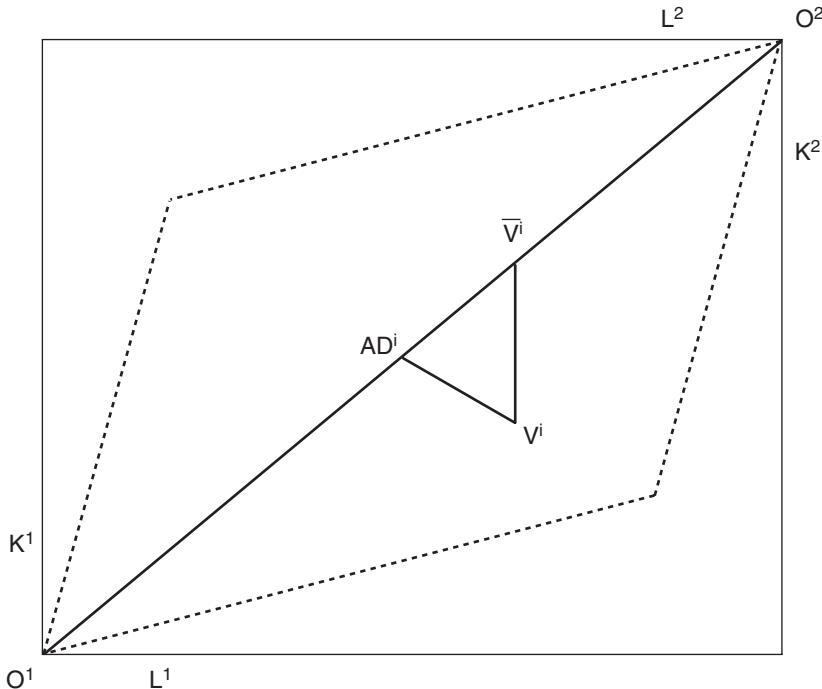


Figure 12.2

have illustrated, country 1 is labor abundant, and the move from V^i to AD^i involves country 1 importing the capital-intensive good. Assume that good 1 is capital intensive, so that $\partial r(p_1, p_2)/\partial p_1 > 0$.

Now allow for capital mobility between the countries, and suppose that country 1 applies a specific import tariff of $t > 0$. This raises the domestic price of good 1 in that country, and since this good is capital intensive, by the Stolper-Samuelson theorem it increases the return to capital, $r(p_1 + t, p_2) > r(p_1, p_2)$. It follows that capital will flow from country 2 to country 1. In figure 12.2, this shifts the effective endowment point vertically upward from point V^i . The question is, where will this process end? The new equilibrium allowing for capital mobility requires that the return to capital be *equalized* across the countries. But with the tariff driving a wedge between the prices in the two countries, it is *impossible* for rentals to be equalized: $t > 0$ always implies that $r(p_1 + t, p_2) > r(p_1, p_2)$. Obviously, something in this equation must change to achieve a new long-run equilibrium!

To see what changes, let us think about what happens as the endowments of the countries move upward from point V^i . Capital is flowing towards country 1, which increases the output of industry 1 and reduces that of industry 2. Since good 1 is imported, the increased supply of that good *lowers* imports, and the reduced supply of good 2 also lowers exports. So as we move upward from point V^i , trade between the countries is being reduced. This will continue until the endowments reach the diagonal, indicated by the point \bar{V}^i , at which point the capital/labor ratios in the two countries are identical. This means that there is no reason for trade: both economies are in autarky.

What about the rental on capital in each country? Since both countries are in autarky, with identical capital/labor ratios and identical homothetic tastes, this means that the equilibrium prices are the same across countries. It follows therefore, that the equilibrium rental on capital is *also* the same across countries. The tariff that is being applied by country 1 is redundant: since there are no imports, the tariff *no longer* drives a wedge between the prices and rentals. So we conclude that the point \bar{V}^i is the new equilibrium with capital mobility. Product prices and the rental on capital are unchanged from their original values *before* the tariff (since this is a property of a change in endowments within the FPE cone). Furthermore, the entire increase in the GDP of country 1 from point V^i to point \bar{V}^i is repaid to the foreign capital located there.⁵ If we deduct this amount from GDP, country 1 ends up back at its *original* consumption point AD^i , and so its welfare is the same as before the tariff was applied!

This analysis by Mundell (1957) is important for several reasons: it shows that a tariff can lead to a large enough inflow of capital to eliminate trade; and it shows that the capital flow eliminates the deadweight loss due to the tariff. Both of these observations are realistic, and can be illustrated by the VER applied to Japanese exports of automobiles into the United States, as discussed in chapter 9. While this trade policy took the form of a quota rather than a tariff, it still led a large inflow of Japanese auto firms into the United States—enough to make the VER redundant after 1987 (i.e., there were fewer imports from Japan than allowed by the VER). This means that its welfare cost was also reduced.⁶

The results of Mundell (1957) are sometimes summed up with the phrase “trade in factors is a substitute for trade in goods,” since the capital inflow due to the tariff has the effect of eliminating trade. This hypothesis has been empirically tested. Blonigen (2001) investigates the extent to which FDI and imports are substitutes (with inflows of capital reducing imports) or complements (FDI leading to higher imports). He finds evidence for both hypotheses, as is also confirmed by Swenson (2004) for the United States and Goldberg and Klein (2000) using data from Latin America. The effect of U.S. antidumping duties on inward FDI is investigated by Blonigen (2002). He finds that this U.S. investment is a realistic option only for firms with prior multinational experience. While the raw data show a particularly high response rate for Japanese firms shifting their production to the United States, this is mainly explained by the fact that many of these firms are already multinationals.

VERTICAL AND HORIZONTAL MULTINATIONALS

We turn now from a discussion of capital flows to the modeling of multinational activity, where a firm chooses to operate plants in multiple countries. We distinguish the cases where the multinational chooses to operate its headquarters in one country and its production facilities in another—the so-called *vertical* multinational—from those where it operates its headquarters in one country but has production facilities in multiple countries—the *horizontal* multinational.

⁵ To see this, measure GDP as the value of factor earnings. With unchanged product prices and factor prices, the new GDP at \bar{V}^i is $GDP = wL^i + rK^i + rK^*$, so deducting payment to foreign capital rK^* , we obtain the original GDP of $wL^i + rK^i$.

⁶ Neary (1988) examines more systematically the welfare costs of tariffs and quotas under capital mobility.

Vertical Multinationals

The vertical case was first analyzed by Helpman (1984c) in a two-factor framework with monopolistic competition. The incentive to operate headquarters in one country and production in another arises due to factor price differences across the countries. As we shall see, however, the ability of a multinational to spread its facilities across several countries serves to ameliorate these factor price differences, and *enlarges* the region of factor price equalization.

Suppose that sector 1 produces a homogeneous good, while sector 2 produces a differentiated good with a CES (constant elasticity of substitution) utility function. A variety of the differentiated good requires α units of high-skilled labor H in its fixed costs (which we think of as “headquarters” services), and β units of low-skilled labor L in its variable costs. Denoting the output of a differentiated variety by y_{2i} , the costs of production are $C_{2i}(w_H, w_L, y_i) = w_H \alpha + w_L \beta y_{2i}$, where w_H and w_L are the wages of high-skilled and low-skilled labor, respectively. As discussed in chapter 5, with a CES utility function there is a unique output $y_{2i} = \bar{y}$ at which profits are zero, where free entry also determines the number of varieties N_2 in sector 2. Then the demand for high-skilled labor for each variety produced is α , and the demand for low-skilled labor is $\beta\bar{y}$.

Technologies are the same across countries, and we will focus on the “integrated world equilibrium” at which factor price equalization (FPE) is obtained. In sector 1, producing the homogeneous good, the requirements of low-skilled and high-skilled labor to produce *one unit* of output are denoted by a_{1L} and a_{1H} . In sector 2, the requirements to produce *one variety* of the differentiated product are $a_{2L} \equiv \beta\bar{y}$ and $a_{2H} \equiv \alpha$. Then we can let $A = \begin{bmatrix} a_{1L} & a_{2L} \\ a_{1H} & a_{2H} \end{bmatrix}$ denote the matrix of labor requirements in each industry, which applies in either country under FPE. Let $Y^j = (y_1^j, N_2^j)'$ denote the vector whose first component is the output of industry 1 in country j , and whose second component is the *number* of varieties of the differentiated product in sector 2 produced in country j . Then full employment of resources in each country requires that $AY^j = V^j$, where V^j is the vector of endowments in country $j = 1, 2$.

Let us assume initially that the fixed costs of α and variable costs of $\beta\bar{y}$ for each differentiated good have to be performed in the *same* country. Then the integrated world equilibrium can be illustrated in a fashion very much analogous to the two-factor HO model. In figure 12.3, the FPE set is shown by the parallelogram $O^1QO^2Q'O^1$. Starting at any endowment point such as V^j within this parallelogram, free trade in goods allows the economies to consume at points D^j that are proportional across countries, so that the factors embodied in consumption AD^j lie along the diagonal of the box.⁷ The slope of the line between V^j and AD^j (in absolute value) is the equilibrium ratio of factor prices w_L/w_H .

Let us further assume that industry 1, producing the homogeneous good, is low-skilled-labor intensive, so that $a_{1L}/a_{1H} > a_{2L}/a_{2H} = \beta\bar{y}/\alpha$. Then from the endowment and consumption points we can readily compute the trade balance in each industry. Thus, in *production*, country 1 uses the amount of labor O^1B in industry 1 and O^1B' in industry 2, which sum to the endowment point V^j . Similarly, in *consumption*, country 1 uses the amount of labor O^1C in industry 1 and O^1C' in industry 2, which sum to the point AD^j . It follows that country 1 is an importer of good 1 ($OB < OC$), and a

⁷ Note that the consumption in each country equals $D^j = [d_1^j, (N_2^j + N_2^2)s^j]'$, where the first component is the quantity consumed in country j of the homogeneous good, and the second component is the *world* variety of the differentiated good times country j 's share of world GDP.

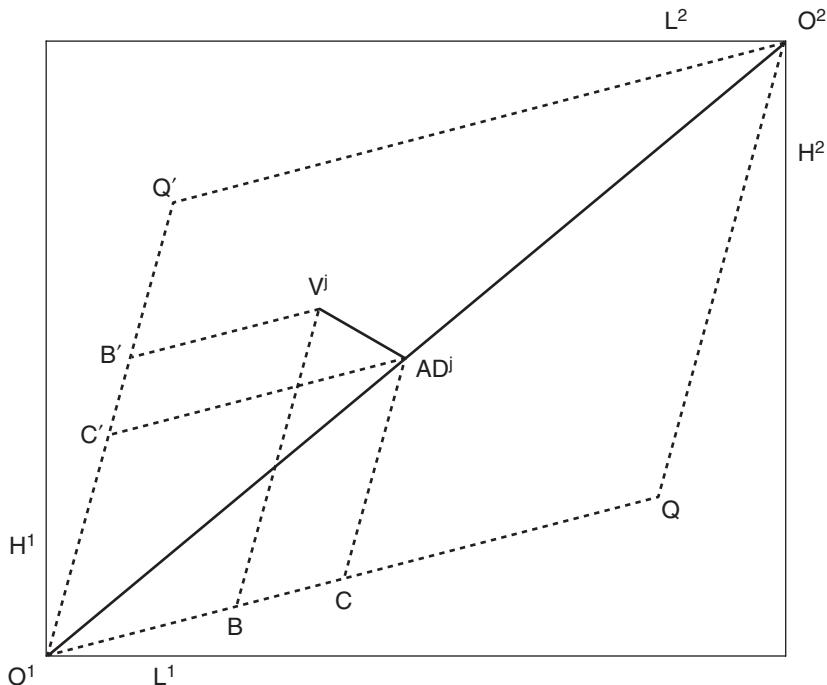


Figure 12.3

net exporter of the differentiated good ($OB' > OC'$). In addition, there is intraindustry trade in the differentiated goods sector.

Now let us change our assumptions and suppose that the fixed costs of a high-skilled labor and variable costs of βy low-skilled labor for each differentiated good can be performed in *different* countries by a multinational firm. Then how does the ability to “fragment” the headquarters and production activities across countries affect the FPE set? A quick (and entirely correct) answer can be obtained by appealing to the FPE set with *multiple goods* developed by Dixit and Norman (1980), as discussed in chapter 3. Rank the activities beginning with the variable costs of producing the differentiated good, which uses only low-skilled labor; followed by the production of the homogeneous product, which uses both factors; followed by the fixed costs of producing the differentiated good, which uses only high-skilled labor. This gives *three* rather than two factor requirements vectors, which are $(\beta y, 0)$, (a_{LH}, a_{IH}) , and $(0, \alpha)$. Multiplying each of these by the total worldwide amount of each activity in the integrated equilibrium, and summing them, we will obtain the FPE set $O^1RSO^2R'S'O^1$ in figure 12.4. This FPE region is clearly *larger* than the original parallelogram $O^1QO^2Q'O^1$.

To understand how FPE is achieved in this expanded set, suppose that the factor endowments of the two countries are given by V' in figure 12.4. This point lies *outside* the original FPE set, so that any firm in sector 1 would choose the country with minimum wage of high-skilled labor for headquarters, and the country with minimum wage of low-skilled labor for production. Since country 1 is high-skilled labor abundant, we can presume that it would have the lower relative wage of high-skilled labor, and firms from either country would want to establish headquarters there. This will increase the demand for high-skilled labor in country 1, to the point where the

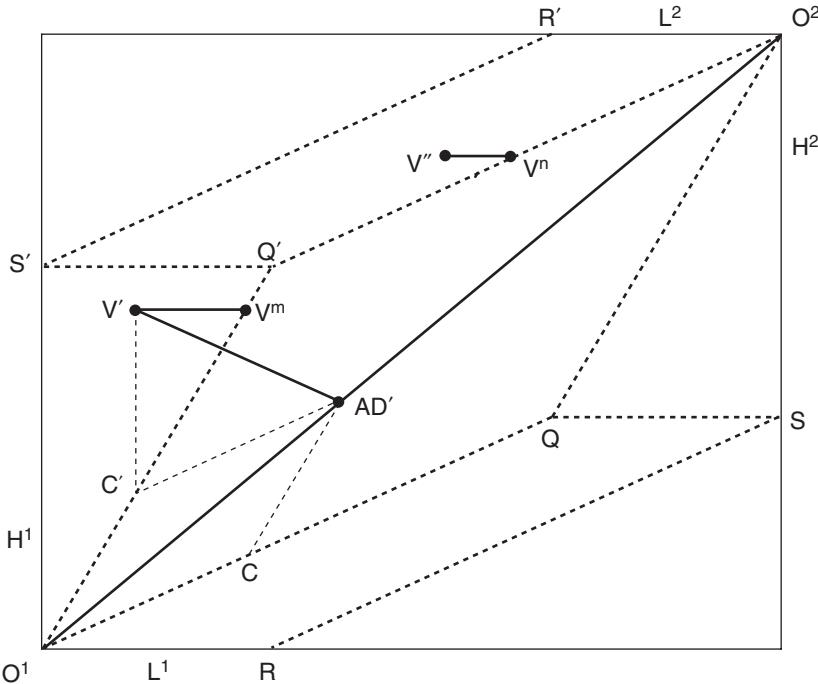


Figure 12.4

entire endowment V' can be employed at the *same* factor prices as prevailing abroad. Intuitively, this is what allows FPE to hold in the expanded set.

Because there are three production activities and only two factors, there are actually many combinations of the activities across countries that are consistent with full employment and FPE. As in Helpman (1984c), we will focus on those combinations that lead to a minimum amount of trade. With factor endowments given by V' in figure 12.4, country 1 could produce varieties of the differentiated good using resources shown by O^1C' , and then devote its remaining endowment of high-skilled labor, shown by $C'V'$, to headquarters services. These headquarters services would be combined with *low-skilled* labor in country 2, of the amount VV^m , in order to produce varieties of the differentiated product in that country. Additional resources in country 2, of the V^nQ' , would also be devoted to produce varieties of the differentiated product, and the remaining resource of $Q'O^2$ would be used to produce the homogeneous good in sector 1. Factor price equalization is achieved through combining high-skilled labor for headquarters services in country 1 with low-skilled labor for production in country 2.

What is the trade pattern implied by the multinational activity? For factor endowments in the region $O^1S'Q'$, as we have been discussing, the homogeneous good is produced only in country 2, so country 1 imports that good. But for the differentiated good, the trade pattern is not so clear. Indeed, in the case we have illustrated in figure 12.4, it turns out that both countries have *balanced net trade* in the differentiated good. This can be seen by comparing the resources used in production and consumption. As discussed just above, country 1 uses the resources O^1C' to produce varieties of the differentiated good. To determine its consumption, we can trace a line from its

endowment point V' to its consumption of factors AD' (this line has the slope w_L/w_H). We then decompose the point AD' into the consumption of factors O^1C in sector 1, and O^1C' in sector 2. Notice that for the case we have illustrated, the *consumption* of factors O^1C' in sector 2 is identical to those used in the *production* of differentiated varieties by country 1. It follows that trade is balanced in that sector, and the same will hold for country 2: they have balanced intra-industry trade in the differentiated good.⁸

How can country 1 be an importer of the homogeneous good, and yet have balanced net trade in the differentiated good? The answer is that this country receives *profits* from its multinationals engaged in production abroad, which would show up as investment income in the current account. Thus, its balance-of-trade deficit (due to importing the homogeneous good) is offset by receipts from its subsidiaries abroad, so that it has an overall current account balance. The *amount* of profits earned in country 2 will be proportional to the low-skilled labor employed there by the multinationals, which is measured by $L^* \equiv VV''$. With $\beta\bar{y}$ units of low-skilled labor needed per variety, the number of varieties produced by multinationals in country 2 is therefore $N^* \equiv L^*/\beta\bar{y}$. The value of headquarters services provided by these multinationals through employing high-skilled labor in country 1 is then $\alpha N^* = \alpha L^*/\beta\bar{y}$.⁹ This is an “invisible export” by country 1, which is used to pay for its imports of the homogeneous good.

As we move to other regions of the FPE set, the trade patterns change. For example, consider endowment within $O^2R'S'Q'$, such as V'' . Once again, multinationals with headquarters in country 1 will use some of the unskilled labor in country 2 so that FPE is achieved. In this case, the amount of unskilled labor employed in country 2 by multinationals is the distance $V''V''$. We can think of the vector O^1V'' as the *employment* of resources by firms that have headquarters in country 1, as distinct from that country’s *endowment* vector O^1V'' . In the absence of multinationals, the employment of resources must equal a country’s endowment, but now they can differ. We can also construct the trade patterns in the region $O^2R'S'Q'$. It turns out the country 1 can be either an importer or an exporter of the homogeneous product in this region, and also export the differentiated product as well as the headquarters services.¹⁰

Setting aside the details of the trade patterns, the most important message to take from figure 12.4 is that vertical multinationals *enlarge* the FPE set. This is a deep insight from Helpman (1984c), and can be compared to our results in chapter 4, where we discussed trade in intermediate inputs and wages. We suggested there that the ability to fragment the production process, and purchase some low-skilled-labor intermediate inputs from another country, could lead to an *increase* in the relative demand and wages of high-skilled labor in *both* countries. In comparison, we have argued above that such offshoring by multinationals will lead to an enlargement of the FPE set, and therefore, a tendency for factor prices to equalize across countries. These two results are not necessarily inconsistent, but it is worth thinking about how the assumptions of Helpman (1984c) differ from those in chapter 4.

In the model of chapter 4 with a continuum of inputs, we showed that an increase in offshoring would occur through the “borderline” activity being transferred from

⁸ It follows that for slightly different choices of the endowment point V' , country 1 could be either a net exporter or a net importer of the differentiated good in the region $O^1S'Q'$.

⁹ Notice that the skilled labor of $\alpha N^* = \alpha L^*/\beta\bar{y}$ in country 1 is equivalently measured by the distance $C'V'$.

¹⁰ Helpman (1984c) carefully solves for the trade pattern in this region, as well as showing how endowments affect the *volume* of trade in all regions of the FPE set.

the home to the foreign country: that activity was *less* skill intensive than any other activity performed at home, but *more* skill intensive than any other activity performed abroad. Therefore, the relative demand for high-skilled labor rose in both countries, as did its relative wage. In contrast, in our discussion of Helpman's (1984c) model above, we treated the production of the differentiated good as using only low-skilled labor, with headquarters services using only high-skilled labor. The production activity is therefore *less* skill intensive than the homogeneous good produced in either country. The transfer of multinational production from country 1 to country 2 raises the relative demand for high-skilled labor in country 1, but lowers it in country 2. This will raise the relative wages of high-skilled labor in country 1, but lower it in country 2, thereby moving these relative wages towards *equality*, so that FPE is achieved.

We see that the assumptions on factor-intensities differ between chapter 4 and what we have used to describe Helpman's (1984c) model.¹¹ Obviously, other assumptions on factor intensities can be incorporated into Helpman's model, as well as alternative assumptions on fixed costs and trade barriers, as we shall explore in the next section.

HORIZONTAL MULTINATIONALS

In Helpman's model, the multinational will never choose to operate production facilities in *both* countries, because with free trade, there is no benefit to doing so. But suppose that there are some trade barriers, such as the “iceberg” transport costs discussed in chapter 5. Suppose further that each plant has some specific fixed costs, which must be borne in the same country as production costs. Then we will say that *horizontal FDI* occurs if a firm chooses to operate production facilities in *multiple countries*, with each plant selling locally to that market. Horizontal FDI requires that the gains from selling locally and thereby avoiding transport costs more than offset the plant-specific fixed costs. This case has been analyzed in a series of studies beginning with Markusen (1984) and culminating in Markusen (2002). These writings cover a whole range of vertical and horizontal multinationals, using low-skilled and high-skilled labor in varying proportions. We will not attempt to summarize all these results, but shall develop one example of how horizontal multinationals can arise.

We suppose that a firm in country i is already producing for its home market and is deciding whether to export to country j or establish a production facility and sell there. This firm has already covered its fixed costs of headquarters services, as well as any plant-specific fixed costs in country i , and there is an additional fixed cost of establishing a plant abroad. We suppose there are “iceberg” transportation costs between the countries, so that $\tau^{ij} > 1$ units of the good must be exported from country i in order for one unit to arrive in j . We let p^i denote the free on board (f.o.b.) price received by the firm in country i . Inclusive of the transportation costs, the cost, insurance, freight (c.i.f.) price abroad is $p^{ij} = \tau^{ij} p^i$.

We borrow the demand structure from our treatment of monopolistic competition in chapter 5. With country j having a CES utility function over the differentiated product, its demand c^{ij} for a variety sent from country i is

¹¹ Helpman (1984c) actually uses more general assumptions on factor intensities but, to illustrate the FPE regions, focuses on the case where the fixed costs use only skilled labor (which he calls a “general purpose” input H), and the variable costs use only unskilled labor. These are the same assumptions we have used above. In related theoretical work, Helpman (1985) allows for multinationals to produce multiple products.

$$c^{ij} = (p^{ij}/P^j)^{-\sigma} (Y^j/P^j), \quad (12.2)$$

where Y^j is country j 's GDP, and P^j refers to country j 's overall price index, defined as

$$P^j = \left(\sum_{i=1}^C N^i (p^{ij})^{1-\sigma} \right)^{1/(1-\sigma)}, \quad (12.3)$$

and where N^i is the number of products exported by each country $i = 1, \dots, C$.

We treat labor as the only factor of production, and do not distinguish by skill type. The *exporting* firm in country i has the production function $L^i = \beta y^i$, where β is its marginal labor costs. Note that its plant-specific fixed costs are *already covered* by virtue of its home sales of this product. Revenue received from producing in country i and exporting to country j is $p^{ij}c^{ij} = p^i y^i$, where output is related to consumption by $y^i = c^{ij}\tau^{ij}$. The profits from exporting are then

$$p^i y^i - w^i \beta y^i = (p^i - w^i \beta) \tau^{ij} c^{ij}. \quad (12.4)$$

Substituting for the CES demand from (12.2), and maximizing (12.4) over the choice of the price $p^{ij} = \tau^{ij} p^i$, we readily obtain the optimal price $p^i [1 - (1/\sigma)] = \beta w^i$. Substituting this back into (12.4), we obtain the maximized value of profits,

$$p^i y^i - w^i \beta y^i = \left(\frac{p^i y^i}{\sigma} \right) = \left(\frac{p^i \tau^{ij}}{\sigma} \right) \left(\frac{p^i \tau^{ij}}{P^j} \right)^{-\sigma} \left(\frac{Y^j}{P^j} \right) = \frac{1}{\sigma} \left(\frac{p^i \tau^{ij}}{P^j} \right)^{1-\sigma} Y^j, \quad (12.5)$$

where the second equality follows using consumption from (12.2) with $y^i = c^{ij}\tau^{ij}$, and the third by simplification. From (12.5), we see that higher transport costs reduce profits from exporting, while higher GDP abroad increase profits.

Rather than exporting from country i , the multinational firm can instead establish a production facility in country j , and sell locally. In that case, it will face the production function $L^j = \alpha^j + \beta y^j$, including the plant-specific fixed costs paid abroad. For simplicity we treat marginal labor costs β as equal across countries, though the wages and fixed costs can differ. There are no transport costs when producing in country j and selling locally, so with the price p^j , the quantity sold is

$$c^j = (p^j/P^j)^{-\sigma} (Y^j/P^j). \quad (12.2')$$

Again, the optimal price is determined as $p^j [1 - (1/\sigma)] = \beta w^j$. Using this and (12.2') to compute profits earned from the subsidiary in country j , we obtain

$$p^j y^j - w^j (\beta y^j + \alpha^j) = \frac{1}{\sigma} \left(\frac{p^j}{P^j} \right)^{1-\sigma} Y^j - w^j \alpha^j. \quad (12.5')$$

Thus, in order for the multinational based in country i to wish to establish a second plant in country j , it must be the case that (12.5') is at least as large as (12.5):

$$\frac{1}{\sigma} \left(\frac{p^j}{P^j} \right)^{1-\sigma} Y^j - w^j \alpha^j \geq \frac{1}{\sigma} \left(\frac{p^i \tau^{ij}}{P^j} \right)^{1-\sigma} Y^j. \quad (12.6)$$

Similarly, in order to be a horizontal multinational, it must be the case that it is *more profitable* to open a plant in country i than to simply export from the single plant in country j . The condition to ensure this is

$$\frac{1}{\sigma} \left(\frac{p^i}{P^i} \right)^{1-\sigma} Y^i - w^i \alpha^i \geq \frac{1}{\sigma} \left(\frac{p^j \tau^{ji}}{P^i} \right)^{1-\sigma} Y^i, \quad (12.6')$$

where α^i is plant fixed costs in country i . The left-hand side of (12.6') is the profits from operating its plant in country i , while the right-hand side is the profits from exporting from country j to i .

So in order to have a horizontal multinational operating plants in both countries, (12.6) and (12.6') must both hold. These conditions can be reexpressed by moving the fixed costs to the right side, multiplying each equation by σ and multiplying them together, in which case we obtain a *single necessary condition* to ensure that the multinational maintains plants in both countries:

$$\begin{aligned} & [1 - (\tau^{ij} \tau^{ji})^{1-\sigma}] (Y^i Y^j) \left(\frac{p^i p^j}{P^i P^j} \right)^{1-\sigma} \\ & \geq \sigma w^i \alpha^i Y^i \left(\frac{p^i \tau^{ij}}{P^j} \right)^{1-\sigma} + \sigma w^j \alpha^j Y^j \left(\frac{p^j \tau^{ji}}{P^i} \right)^{1-\sigma} + \sigma^2 w^i \alpha^i w^j \alpha^j. \end{aligned} \quad (12.7)$$

This condition is more likely to hold in the regions of parameters given by the following theorem.

THEOREM (MARKUSEN AND VENABLES 1998, 2000; MARKUSEN 2002)

Horizontal multinationals will be more likely when (a) transport costs are higher; (b) plant-specific fixed costs are lower; (c) GDPs are higher or more similar across countries.

Conditions (a) and (b) of this theorem follow by inspection of condition (12.7).¹² Condition (c) requires some additional explanation. Notice that product of GDPs can be rewritten as

$$Y^i Y^j = (Y^i + Y^j)^2 s^i s^j = (Y^i + Y^j)^2 [1 - (s^i)^2 - (s^j)^2]/2, \quad (12.8)$$

where $s^i = Y^i/(Y^i + Y^j)$ and $s^j = Y^j/(Y^i + Y^j)$ are county i 's and j 's shares of GDP, with $(s^i + s^j) = 1$. It follows that $1 = (s^i + s^j)^2 = (s^i)^2 + (s^j)^2 + 2s^i s^j$, from which the last equality in (12.8) is obtained. The value $[1 - (s^i)^2 - (s^j)^2]$ is a size dispersion index of the two countries, and it takes its highest value (of one-half) when the countries are the same size. Denoting this index by D^{ij} , we can divide (12.7) by $(Y^i + Y^j)$ and rewrite it as

$$\begin{aligned} & [1 - (\tau^{ij} \tau^{ji})^{1-\sigma}] (Y^i + Y^j) \left(\frac{D^{ij}}{2} \right) \left(\frac{p^i p^j}{P^i P^j} \right)^{1-\sigma} \\ & \geq \sigma w^i \alpha^i s^j \left(\frac{p^i \tau^{ij}}{P^j} \right)^{1-\sigma} + \sigma w^j \alpha^j s^i \left(\frac{p^j \tau^{ji}}{P^i} \right)^{1-\sigma} + \frac{\sigma^2 w^i \alpha^i w^j \alpha^j}{(Y^i + Y^j)}. \end{aligned} \quad (12.7')$$

¹² Notice that the left-hand side of (12.7) is increasing in the transport costs τ^{ij} or τ^{ji} , while the right-hand side is decreasing. In addition, the right-hand side is increasing in the fixed costs α^i or α^j .

We see that the left-hand side of (12.7') is larger when the sum of GDPs ($Y^i + Y^j$) is larger, and the right-hand side is smaller. In addition, the left is larger when the size dispersion index is higher, meaning that the countries are of similar size.¹³ Under these conditions, (12.7') is more likely to hold.

Markusen and Venables (1998) and Markusen (2002, chapter 5) demonstrate the above theorem in a model where firms produce a homogeneous good under Cournot-Nash competition, while Markusen and Venables (2000) and Markusen (2002, chapter 6) use monopolistic competition with a CES utility function. In both cases, the multinationals employ high- and low-skilled labor. So in addition to the similarity of GDPs that we obtain in condition (c), they also find that horizontal multinationals are more likely to hold when the *relative endowments* of high- and low-skilled workers are *similar*. Notice that this is the opposite of the result obtained from Helpman's (1984c) model: since vertical multinationals occur in the regions $O^1S'Q'$ and $O^2R'S'Q'$ of figure 12.4, this requires that the relative endowments of skilled/unskilled workers are sufficiently *different*.

ESTIMATING THE KNOWLEDGE-CAPITAL MODEL

The vertical and horizontal multinationals can be combined in a unified framework, which Markusen (2002, chapters 7, 8) refers to as the “knowledge-capital” model. This term is used because the knowledge created at the headquarters is utilized at plants in multiple countries: it has the nature of a public good within the firm. This is the intangible asset that we identified in the introduction of the chapter as essential to any multinational firm. It is assumed that the ratio of skilled/unskilled labor used in headquarters services *exceeds* that used in plant production by multinationals, which in turn *exceeds* that used in production of the homogeneous competitive good. Then Markusen (2002, chapter 8) finds that increased investment by multinationals raises the relative wage of skilled/unskilled workers in the skill-abundant country, and possibly in the other country as well. The reason for this raise in wages is that the skill-abundant country specializes more in headquarters services, while the other country specializes more in production by multinationals, which raise the relative demand for skilled labor in *both* countries. This outcome applies to either horizontal or vertical investment, and is very similar to what we obtained from offshoring in chapter 4.¹⁴

The knowledge-capital model has been estimated by Carr, Markusen, and Maskus (2001), and hypothesis tests to distinguish the horizontal from the vertical model are performed by Markusen and Maskus (2001, 2002) as well as Blonigen, Davies, and Head (2003).¹⁵ These studies differ somewhat in their precise specifications, and we will combine their various results in our discussion below.¹⁶ The dependent variable

¹³ Changes in the country shares will also affect the right-hand side of (12.7'), but only by a small amount if fixed costs, transport costs, and wages are similar across countries.

¹⁴ Notice that the skilled/unskilled labor used in plant production in the knowledge capital model lies *in between* the factor intensities of headquarters services and of the homogeneous competitive good. This is similar to factor intensity of the “borderline” activity that was transferred between countries in the outsourcing model of chapter 4, but differs from the assumption of Helpman (1984c).

¹⁵ Brainard (1997) also estimates a horizontal versus multinationals model in which firms trade the benefits from being close to customers against the benefits of concentrating production in one location.

¹⁶ In the estimation of (12.9), Blonigen, Davies, and Head (2003) measure the skill difference and GDP difference as absolute values instead. They argue that the absolute values are closer to the specification in

used is the real sales of foreign affiliates of multinational firms, including both their local and export sales. For horizontal multinationals the local sales are most important, while for vertical multinational the export sales (back to the parent corporation) are most important.

The equation estimated by Carr, Markusen, and Maskus (2001) is

$$\begin{aligned} \text{Real Affiliate Sales} = & \beta_0 + \beta_1(\text{GDP}^i + \text{GDP}^j) + \beta_2(\text{GDP}^i - \text{GDP}^j)^2 + \beta_3(\text{Skill}^i - \text{Skill}^j) \\ & + \beta_4(\text{GDP}^i - \text{GDP}^j)(\text{Skill}^i - \text{Skill}^j) + \beta_5(\alpha^j) + \beta_6(\tau^j) + \beta_7(\tau^j)(\text{Skill}^i - \text{Skill}^j)^2 \\ & + \beta_8(\tau^i) + \beta_9 \text{Distance}^{ij}, \end{aligned} \quad (12.9)$$

where country i refers to the headquarters of a multinational and country j to the location of its a foreign affiliate. The first variable on the right side of (12.9) is the sum of country GDPs, which we expect to be positively related to the presence of horizontal multinationals, and therefore real affiliate sales ($\beta_1 > 0$). Likewise, the second variable is the squared difference in real GDPs, which measures the size dispersion, and which we expect to be negatively related to horizontal multinationals ($\beta_2 < 0$). The variable α , reflecting host country investment costs, and τ^i and τ^j , reflecting trade costs, are included as controls.

What about for vertical multinationals? In the model of Helpman (1984c), neither the sum nor the difference of GDPs determines the magnitude of affiliate exports back to the parent country; rather, it is the relative factor endowments that determine the magnitude of these exports. For example, in figure 12.4, the affiliate exports are proportional to the line segments VV'' and $V''V'$, which equal the amount of unskilled labor used for affiliate production. The positions of the endowment points V' and V'' relative to the original FPE set are what determine these exports, not the country GDPs.

Thus, one test of the horizontal versus the vertical model is to compare the expected coefficients $\beta_1 > 0$ and $\beta_2 < 0$ for the horizontal case, with $\beta_1 = \beta_2 = 0$ for the vertical case. The estimation of (12.9) is performed with a dataset on inbound and outbound U.S. foreign direct investment, from 1986 to 1994. For each country pair, either i or j denotes the United States, with the other index being the location of either a U.S. affiliate or a foreign parent company with an affiliate in the United States. The hypotheses that $\beta_1 > 0$ and $\beta_2 < 0$ are strongly confirmed in the estimation, which supports the horizontal over the vertical multinational model.

The next two variables in (12.9) are related to the skill differences between the parent and host country, where Skill^i is measured by the share of the labor force in certain skilled occupations. In the horizontal model, it is expected that investment and affiliate sales are declining in the endowment differences of the countries, and this effect is more pronounced when the country GDPs also differ. Thus, the hypotheses $\beta_3 < 0$ and $\beta_4 < 0$ are obtained from the horizontal model. In contrast, in Helpman's model vertical investment is higher when the endowments of the countries differ more, so that $\beta_3 > 0$ is predicted. Markusen (2002, chapter 12) also finds in his simulations that countries that are both *small* and *skilled-labor abundant* relative to their trading partners, such as the Netherlands, have the greatest concentration of headquarters for

Markusen and Maskus (2002) and Markusen (2002, chapter 12), who allow for separate coefficients on the skill difference depending on its sign, along with separate coefficients on its interaction with the GDP difference.

vertical multinationals. Thus, the interaction between relative country size and the skill endowment is positively related to vertical investment, so that $\beta_4 > 0$ also.

In the estimation of (12.9), the sign pattern $\beta_3 < 0$ and $\beta_4 < 0$ is strongly confirmed, so once again, the horizontal model is accepted and the vertical model is rejected. This outcome is perhaps not surprising in view of the fact that the majority of foreign investments worldwide are between *industrial* countries, and not between developed and developing countries (Markusen 1995; and Lipsey 1999, 2003). These investments would be more typical of horizontal rather than vertical linkages.

HORIZONTAL MULTINATIONALS WITH HETEROGENEOUS FIRMS

Helpman, Melitz, and Yeaple (2004) have extended the model of horizontal multinationals to incorporate heterogeneous firms, as in the Melitz (2003) model discussed in chapter 6. For convenience, they assume there is a homogenous good sector in every country. This good is freely traded and because firms producing it are assumed to have the same productivity in every country, its price and also wages are equalized across countries so we normalize $w^* = w = 1$. We further assume that the home and foreign countries are symmetric, so that $L = L^*$. Besides producing at home and selling at home and abroad, every firm can pay fixed costs f_i to get access to the foreign market through foreign direct investment (FDI). This means that instead of exporting, a firm can build a production facility in the destination country. In this case the firm faces the foreign wage, but does not have to pay transportation costs.

We use CES demand as in the Melitz model in chapter 6, and adopt other notation from that model. Then the profits from FDI are

$$\pi_I(\varphi) = \underbrace{\left[\frac{L\sigma^{-\sigma}}{(P_d^*)^{1-\sigma}(\sigma-1)^{1-\sigma}} \right]}_{B_d} \varphi^{\sigma-1} - f_i.$$

We assume that $f_i > (\tau)^{\sigma/(\sigma-1)} f_x > f_d$, so that the fixed costs of FDI are higher than for exporting (including the iceberg costs τ) or for domestic production. We define the term B_d as shown above, which acts like a shift parameter in demand. This profit function can be graphed in figure 12.5, which builds upon our treatment in chapter 6 (see figure 6.2).

The term B_d is identical to the demand shifter that applies to domestic sales, because we have assumed that the home and foreign countries are symmetric, $L = L^*$. It follows that the lines π_I and π_d are parallel, but π_I lies below π_d because of the higher fixed costs of FDI. Furthermore, the term B_d exceeds the demand shifter that applies to foreign export sales, since those sales are subject to the iceberg trade costs $\tau > 1$. It follows that the line π_I is steeper than π_x , but again has a lower intercept under our assumption on fixed costs. These relationships between π_d , π_x , and π_I imply that selling in only the domestic market is profitable only for low-productivity firms, while exporting is profitable for higher-productive firms, and the most productive firms would choose FDI instead of exporting (both the exporters and firms choosing FDI also sell in the domestic market). The reason this ranking is obtained is because firms with the highest productivity are able to set lower prices and hence have higher demand for their product. It allows them to use the advantage of lower marginal costs that FDI gives, and they use the higher demand to cover higher fixed costs. The cut-off between

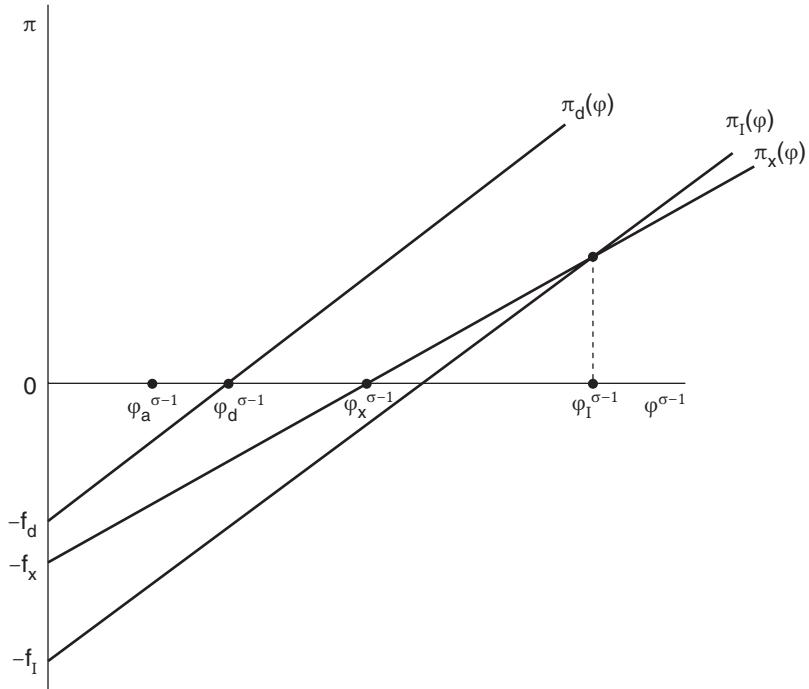


Figure 12.5

firms that export and firms that engage in FDI is denoted as $\varphi_I^{\sigma-1}$ in figure 12.5. This cut-off can be found from the equation

$$(1 - \tau^{1-\sigma}) \varphi_I^{\sigma-1} B_d = f_I - f_x. \quad (12.10)$$

To test this model empirically, let s_x and s_I be market shares abroad of exporters and firms performing FDI. Then Helpman, Melitz, and Yeaple (2004) show that the share of exports relative to foreign sales by FDI equals $s_x/s_I = \tau^{1-\sigma} \{ [V(\varphi_x)/V(\varphi_I)] - 1 \}$, where $V(\varphi) = \int_0^\varphi z^{1-\sigma} dG(z)$ and $G(\varphi)$ is the distribution function for firms' productivities. It follows from this relationship that when the distribution of productivities is more dispersed, the share of exports relative to FDI in that industry will decline. Helpman, Melitz, and Yeaple (2004) find that this hypothesis is supported using data on exports and FDI activities across U.S. industries: those industries with a greater variance of productivities have a greater share of foreign sales by FDI relative to exports. They conclude that firm-level heterogeneity is an important determinant of relative export and FDI flows.

ORGANIZATION OF THE FIRM

Both the vertical and horizontal multinational models involved a firm deciding whether to open an additional plant in a foreign country, but neither of these models addressed the question of whether the foreign plant had to be *owned by* or *internalized within* the multinational or not. What would prevent the home firm from entering

into a joint venture or some other contractual arrangement with a foreign firm rather than owning a foreign plant? In order to explain this, we need to add a richer structure to the models we have considered so far. Generally, Perry (1989, 187) identifies three reasons for firms to vertically integrate: technological externalities; market imperfections, which can be due to the exercise of market power or imperfect information; and transactions costs. We shall discuss each of these in turn.

Technological Externalities

Increasing returns to scale are one reason for firms to integrate: if two firms with given resources can produce more by combining, then there is a technological argument for doing so. Perry (1989, 187) uses the example of an integrated steel mill saving resources by not having to reheat its manufactured steel in order to make sheet metal. The model of horizontal multinationals we considered above had such a technological externality: the fixed cost of headquarters services had to be paid only in one country, whereas production could occur in multiple countries. It was assumed, however, that these headquarters services could be transferred only *within* a firm, thereby ruling out other contractual relationships between the firms. It would be preferable to derive the ownership structure endogenously in a framework where the firms might be integrated within a multinational, but might also engage in other contractual relationships.

Market Power

A second reason for firms to integrate is to offset the exercise of market power. In the classic paper entitled “The Division of Labor Is Limited by the Extent of the Market,” Stigler (1951) describes the example of a coal cartel at the end of the nineteenth century, when steel companies kept acquiring mines to avoid paying the cartel’s prices. That paper is primarily devoted to discussing a theorem of Adam Smith, whose statement is used as the title, that firms will become less integrated (spinning off specialized activities) as the market grows. The example of backward integration by the steel companies is used as an instance where vertical integration is caused by a failure of the competitive price system, rather than by the extent of the market. So firms may integrate to offset the exercise of market power by suppliers, and that activity can interact with the size of the market. Perry (1989) describes several models of that type within a single country. The closest application to multinationals comes from the studies of Horstmann and Markusen (1992) and Horn and Persson (2001), and further work in this area would be desirable.

Another case where the exercise of market power appears to be important is in the formation of *business groups* that are found in many countries: examples are the *keiretsu* in Japan and the *chaebol* in South Korea. The existence of these groups is often attributed to capital market imperfections, whereby the groups form to provide financing to their member firms. But Feenstra, Huang, and Hamilton (2003) argue that these groups, especially in Taiwan and Korea, can alternatively be described by their desire to sell inputs internally within a group at marginal cost, while charging optimally high markups for external sale of inputs.¹⁷ By allowing business groups to optimally choose these markups, while also having free entry of business groups, it means

¹⁷ Alternatively, Spencer and Qui (2001) develop a transactions cost model of the *keiretsu* in Japan, which is empirically evaluated by Head, Ries, and Spencer (2004).

that there are *multiple equilibria* in the structure of business groups: an economy can have a small number of large business groups, charging high prices for inputs sold to other groups; or a larger number of smaller groups, charging lower prices for the sale of intermediate inputs. Therefore, historical conditions must be considered to determine which equilibria occurs in a country, and Feenstra and Hamilton (2006) describe why South Korea has ended up with large business groups, while Taiwan has smaller groups. The idea of market power leading to vertical integration has also been used in a model of multinationals by Garett (2013).

Transactions Costs

A general reason for the existence of firms provided by Ronald Coase (1937) is that there are costs involved with making transactions on the markets, and when these costs become too high, the activities are instead internalized within the firm. This idea was greatly elaborated by Oliver Williamson (1975, 1985), and is associated with the twin concepts of “asset specificity” and the “holdup problem.” To explain these concepts, suppose that firm A needs some new input. It could ask firm B to produce it, but that would be expensive for B to design and test. If firm B managed to successfully produce it, and then sell it to A, then what guarantee does firm B have that it will be paid enough to cover its costs of design, let alone its costs of production? If firm A and B could contract in the beginning to specify what the input would cost to develop and produce, that would resolve the problem, but it may be difficult to foresee in the beginning what these costs would be.

In the absence of complete contracts, firm B will be left with an input which may or may not be useful to buyers other than firm A: this is called *asset specificity*. We can presume that the better suited the input to firm A’s needs, then the less valuable it is to alternative buyers. But in that case, firm B has very little bargaining power in refusing to sell the input to firm A: this is the “holdup problem.” Because of the holdup problem, firm B might decide to produce an input that is not very well suited to firm A, but instead can be sold to other purchasers. So the incomplete contracts lead to an inefficient outcome. If this inefficiency is too great, firm A might decide to just design the input itself; that is, it will vertically integrate into the production of that good, and thereby become a bigger firm.

From this line of reasoning, we can see that the decision of a firm to integrate depends on the cost of making market transactions under incomplete contracts, and in particular, asset specificity and the holdup problem. These insights of Coase and Williamson have been formalized into models of the firm, and key references are Grossman and Hart (1986), Hart and Moore (1990, 1999), Holmstrom and Milgrom (1994), and Aghion and Tirole (1997). These ideas have also been incorporated into general equilibrium models of trade, and we shall discuss the transactions cost model of Antràs (2003), as extended by Antràs and Helpman (2004) and Fernandes and Tang (2012).

In the Antràs-Helpman model, a firm makes a decision about where and how to engage in a portion of its production process: the assembly of components, for example, or research and development. We represent the choices of this firm in figure 12.6. Along the side of the diagram, the firm can engage in this process at home (in the North) or in a foreign country (South). Along the top of the diagram, the firm can decide whether to keep the process within the firm, which is “vertical integration,” or whether to “outsource,” which is contracting with another firm. In the case where the process is kept at home, we have either “domestic integration” within the firm, or

		Ownership of Plant, $k \in \{V, O\}$	
		Vertical Integration	Outsource
Location of the Plant, $\ell \in \{N, S\}$	Home (North)	Domestic Integration $\beta_V^N = (\delta^N)^\alpha + \beta[1 - (\delta^N)^\alpha]$	Domestic Outsourcing $\beta_O^N = \beta$
	Foreign (South)	Multinational $\beta_V^S = (\delta^S)^\alpha + \beta[1 - (\delta^S)^\alpha]$	Foreign Outsourcing $\beta_O^S = \beta$

Figure 12.6: Organizational Choices for the Firm

“domestic outsourcing” outside the firm. Conversely, if the production process is sent abroad, then the production occurs either within a vertically integrated “multinational firm,” or in the final case where production is sent abroad outside the firm, we have “foreign outsourcing.” This exhausts the four possibilities.

We might ask where does “offshoring,” as we discussed in chapter 4, enter the picture? A narrow definition of offshoring, used sometimes in business circles, is when a firm sends a portion of its production process abroad but keeps it in-house, so it becomes a multinational: this is the lower-left cell in figure 12.6. Under this narrow definition, offshoring and foreign outsourcing are two distinct strategies of the firm. But a more common definition of offshoring is that it encompasses both the multinational strategy *and* foreign outsourcing, in other words, it refers to any transfer of production overseas, whether it is done within or outside the firm. Under this definition, offshoring encompasses both choices in the lower row of figure 12.6, and this is how we used the term “offshoring” in chapter 4. We are now interested in not only offshoring, but also *all* the choices for the firm as shown in figure 12.6: where to locate, denoted by $\ell \in \{N, S\}$ for a North and South country; and to vertically integrate or not, denoted by $k \in \{V, O\}$ for vertical integration and outsourcing.

Choice of Ownership and Location

The starting point for all transactions costs models is the concept of Nash Bargaining. Proposed by John Nash (1950), this approach gives a simple solution to a bargaining game satisfying certain axiomatic properties. In particular, suppose that there are two agents labeled as h and m . Specifically, we will think of h as the “headquarters” of a company that is negotiating with a plant “manager” m to build a product. For now, we will keep the application simple and suppose that h and m are bargaining over some reward, or rent, denoted by R . We specify that the two agents might have different bargaining weights, which are denoted by β for the headquarters and $(1 - \beta)$ for the manager.

These two agents also have different payoffs that they can get if the bargaining breaks down altogether. In that case, they receive their outside options or threat-point

payoffs of R_0^h and R_0^m . Under the *property rights* theory of the firm, the owner has the residual right of control to the output of the firm. Specifically, denoting the output of the firm by y , Antràs and Helpman (2004) suppose that under vertical integration the headquarters can recover a fraction δ_V^ℓ of the output, $\ell \in \{N, S\}$, with $0 < \delta_V^S < \delta_V^N < 1$. In other words the headquarters obtains more under vertical integration if the plant is located in its home country (North), than if the plant is located in a foreign country (South), where legal institutions are likely to be less familiar and receptive to claims of the headquarters. Antràs and Helpman suppose that the manager receives zero in either case, and that *both agents* have outside options of zero under the outsourcing regime.

In general, Nash proposed that the solution to the bargaining game between the two agents would be given by

$$\max_{R^h, R^m} (R^h - R_0^h)^\beta (R^m - R_0^m)^{(1-\beta)} \text{ subject to } R^h + R^m = R. \quad (12.11)$$

This objective function is like a Cobb-Douglas function with weights β and $(1-\beta)$. The reader will guess that the solution gives the payoffs $(R^h - R_0^h)$ and $(R^m - R_0^m)$ in the relative shares β and $(1-\beta)$, so that $(R^h - R_0^h) = \beta \tilde{R}$ and $(R^m - R_0^m) = (1-\beta) \tilde{R}$. These solutions will satisfy the constraint $R^h + R^m = R$ provided that $R^h + R^m = \tilde{R} + R_0^h + R_0^m = R$, so that $\tilde{R} = R - R_0^h - R_0^m$. The magnitude \tilde{R} is sometimes called the “quasi-rents” of the game, since it is the rents available to the agents over and above their threat-point payoffs. It follows that the Nash-bargaining solution is

$$R^h = R_0^h + \beta(R - R_0^h - R_0^m), \quad \text{and} \quad R^m = R_0^m + (1-\beta)(R - R_0^h - R_0^m).$$

In the Antràs-Helpman model, the outside option of the headquarters is a fraction δ_V^ℓ of the firm’s output under vertical integration, $\ell \in \{N, S\}$, and it turns out that this corresponds to earnings of $R_0^\ell = (\delta_V^\ell)^\alpha R$, for a parameter $0 < \alpha < 1$ that is defined below. All other outside options are zero. It follows that we can write the payoffs to the headquarters as $\beta_k^\ell R$ and to the manager as $(1-\beta_k^\ell)R$, where

$$\beta_k^\ell = \begin{cases} (\delta_V^\ell)^\alpha + \beta[1 - (\delta_V^\ell)^\alpha] & \text{for } k = V \\ \beta R & \text{for } k = O \end{cases}$$

This structure of payoffs is shown in figure 12.6. They are important, because they shows that the headquarters will take actions in order to maximize not the entire rent R , but rather, the portion $\beta_k^\ell R$. Likewise, the manager m will take actions in order to maximize the portion $(1-\beta_k^\ell)R$. We refer to β_k^ℓ and $(1-\beta_k^\ell)$ as the *ex-post bargaining weights* of the two agents, and they sum to unity by construction. Because we have assumed that $0 < \delta_V^S < \delta_V^N < 1$ and also that $0 < \alpha < 1$, it follows that

$$\beta = \beta_O^N = \beta_O^S < \beta_V^S < \beta_V^N \quad (12.12)$$

In other words, the bargaining weight for the headquarters is lowest when it outsourced to either the North or South, is greater when it is vertically integrated in the South (i.e., a multinational), and is greatest when it is vertically-integrated in the North (domestic integration).

The model is completed by specifying the production process that combines inputs from the headquarters h and the plant manager m . Suppose that their x_h and x_m are

combined in a Cobb-Douglas production function with shares η_h and η_m to produce a final good y ,

$$y = Ax_h^{\eta_h}x_m^{\eta_m}, \text{ with } \eta_h + \eta_m = 1.$$

The headquarters is always located in the North and faces the wage of w^N for the input x_h that is used, while the plant manager is located in either the North or the South (depending on the location of production), with wage w^ℓ , $\ell \in \{N, S\}$. We naturally assume that $w^N > w^S$ which gives the incentive to offshore production to the South. Note that the parameter A multiplying the production function reflects the technical efficiency of the plant. Antràs and Helpman (2004) allow for plants of varying productivity, in the style of Melitz (2003), but we ignore that extension here for simplicity.

This good is sold under monopolistic competition subject to the inverse demand,

$$p_y = By^{-1/\sigma}, \quad (12.13)$$

where $\sigma > 1$ is the constant elasticity of substitution and demand, and B is a shift parameter. It follows that revenue is

$$R(x_h, x_m) = p_y y = By^\alpha = BA^\alpha \alpha x_h^{\alpha \eta_h} x_m^{\alpha \eta_m}, \quad \alpha = \frac{(\sigma - 1)}{\sigma}.$$

Therefore, given the ownership structure $k \in \{V, O\}$, and the location of the plant $\ell \in \{N, S\}$, the problem for the headquarters is to choose x_h to maximize its surplus

$$x_h^*(k, \ell) \equiv \underset{x_h}{\operatorname{argmax}} \quad \beta_k^\ell R(x_h, x_m) - w^N x_h - f_k^\ell, \quad (12.14)$$

where f_k^ℓ are fixed costs for the headquarters. Antràs-Helpman assume a particular pattern on the fixed costs,

$$f_O^N < f_V^N < f_O^S < f_V^S. \quad (12.15)$$

In words, the headquarters' fixed costs are always lower in the North, and are also lower for outsourcing (when the manager handles problem of worker supervision, etc.) than for vertical integration. Finally, the manager solves the problem

$$x_m^*(k, \ell) \equiv \underset{x_m}{\operatorname{argmax}} \quad (1 - \beta_k^\ell) R(x_h, x_m) - w^\ell x_m. \quad (12.16)$$

Notice that in both the headquarters and manager's inputs are chosen at *below* the first-best levels, which would correspond to maximizing the total surplus

$$\underset{x_h, x_m}{\operatorname{argmax}} \quad R(x_h, x_m) - w^N x_h - w^\ell x_m - f_k^\ell. \quad (12.17)$$

Putting together these solutions for the headquarters and plant manager, we obtain the joint surplus in the Nash-bargaining solution

$$R[x_h^*(k, \ell), x_i^*(k, \ell), x_m^*(k, \ell)] - w^N x_h^* - w^\ell x_m^* - f_k^\ell. \quad (12.18)$$

We assume that the ownership and the location of the plant are chosen so as to maximize (12.18). The surplus in (12.18) will generally be less than that in (12.17), because the headquarters and manager are not acting so as to maximize total surplus in (12.17). The extent of inefficiency involved with the solutions to (12.14) and (12.16), and therefore the value of (12.18) as compared to (12.17), will depend on how important the headquarters services are as compared to the manager's services in the production function. Specifically, suppose that the headquarters' share η_h in the Cobb-Douglas production function is high. Then it will be important to ensure that the headquarters h is acting as close to first-best efficiency as possible. That outcome will be achieved by ensuring that the headquarters has a bargaining weight β_k^ℓ that is as high as possible. Conversely, if the manager's share $\eta_m = 1 - \eta_h$ in the Cobb-Douglas production function is high, then it will be important to ensure that the manager is acting as close to first-best efficiency as possible. So in that case we would want to choose the manager's bargaining weight $(1 - \beta_k^\ell)$ high so that the headquarters' weight β_k^ℓ is low.

These results are stated more formally as follows.

THEOREM (ANTRÀS 2003; ANTRÀS AND HELPMAN 2004)

- a) The Nash-bargaining weight $\beta_k^\ell(\eta_h)$, that maximizes (12.18) is an increasing function of η_h , the headquarters' share in costs, with $\beta_k^\ell(0) = 0$ and $\beta_k^\ell(1) = 1$.
- b) If the *headquarters* has a high share of costs η_h , then it is optimal to give it ownership, and the plant will be located in the North (domestic integration) if $f_v^N \ll f_v^S$, and located in the South (multinational) if $f_v^N \approx f_v^S$.
- c) If the *manager* has a high share of costs η_m , then it is optimal to give that agent ownership, and the plant will be located in the North (domestic outsourcing) if $f_o^N \ll f_o^S$, and located in the South (foreign outsourcing) if $f_o^N \approx f_o^S$.

Part (a) of this theorem has already been argued above. Parts (b) and (c) follow as corollaries. Whichever agent—the headquarters or the manager—has a higher share in costs will be given the greatest incentive by owning the production process. On the other hand, production is always cheaper if it is done in the South, because the wage is lower there. So regardless of which agent has ownership, production is located in the South unless the fixed costs of locating there are too high as compared to the fixed costs in the North.

There are several testable hypotheses that follow from these results. Antràs (2003) develops the idea that the headquarters is *capital intensive* as compared to the plant. If we focus on the case where the fixed costs in the North and South are similar, then we see from comparing parts (b) and (c) that goods that are more capital intensive (with high η_h) will be made by multinationals rather than produced abroad under foreign outsourcing. That hypothesis is strongly confirmed by looking at imports into the United States: in capital-intensive industries, greater shares of imports are *intra-firm* imports, that is, occurring within multinationals. Likewise, imports from capital-abundant countries are more likely to be intra-firm imports. These hypotheses are also confirmed by Nunn and Trefler (2008), who further add to the empirical analysis by allowing for differing productivities of firms, as examined theoretically by Antràs and

Helpman (2004).¹⁸ Recently, Antràs and Chor (2013) have extended the model to determine the outsourcing decision at different points along the value chain and provide evidence from the United States. Another empirical application, dealing with China, is developed in the next section.

Model of Processing Trade

An important extension of the Antràs-Helpman model is developed by Fernandes and Tang (2012), who apply it to the issue of processing trade in China. Chinese trade statistics keep track of whether imports and exports are for ordinary or processing trade, and *within* processing trade, whether it falls into the categories “processing with supplied materials” or “processing with imported materials.” Feenstra and Hanson (2005) called the former category “pure assembly,” meaning that the inputs are provided by the firm outside of China, and the latter category “import and assembly,” meaning that the firm in China makes the decision of what inputs to buy. We refer to “import and assembly” as allowing the Chinese firm/manager to have more *control* over the input decision, in the sense that they are selecting the input to purchase. So focusing just on processing trade, we obtain the organizational choices shown in figure 12.7.

Analogous to Antràs and Helpman (2004), we will let $k \in \{V, O\}$ denote the ownership structure of the plant in China, which we label as the “South”, and we now let $\ell \in \{N, S\}$ denote the party that has control over the inputs decision. For example, vertical integration ($k = V$) in the first column indicates that the foreign or “Northern” headquarters owns the plant, and so it is a foreign-owned firm operating in China. In the second column, outsourcing ($k = O$) means that the foreign firm (in the “North”) contracts with the Chinese firm (in the “South”) to assemble a product. In either case, the inputs used in the assembly can be under the control of the foreign (Northern) headquarters, $\ell = N$, or under the control of the Chinese (Southern) manager, $\ell = S$.

Feenstra and Hanson (2005) developed a theory of firm ownership and control that fits the choices shown in figure 12.7. Their theory uses the property rights approach, but is partial equilibrium. Recently, Fernandes and Tang (2012) developed a version of this theory that combines elements of both Antràs and Helpman (2004) and Feenstra and Hanson (2005). We present here a simplified version of the Fernandes and Tang (2012) model that allows us to analyze the organizational choices illustrated in figure 12.7.

As in our earlier discussion, we assume that the headquarters has Nash bargaining weight β and the manager has weight $(1 - \beta)$. Then we need to specify the outside options of the parties under each form of organization to determine the ex-post bargaining weights. Consider first the case of pure assembly. Denoting the revenue by R , in the case of disagreement we suppose that the Northern headquarters of a vertically integrated firm obtains the outside option $R_V^N = \delta^\alpha R$. The manager in this case receives an outside option of zero. Using the notation introduced above, it follows that the ex-post bargaining weight of the headquarters in the case of vertical integration and pure assembly is

$$\beta_V^N \equiv \delta^\alpha + \beta(1 - \delta^\alpha).$$

This ex-post bargaining weight is shown in the upper-left cell of figure 12.7.

¹⁸ See the empirical exercises to this chapter, where you are asked to reproduce regression results from Antràs (2003), as well as explore other evidence for how relationship-specificity and incomplete contracts affect exports from Nunn (2007) and Chen and Feenstra (2008).

Ownership of Chinese Plant, $k \in \{V, O\}$

		North/Foreign (Vertically-Integrate)	South/Chinese (Outsource)
		Foreign-owned firm, choosing its inputs $\beta_V^N = \delta^\alpha + \beta(1 - \delta^\alpha)$	Chinese-owned firm, foreign choice of inputs $\beta_O^N = \gamma^\alpha + \beta(1 - \gamma^\alpha)$
Control Over Input Choices, $\ell \in \{N, S\}$	North/Foreign (Pure-Assembly)	Foreign-owned firm, Chinese choice of inputs $\beta_V^S = \delta^\alpha + \beta(1 - \delta^\alpha - \gamma^\alpha)$	Chinese-owned firm, choosing its inputs $\beta_O^S = \beta(1 - \gamma^\alpha)$
	South/Chinese (Import-and-Assembly)		

Figure 12.7: Organizational Choices for Processing Trade

Continuing to assume that the plant in the South is vertically integrated, suppose that the Southern manager now makes the decision of what inputs to purchase, in the second row of figure 12.7. The Northern headquarters is still entitled to the residual output in the case of disagreement with the Southern manager, but Feenstra and Hanson (2005) assume that the manager also has an outside option that is nonzero in this case. Since the manager has control over the input purchases, this knowledge will be useful even if the Southern plant does not reach agreement with the Northern headquarters, in which case the manager can still earn the outside option $\gamma^\alpha R$. We assume that $\gamma < \delta$, reflecting the idea that experience in purchasing inputs is less important than plant ownership. In this case, the ex-post bargaining weight of the Northern headquarters is shown in the lower-left cell of figure 12.7,

$$\beta_V^S \equiv \delta^\alpha + \beta(1 - \delta^\alpha - \gamma^\alpha) < \beta_V^N.$$

Now consider the case where the Northern headquarters does not own the plant, in the second column of figure 12.7. Under pure assembly, the Northern headquarters will be making the choice of inputs, so we assume that it has some outside option which is again worth $\gamma^\alpha R$. The ex-post bargaining weight of the Northern headquarters becomes

$$\beta_O^N \equiv \gamma^\alpha + \beta(1 - \gamma^\alpha) < \beta_V^N,$$

where this inequality follows from our assumption that $\gamma < \delta$. So both β_V^S and β_O^N are less than β_V^N , which is the bargaining weight of the Northern headquarters when it owns the plant and controls the inputs choice. Fernandes and Tang (2012) make further assumptions to ensure that $\beta_V^S < \beta_O^N$, which holds when $\gamma < \delta$ but γ is not too small, in which case we obtain the ranking

$$\beta_V^S < \beta_O^N < \beta_V^N.$$

The final case we need to consider is when the plant is not owned by Northern headquarters *and* the local manager has control over the input decision. In this case, we suppose that the outside option of the headquarters is zero, while the Chinese manager can still earn $\gamma^a R$, so that the ex-post bargaining weights of the headquarters become

$$\beta_O^S \equiv \beta(1 - \gamma^a) < \beta_V^S < \beta_O^N < \beta_V^N. \quad (12.12')$$

Thus, the model of Fernandes and Tang (2012) gives a well-determined ranking of the ex-post bargaining weights of the headquarters, depending on the ownership of the plant and control over the inputs, though this ordering differs from that shown in (12.12). Following Antràs and Helpman, Fernandes and Tang further assume that plant production in each case is subject to fixed costs, and impose the following ranking:

$$f_O^S < f_V^S < f_O^N < f_V^N. \quad (12.15')$$

That is, the fixed costs are highest for the vertically integrated plant where input decisions are made by the foreign firm (in the North), reflecting the difficulty of governing that plant, and lowest for an outsourcing plant where input decisions are made entirely in China (the South).

The model is completed by specifying the production process that combines inputs from the headquarters h and the plant manager m . Suppose that their inputs x_h and x_m are combined in a Cobb-Douglas production function with the inputs x_i to produce a final good y ,

$$y = Ax_h^{\eta_h} x_i^{\eta_i} x_m^{\eta_m}, \text{ with } \eta_h + \eta_i + \eta_m = 1.$$

The headquarters is always located in the North and faces the wage of w^N for the input x_h that is used, while the plant manager is always located in the South, with wage w^S , and naturally we assume that $w^N > w^S$, which gives the incentive to offshore production to the South. Control over the input choice x_i can be given to the headquarters or the plant manager, and in either case the input is purchased on the world market at the price of p_i .

As before, the good is sold under monopolistic competition subject to the inverse demand shown in (12.13). It follows that revenue is

$$R(x_h, x_i, x_m) = p_y y = By^\alpha = BA^\alpha \alpha x_h^{\alpha\eta_h} x_i^{\alpha\eta_i} x_m^{\alpha\eta_m}, \quad \alpha = \frac{(\sigma-1)}{\sigma}.$$

Therefore, given the ownership structure $k \in \{V, O\}$, if control of the inputs choice rests with the Northern headquarters ($\ell = N$, or pure-assembly), then the problem for the headquarters is to choose both the inputs x_h and x_i to maximize its surplus

$$\begin{aligned} x_h^*(k, N) &\equiv \underset{x_h}{\operatorname{argmax}} \quad \beta_k^N R(x_h, x_m) - w^N x_h - p_i x_i - f_k^N, \\ x_i^*(k, N) &\equiv \underset{x_i}{\operatorname{argmax}} \quad \beta_k^N R(x_h, x_i, x_m) - w^N x_h - p_i x_i - f_k^N, \end{aligned}$$

while the manager solves the problem

$$x_m^*(k, N) \equiv \operatorname{argmax}_{x_m} (1 - \beta_k^N) R(x_h, x_i, x_m) - w^N x_m.$$

In contrast, if control over the input choices rests with the Southern manager ($\ell = S$, or import-and-assembly), then the headquarters chooses only x_h to maximize its surplus

$$x_h^*(k, S) \equiv \operatorname{argmax}_{x_h} \beta_k^S R(x_h, x_m) - w^S x_h - f_k^N,$$

while the manager chooses both the inputs x_m and x_i to maximize the surplus

$$x_m^*(k, S) \equiv \operatorname{argmax}_{x_m} (1 - \beta_k^S) R(x_h, x_i, x_m) - w^S x_m - p_i x_i,$$

and

$$x_i^*(k, S) \equiv \operatorname{argmax}_{x_i} (1 - \beta_k^S) R(x_h, x_i, x_m) - w^S x_m - p_i x_i.$$

Putting together these solutions for the headquarters and plant manager, we obtain the joint surplus in the Nash-bargaining solution

$$R[x_h^*(k, \ell), x_i^*(k, \ell), x_m^*(k, \ell)] - w^N x_h^* - p_i x_i^* - w^S x_m^* - f_k^\ell. \quad (12.19)$$

Fernandes and Tang (2012) characterize the ownership-control structure that maximizes (12.19), depending on the share of headquarters services η_h needed in production. As before, it is convenient to solve for the bargaining weight β_k^ℓ that maximizes (12.19), and then choose the ownership structure that corresponds as closely as possible to that bargaining weight. This approach gives us the following result.

THEOREM (FERNANDES AND TANG 2012)

- (a) The Nash-bargaining weight $\beta_k^S(\eta_h)$ that maximizes (12.19) under import-and-assembly, is an increasing function of η_h , the headquarters' share in costs, with $\beta_k^S(0) = 0$ and $\beta_k^S(1) = 1$.
- (b) The Nash-bargaining weight $\beta_k^N(\eta_h + \eta_i)$ that maximizes (12.19) under pure-assembly is an increasing function of $\eta_h + \eta_i$, the headquarters' share in costs including the input costs, with $\beta_k^N(0) = 0$ and $\beta_k^N(1) = 1$.

Part (a) of this result is analogous to what Antràs and Helpman (2004) find in their analysis of ownership and location: when the headquarters' services are more important in costs, then it should have greater bargaining weight to obtain a value for surplus that is as high as possible. Part (b) generalizes this to the three-input case. Under pure-assembly, the headquarters chooses its own services along with the inputs, and the Nash bargaining weight β_k^N that maximizes (12.19) increases in the sum of $\eta_h + \eta_i$. Under import-and-assembly in part (a), the headquarters chooses only its own

services, and the Nash bargaining weight β_k^S that maximizes (12.19) is increasing in its own share η_h .

An immediate implication of this result is the following corollary.

COROLLARY

- (a) If the headquarters has a *high* share of costs η_h (or respectively, a *low* share of costs), then it is optimal to give ownership to the *headquarters h* (or respectively, the *manager m*).
- (b) If input search has a *high* share of costs η_i , then it is optimal to give control over input search to the *same* party that has ownership.

That is, as the headquarters share of costs η_h rises in the above theorem, parts (a) or (b), then so too will the Nash-bargaining weight of the headquarters, $\beta_k^S(\eta_h)$ or $\beta_k^N(\eta_h + \eta_i)$, that maximizes the joint surplus (12.19). In other words, the headquarters must be given a high ex-post bargaining weights so that it has good incentives. That is achieved by giving the headquarters ownership of the vertically integrated firm, putting it in the left-hand column of figure 12.7, where $\beta_v^N > \beta_o^N$ and $\beta_v^S > \beta_o^S$, depending on whether the headquarters (North) or Chinese manager (South) makes the input decisions.

Part (b) of this corollary addresses the question of which agent should be making the input decision: if the share of input η_i is high, then it should be the *same* agent that has ownership of the production process. This follows again from the logic of the above theorem. To see this, suppose that $\eta_i \rightarrow 1$. In part (a), this means that $\eta_h \rightarrow 0$ and so $\beta_k^S \rightarrow 0$. Then the ownership and control structure should give the headquarters the minimum ex-post bargaining power possible, which occurs when the Chinese manager owns the plant and makes the input decisions. Alternatively, suppose that $\eta_i \rightarrow 1$ in part (b), so that $\eta_h + \eta_i \rightarrow 1$ and so $\beta_k^N \rightarrow 1$. Then the ownership and control structure should give the headquarters the maximum ex-post bargaining power possible, which occurs when the headquarters owns the plant and makes the input decisions.

Feenstra and Hanson (2005) refer to the case where the same agent owns the plant and makes the input decisions as *concentrated ownership and control* (i.e., the Chinese manager operating under outsourcing with import-and-assembly, β_o^S , or the foreign headquarters operating under vertical integration with pure-assembly, β_v^N). This can be contrasted to *divided ownership and control*, when different parties each have responsibility (i.e., either the Chinese manager chooses the inputs under vertical integration, β_v^S , or the foreign headquarters chooses the inputs under Chinese ownership and outsourcing, β_o^N). From (12.12'), we have the following ranking of bargaining weights for the headquarters: $\beta_o^S < \beta_v^S < \beta_o^N < \beta_v^N$. Notice that the two extreme cases in this ranking have *concentrated ownership and control*: the same party owns the plant and it controls the input decision. These two cases are shown on the *diagonal* in figure 12.7. In contrast, the two middle cases in the above ranking illustrate *divided ownership and control*: different parties own the plant and control the input decision, which occurs along the *off-diagonal* in figure 12.7.

Estimating the Property-Rights Model for China

Let us now see how these predictions line up with the evidence from China. Predictions of the model can be directly applied to the data, denoting South as China and

TABLE 12.1: Processing Exports by Input Control and Factory Ownership Regime

Control over Inputs:	All Provinces		Interior Provinces	
	Ownership of Factory:		Ownership of Factory:	
	Foreign	Chinese	Foreign	Chinese
Foreign Buyer (pure-assembly)	0.083	0.271	0.177	0.110
Chinese Factory (import-and-assembly)	0.496	0.146	0.416	0.297

Notes: This table shows means for shares of processing exports by factory ownership (foreign versus Chinese) and input-control regime (pure-assembly versus import-and-assembly) for observations by year, industry, destination country, origin province, and trade zone status. The first two columns show results for all provinces; the second two columns show results for interior provinces (all provinces excluding the southern coastal provinces of Shanghai, Zhejiang, Fujian, Guangdong, Hainan). Data are for 1997–2002.

TABLE 12.2: Processing Exports depending on Trade Zone Status

Control over Inputs:	Inside SEZs		Outside SEZs	
	Ownership of Factory:		Ownership of Factory:	
	Foreign	Chinese	Foreign	Chinese
Foreign Buyer (pure-assembly)	0.081	0.076	0.084	0.317
Chinese Factory (import-and-assembly)	0.745	0.098	0.442	0.157

Source: Feenstra and Hanson (2005).

Notes: This table shows mean shares of processing exports by factory ownership and input-control regime for goods produced inside (columns 1 and 2) and outside (columns 3 and 4) of Special Economic Zones. See also notes to table 12.1.

North as the rest of the world (foreign). In table 12.1, we show the shares of processing exports from China (over 1997–2002) coming from the various ownership/control regimes. It is immediately apparent that for all provinces, shown in the left panel of the table, close to 75 percent of exports come from cases of *divided* ownership and control, that is, on the off-diagonal cells. The right panel of table 12.1 shows just interior provinces. In that case, there is only about 50 percent of exports coming from divided ownership and control, with a similar amount coming from *concentrated* ownership and control, on the diagonal of the right panel. From the above corollary, part (b), such cases of concentrated ownership and control can arise when the share of costs devoted to input search are high. We might hypothesize that input search in the interior provinces is more time-consuming and costly, due to thinner markets, than in the coastal provinces. That would explain why a greater proportion of exports from the interior provinces have the concentrated ownership and control regime, on the diagonal, than do the coastal provinces.

This difference between the coastal and the interior provinces also show up when we examine processing activity inside and outside of the Special Economic Zones (SEZ's), in table 12.2. For processing trade within the SEZ's, in the left panel of the table, only 18 percent of exports come from cases of concentrated ownership and control, on the diagonal; while for processing exports from outside the SEZ's there are somewhat more, with 24 percent of exports coming from cases of concentrated ownership and control, on the diagonal. This difference can be explained by higher costs of input search outside of the SEZ, due to a thinner market than within the SEZ. The other difference between

the left and right panels of table 12.2 is that within the SEZ a great deal of exports—nearly 75 percent—come from Chinese-owned factories that also choose the inputs under import-and-assembly; whereas outside of the SEZ, only about 45 percent of exports come from that regime. From the corollary, part (a), that difference is consistent with the Chinese processing firm having a higher share of costs within the SEZ, and therefore more likely to retain ownership of the plant to improve its incentives.

A final comparison is made between low-tech and high-tech industries, in tables 12.3 and 12.4. The low-tech industries are apparel and footwear, and the high-tech industries are office machines and electrical machinery. For the low-tech industries, between 24 and 41 percent of exports come from concentrated ownership and control (on the diagonal), whereas for the high-tech industries only 9 to 18 percent of processing exports come from this regime. That difference can arise from greater costs associated with input search in low-tech industries. That hypothesis may sound like a paradox, since high-tech goods presumably require high-quality components, but it is plausible that the market for high-tech inputs is more reliable in its quality than the market for low-tech components, and therefore it is easier to identify and purchase the high-tech inputs. The other difference between tables 12.3 and 12.4 is in the proportion of Chinese-owned factories that choose the inputs under import-and-assembly: more exports—69 versus 30 percent—come from such factories assembling

TABLE 12.3: Processing Exports for Low Value-Added Industries

Control over Inputs:	Apparel		Footwear	
	Ownership of Factory: Foreign	Ownership of Factory: Chinese	Ownership of Factory: Foreign	Ownership of Factory: Chinese
Foreign Buyer (pure-assembly)	0.225 (0.003)	0.286 (0.003)	0.104 (0.007)	0.178 (0.007)
Chinese Factory (import-and-assembly)	0.302 (0.003)	0.187 (0.002)	0.579 (0.009)	0.140 (0.003)

Notes: This table shows mean shares of processing exports by factory ownership and input-control regime for Apparel (SITC 84, columns 1 and 2) and Footwear (SITC 85, columns 3 and 4). See also notes to Table 12.1.

TABLE 12.4: Processing Exports for High Value-Added Industries

Control over Inputs:	Office Machines		Electrical Machinery	
	Ownership of Factory: Foreign	Ownership of Factory: Chinese	Ownership of Factory: Foreign	Ownership of Factory: Chinese
Foreign Buyer (pure-assembly)	0.033 (0.006)	0.211 (0.013)	0.070 (0.006)	0.246 (0.005)
Chinese Factory (import-and-assembly)	0.693 (0.016)	0.062 (0.008)	0.575 (0.007)	0.109 (0.004)

Source: Feenstra and Hanson (2005).

Notes: This table shows mean shares of processing exports by factory ownership and input-control regime for Office Machines (SITC 75, columns 1 and 2) and Electrical Machinery (SITC 77, columns 3 and 4). See also notes to table 12.1.

office machines than those assembling apparel (though these percentages are about the same for electrical machinery and footwear). From the corollary, part (a), that difference can arise from the Chinese processing firm having a higher share of costs when assembling office machines than when assembling apparel, due to the greater attention to quality needed in the production of office machines.

These examples in tables 12.1–12.4 illustrate the variation in the data that can be used to estimate the model of Fernandes and Tang (2012). They test the model using Chinese trade data, as do Feenstra and Hanson (2005) by using the shares of city-level exports from China broken down by the four ownership/control regimes: export share of X_V^N (vertical integration with Northern/foreign control of inputs), X_O^N (outsourcing to the Chinese firms with Northern/foreign control of inputs), X_V^S (vertical integration with Southern/Chinese control of inputs), and X_O^S (outsourcing with Southern/Chinese control of inputs). As in the above model, the share of costs devoted to headquarters services, input search, and processing activity are denoted by η_h , η_i , and η_m . Then we hypothesize the following relationship between exports and these cost-share variables:

$$\ln X_V^N = \alpha_{0V}^N + \underbrace{\alpha_{1V}^N}_{(+)} \frac{\eta_h}{\eta_m} + \underbrace{\alpha_{2V}^N}_{(+)} \frac{\eta_i}{\eta_m},$$

$$\ln X_O^N = \alpha_{0O}^N + \underbrace{\alpha_{1O}^N}_{(-)} \frac{\eta_h}{\eta_m} + \underbrace{\alpha_{2O}^N}_{(-)} \frac{\eta_i}{\eta_m},$$

$$\ln X_V^S = \alpha_{0V}^S + \underbrace{\alpha_{1V}^S}_{(+)} \frac{\eta_h}{\eta_m} + \underbrace{\alpha_{2V}^S}_{(-)} \frac{\eta_i}{\eta_m},$$

and

$$\ln X_O^S = \alpha_{0O}^S + \underbrace{\alpha_{1O}^S}_{(-)} \frac{\eta_h}{\eta_m} + \underbrace{\alpha_{2O}^S}_{(+)} \frac{\eta_i}{\eta_m}.$$

The sign pattern of the coefficients is obtained from the above corollary. From part (a), having a higher share of cost in headquarters services makes it more likely that the foreign firm has ownership of the Chinese plant, under vertical integration, leading to $\alpha_1 > 0$ in the first and third equations but $\alpha_1 < 0$ in the second and fourth. From part (b), having a higher share of costs in input search makes it more likely that the same party has ownership and control, leading to $\alpha_2 > 0$ in the first and last equations but $\alpha_2 < 0$ in the second and third.

So we see that the model leads to easily testable predictions on the coefficients of the relative input shares. But these equation are not really independent, because the share of exports coming from each ownership/control type sums to unity across the four possibilities. So Feenstra and Hanson take the first case and subtract it from the others, obtaining:

$$\ln X_O^N - \ln X_V^N = (\alpha_{0O}^N - \alpha_{0V}^N) + \underbrace{(\alpha_{1O}^N - \alpha_{1V}^N)}_{(-)} \frac{\eta_h}{\eta_m} + \underbrace{(\alpha_{2O}^N - \alpha_{2V}^N)}_{(-)} \frac{\eta_i}{\eta_m},$$

$$\ln X_V^S - \ln X_V^N = (\alpha_{0V}^S - \alpha_{0V}^N) + \underbrace{(\alpha_{1V}^S - \alpha_{1V}^N)}_{(?)}) \frac{\eta_h}{\eta_m} + \underbrace{(\alpha_{2V}^S - \alpha_{2V}^N)}_{(-)} \frac{\eta_i}{\eta_m},$$

$$\ln X_O^S - \ln X_V^N = (\alpha_{00}^S - \alpha_{0V}^N) + \underbrace{(\alpha_{10}^S - \alpha_{1V}^N)}_{(-)} \frac{\eta_h}{\eta_m} + \underbrace{(\alpha_{20}^S - \alpha_{2V}^N)}_{(?)} \frac{\eta_i}{\eta_m}.$$

Because we have chosen the first case—the vertically integrated plant with Northern/foreign control over the input choice (i.e., pure-assembly)—as the reference, we find that all coefficients in this revised system are negative or uncertain. The intuition is that higher headquarters costs or input shares make it less likely to observe any of the other cases (or uncertain) relative to the Northern ownership and control reference.

With these prediction on the coefficients, the system of equations can be estimated using data for Chinese processing trade by location (city in China), products (eight-digit Harmonized System, HS), and destination countries. Feenstra and Hanson call the variable η_h/η_m the “outbound value-added ratio” and η_i/η_m the “inbound value-added ratio.” The denominator of both these ratios, which is η_m , is measured by the share of Chinese value-added in processing exports at the one-digit Standard International Trade Classification (SITC) level. These shares are obtained from estimated provincial input-output tables for processing trade. The numerators are obtained from the markups on goods reexported to, or imported from, Hong Kong. Specifically, the markup on goods exported from China to Hong Kong and then reexported to the rest of the world is used to measure the headquarters services η_h , while the markup on goods imported by Hong Kong and then reexported to China is used to measure the input search costs η_i (both these markups are measured relative to the value of the good in China).

Estimation is then performed over all processing trade in China during 1997–2002, with the results shown in table 12.5. The first estimating equation above is shown in column (1) of table 12.5, where the expected negative signs on both coefficients are obtained, though only the coefficient on the inbound value-added ratio, η_i/η_m , is significantly different from zero. In the second equation in column (2), the expected signs are uncertain on the outbound ratio and negative on the inbound ratio, which is confirmed, though the estimate is not significant. For the third equation in column (3), we expect the outbound value-added ratio to have a negative coefficient but find instead that it is positive, while the inbound ratio turns out to have the negative coefficient, both of which are significant. In summary, these estimates provide some support for the property rights model as applied to Chinese processing trade.

To summarize, in this section we have explored a property-rights theory of the firm applied to its international decisions, as elaborated by Antràs (2003) and Antràs and Helpman (2004). Fernandes and Tang (2012) applied that general equilibrium model to the case of processing trade in China. Feenstra and Hanson (2005) developed a simpler partial-equilibrium model with some of the same features: in particular, the decision of who should own the firm and who should control the input decisions depends on the shares of the headquarters, Chinese manager, and input search in total costs. The empirical results from Feenstra and Hanson point to an important role for the property rights model, and transactions costs more generally, in understanding the structure of processing trade within China.

IMPERFECT INFORMATION

A final way to model the ownership decision is to introduce informational flows between the firms in the two countries, as done by Markusen (2002, chapters 13–15) using several models of imperfect information. In one case (Horstmann and Markusen

TABLE 12.5: Estimates of the Property Rights Model for China

	Northern/Foreign ownership & Chinese control over inputs (1)	Chinese ownership & Northern/Foreign control over inputs (2)	Southern/Chinese ownership & control over inputs (3)
Outbound	-0.007	0.013	0.035*
VA ratio	(0.014)	(0.021)	(0.017)
Inbound	-0.34**	-0.036	-0.17**
VA ratio	(0.054)	(0.072)	(0.050)
Years:	1997–99	2000–02	1997–99
Hong Kong	-0.061 (0.20)	0.24 (0.15)	0.16 (0.23)
SEZ	-0.21 (0.15)	0.47** (0.11)	0.19 (0.31)
North coast	0.32 (0.21)	-0.18 (0.16)	-0.25 (0.24)
Beijing area	0.11 (0.20)	-0.21 (0.15)	-0.12 (0.20)
Shanghai area	-0.14 (0.19)	0.57** (0.14)	-0.56** (0.22)
South coast	0.75** (0.29)	1.94** (0.20)	0.25 (0.37)
			2000–02

Source: Feenstra and Hanson (2005).

Notes: Standard errors are in parentheses.

* significant at the 5 percent level; ** significant at the 1 percent level.

$N=1,118,883$, across eight-digit harmonized system (HS) products, provinces, years, and destination markets. Additional control variables are as follows: Hong Kong equals one if the good is re-exported through Hong Kong, and the variable SEZ equals one if the HS products are produced in a Special Economic Zone (including high-technology development zones). All regressions include indicator variables for the year. The northern coast consists of Heilongjiang, Jilin, and Liaoning; the area around Beijing includes Beijing/Tianjin, Hebei, and Shandong; the area around Shanghai includes Shanghai, Jiangsu, and Zhejiang; and the southern coastal provinces consist of Fujian, Guangdong, and Hainan Island.

1987), they suppose that the good produced by the multinational has a quality that is imperfectly observed by consumers, so that resources must be spent to maintain its quality reputation. If the multinational contracts with a foreign firm, the latter will have an incentive to shirk on this activity, and so the parent company must transfer some rents to the foreign firm to avoid this shirking. If these rents become too costly, the multinational would just set up its own foreign affiliate instead. In another version of this story (Ethier 1986; Ethier and Markusen 1996; Markusen 2001), the imperfect information comes in the form of the foreign plant manager learning the production technology, and then potentially leaving the firm. Again, the parent company must transfer rents to the manager to prevent this from occurring, and if these become too high, the multinational will prefer to establish its own foreign affiliate.

Both of these models rely on *moral hazard*: the foreign manager or firm is taking some action that cannot be directly observed by the multinational, and it becomes expensive to control this activity. A third model (Horstmann and Markusen 1996) uses *asymmetric information* about market conditions abroad, which are known to a foreign agent but not the multinational. In that case, if the multinational enters into

a licensing agreement with the foreign agent, the multinational will have to transfer rents to elicit truthful reports of the foreign market conditions. If these rents are too high, the multinational will again establish its own foreign affiliate.

In a following section we will summarize another informational model due to Rauch and Trindade (2003), in which the integration of firms allows them to share information that each has about its own national markets.¹⁹ For example, joint ventures between firms in different countries can bring with them knowledge of the supplier and distribution networks in each country, and any other intangible assets owned by either firm. But there is some inherent uncertainty in the ability to find a good partner. The efficacy of achieving matches determines the extent to which markets are integrated, that is, the extent to which factor price equalization obtains. This model is not solved using either a moral hazard or asymmetric information framework but, as we shall see, relies on the “matching” of firms across countries. A more recent example of a matching model that is used to motivate the “network” structure of international trade is provided by Chaney (2014), with accompanying empirical evidence for French firms.

Matching of Firms Across Countries

The starting point for Rauch and Trindade (2003) is similar to the one-sector model discussed at the beginning of the chapter, where capital combines with labor to produce a homogeneous good. We now suppose, however, that two producers must form a joint venture in order to successfully obtain output. These producers may both be in the same country (as in the autarky equilibrium) or may come from different countries (in the trade equilibrium with joint ventures). Let us initially focus on the autarky equilibrium.

The amount of output obtained by two producers depends on their complementarity with each other. For convenience, we will array the producers in each country around a circle of circumference unity, as illustrated in figure 12.8. Let z_{ij} denote the shortest arc-distance between producers of type i and j , which measures their complementarity. Notice that the maximum value of z_{ij} is $1/2$. Then we assume that the output obtained by a joint venture of these producers is $Y_{ij} = f(L_{ij}, z_{ij})$, where L_{ij} is the labor used. We do not include capital in this production function since the amount available from each producer is equal and fixed. We assume that $f(L_{ij}, z_{ij})$ is increasing and homogeneous of degree one in (L_{ij}, z_{ij}) , so that producers farther away from each other obtain higher output.

If labor is available at the wage w , the profits obtained from this joint venture are

$$g(w, z_{ij}) \equiv \max_{L_{ij}} f(L_{ij}, z_{ij}) - wL_{ij}, \quad (12.20)$$

where we have normalized the price of the output at unity. Under our assumption that $f(L_{ij}, z_{ij})$ is homogeneous of degree one in its arguments, then $g(w, z_{ij})$ will be homogeneous of degree one in z_{ij} ; increasing the distance z_{ij} between producers will lead to an equiproportionate increase in labor hired, and in output and profits. It follows that we can write total profits earned as $g(w, z_{ij}) = z_{ij}\pi(w)$, where $\pi(w) \equiv g(w, 1)$. By the

¹⁹ The work of Rauch and Trindade (2003) builds upon that of Casella and Rauch (2002) and Rauch and Casella (2003).

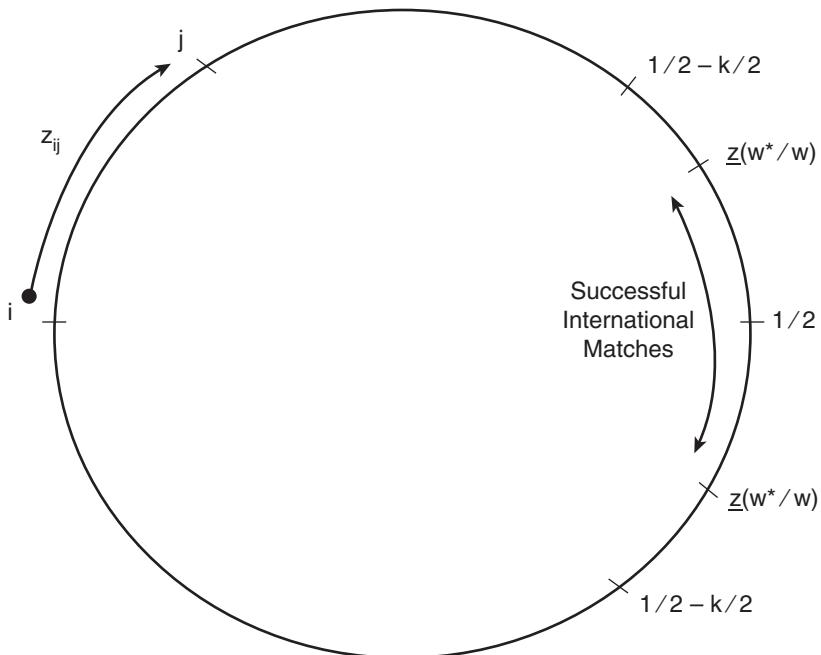


Figure 12.8

envelope theorem, the amount of labor hired by the joint venture is $L_{ij} = -z_{ij}\pi''(w) > 0$. Increasing the wage must reduce the labor hired, so that $dL_{ij}/dw = -z_{ij}\pi'''(w) < 0$, which means that $\pi''(w) > 0$ so π is convex. For convenience in our discussion, we follow Rauch and Trindade in assuming that $\pi(w)$ has the iso-elastic form $\pi(w) = bw^{-\gamma}$, with $b, \gamma > 0$. This form will apply if the function $f(L_{ij}, z_{ij})$ is Cobb-Douglas.²⁰

To solve for the autarky equilibrium, we need to match up producers in a single country. We assume that producers are uniformly distributed around the circle in figure 12.8, and that at each point on the circle there is a unit-mass of producers: so the total distribution of them can be thought of as a “unit cylinder.” The best matches are obtained when each producer has a joint venture with the one located on the opposite side of the circle, so that $z_{ij} = 1/2$. The profits earned from each of these matches are $z_{ij}\pi(w) = \pi(w)/2$. We assume that these are equally divided by the two participants, so each producer earns $\pi(w)/4$ in the autarky equilibrium.

The equilibrium wage is determined by equating labor demand and supply. One-half of the producers are matching with the other half (on the opposite side of the circle), and so the total number of matches is 1/2. It follows that home autarky labor demand is

$$-z_{ij}\pi'(w^a)/2 = -\pi'(w^a)/4 = L, \quad (12.21)$$

where the first equality is obtained because producers are matching with those on the opposite side of the circle ($z_{ij} = 1/2$), and the second equality sets labor demand equal to the endowment L to determine the autarky wage w^a .

²⁰ See problem 12.4.

Let us now introduce a foreign country, with the same technology but having a *higher* labor endowment $L^* > L$. Under autarky, the foreign wage satisfies $-\pi(w^*)/4 = L^*$, and would therefore be lower than at home: $w^* < w^a$ since $L^* > L$, using the convexity of π . In the trade equilibrium we allow joint ventures to form between producers *across* countries. We assume that in such a joint venture, the producers can employ labor in either country: given that foreign wages are lower, they will choose to locate there. This will tend to shift labor demand away from the labor-scarce home toward the labor-abundant foreign country, and therefore diminish the difference in their wages. How far will this process of wage equalization proceed? In the one-sector model used at the beginning of the chapter, *full* equalization of home and foreign wages would be achieved through capital mobility. That is, equalization of the rental rates across countries combined with identical technologies implies that wages are equalized, too.²¹ The goal of the matching model is to see how this result is modified when there are some frictions due to the information problems in finding a good partner.

To this end, suppose that producers can enter into a joint venture across countries. If these joint ventures prove to be unprofitable, then their “status quo” position is to enter into a joint venture domestically. Presuming that they still match with a domestic partner at the opposite side of the circle, then this would result in status quo payoffs of $\pi(w)/4$ and $\pi(w^*)/4$ for a producer in each country. Notice that these are evaluated at the equilibrium wages with international joint ventures, which will differ from the autarky wages. The international joint venture will yield profits of $z_{ij}\pi(w^*)$, where z_{ij} is the distance between the home and foreign producers and they employ labor in the foreign country. The two producers will divide the payoff $z_{ij}\pi(w^*)$ between themselves according to some rule, the details of which need not concern us.²²

Given that the *total* payoff to a home and foreign producer who match domestically is $[\pi(w)/4 + \pi(w^*)/4]$, it is apparent that the international joint venture is equally profitable if it occurs between a home and foreign producer who are distance \underline{z} apart, with

$$\underline{z}\pi(w^*) = [\pi(w)/4 + \pi(w^*)/4] \Leftrightarrow \underline{z}(w^*/w) = (1/4)[1 + \pi(w)/\pi(w^*)]. \quad (12.22)$$

Under our assumption that $\pi(w)$ takes on the iso-elastic form $bw^{-\gamma}$, with $b, \gamma > 0$, we can write $\pi(w)/\pi(w^*) = (w^*/w)^\gamma$ as an increasing function of the relative foreign wage, and likewise for $\underline{z}(w^*/w)$. We will confirm below that foreign wages are less than those at home in the trade equilibrium with joint ventures, so that $\underline{z} < 1/2$. Recalling that the maximum distance between any two producers is $1/2$, then those international matches in the range $z_{ij} \in (z, 1/2]$ will be strictly preferred to domestic matches.

Following Rauch and Trindade, let us add some structure to the international matches. We will assume that when a home producer i searches for a potential foreign partner j , the distance between them is randomly drawn from a uniform distribution defined over the interval $[1/2 - k/2, 1/2]$. Notice that when $k = 1$, this interval is $[0, 1/2]$, which means that partners are randomly distributed over the entire circle (recall again that the maximum distance is $1/2$). On the other hand, when $k = 0$, this interval becomes the point $1/2$, which is the best match possible. So the parameter $k \in [0, 1]$ indexes the extent of *uncertainty* in international joint ventures.

This structure of international matches is illustrated in figure 12.8. For a domestic agent located at the point labeled i , the best match would be at the point labeled $1/2$

²¹ See problem 12.5.

²² Rauch and Trindade (2003) assume that the payoff $z_{ij}\pi(w^*)$ is divided according to the Nash bargaining solution.

on the opposite side of the circle. The range of possible matches are in the interval $[1/2 - k/2, 1/2]$ on either side of the point $1/2$. On the other hand, the *successful* international matches must satisfy (12.22), which must be at least the distance $\underline{z}(w^*/w)$ away from the point i .²³

We can now compute that probability P that an international match is successful. This is simply the length of the interval as $[\underline{z}(w^*/w), 1/2]$ as compared to $[1/2 - k/2, 1/2]$, or

$$P = \frac{1/2 - \underline{z}(w^*/w)}{1/2 - (1/2 - k/2)} = \frac{1 - 2\underline{z}(w^*/w)}{k}$$

Thus, the probability of a successful international match is higher when the ratio of wages w^*/w is lower (meaning that \underline{z} is lower), or when the uncertainty k associated with international joint ventures is reduced.

To determine wages, we use labor market equilibrium in both countries. Home labor is demanded only by those domestic producers whose international matches were *not* successful, or a fraction $(1 - P)$ of home producers. The home demand for labor is therefore

$$-(1 - P)\pi'(w^*)/4 = L, \quad (12.24)$$

which can be compared to (12.21). Home labor demand has fallen from autarky due to producers employing labor abroad, so this results in a fall in the domestic wage, $w < w^a$.

Foreign labor demand has two components. The first is the demand from foreign producers whose international matches are not successful, and by analogy with (12.24), this component is $-(1 - P)\pi'(w^*)/4$. However, there is also labor demand from all the successful international matches. This is computed by taking the expected value of labor demand $-z_{ij}\pi'(w^*)$ over the successful matches in the range $z_{ij} \in [\underline{z}(w^*/w), 1/2]$. Therefore, the equilibrium condition in the foreign labor market is

$$-(1 - P)\pi'(w^*)/4 - \left(\frac{2}{k}\right) \int_{\underline{z}(w^*/w)}^{1/2} z_{ij}\pi'(w^*)dz_{ij} = L^*. \quad (12.25)$$

Notice that in the integral on the left of (12.25) we have multiplied by $(2/k)$, which is the density of the uniform distribution of international matches z_{ij} over its interval $[1/2 - k/2, 1/2]$, and converts that integral into an expected value.

Then, dividing (12.25) by (12.24), we obtain a single equation to determine the relative home wage:

$$(w/w^*)^{\gamma+1} \left[1 + \frac{8}{k(1-P)} \int_{\underline{z}(w^*/w)}^{1/2} z_{ij} dz_{ij} \right] = L^*/L, \quad (12.26)$$

where we have simplified this expression using the iso-elastic form for $\pi(w)$, so that $\pi'(w^*)/\pi'(w) = (w/w^*)^{\gamma+1}$. Despite the complexity of (12.16), it is readily shown that the left-hand side is a strictly increasing function of (w/w^*) , which starts at unity when

²³ It can be confirmed that $\underline{z}(w^*/w)$ must lie within the interval $[1/2 - k/2, 1/2]$ in equilibrium. If this were not the case (i.e., if $\underline{z}(w^*/w) < 1/2 - k/2$), then all home producers would enter into joint ventures with foreign producers, and the demand for home labor would be zero.

$(w/w^*)=1$ and approaches infinity as $(w/w^*) \rightarrow \infty$.²⁴ It follows that there is a unique solution for (w/w^*) from (12.26). This establishes the first part of the following theorem.

THEOREM (RAUCH AND TRINDADE 2003)

- (a) The trade equilibrium with international joint ventures exists and is unique;
- (b) As $k \rightarrow 0$ so there is no uncertainty about international partners, then $(w/w^*) \rightarrow 1$;
- (c) $dw/dk > 0$ and $dw^*/dk < 0$;
- (d) $d\ln(w/w^*)/d\ln(L^*/L) < 1/(1+\gamma)$, is increasing in k , and $\lim_{k \rightarrow 0} d\ln(w/w^*)/d\ln(L^*/L) = 0$.

Part (b) of this theorem is obtained by inspection of the equilibrium condition (12.26). As $k \rightarrow 0$, the term multiplying the integral goes to infinity (meaning that w/w^* goes to zero), *unless* the lower limit of integration approaches $1/2$, and so the integral goes to zero. The former case is easily ruled out as an equilibrium, so as $k \rightarrow 0$, we must have $\underline{z} \rightarrow 1/2$ in the lower limit of integration, which implies that the wages approach equality. Recall that this would be the outcome in the one-sector model used earlier in the chapter under perfect capital mobility. Therefore, the matching model approaches the outcome with perfect capital mobility as the uncertainty about international partners is eliminated. Part (c) of the theorem confirms that as international uncertainty increases, the equilibrium gets monotonically further away from wage equalization: raising k increases domestic wages and lowers foreign wages, as fewer joint ventures are pursued.

In part (d), Rauch and Trindade investigate the sensitivity of relative wages to changes in the labor endowments. Notice that from (12.26), if we held \underline{z} constant (so that P is also constant), then we readily obtain $d\ln(w/w^*)/d\ln(L^*/L) = 1/(1+\gamma) > 0$. Relative wages are increasing in the relative labor endowment of the other country for fixed \underline{z} , just as they are in autarky, because more labor abroad lowers foreign wages. In contrast, when wages are equalized across countries, $d\ln(w/w^*)/d\ln(L^*/L) = 0$, as occurs in the one-sector model under perfect capital mobility used earlier in the chapter. We can interpret this as saying that perfect capital mobility leads to complete integration of the labor markets. Then part (d) states that the matching model leads to greater integration of the labor markets than in autarky, and monotonically approaches the case of perfect capital mobility as uncertainty k approaches zero.

There are many potential empirical and theoretical applications of this matching model. It is well known that wages do respond to some degree to changes in local labor supplies (see Bernard and Jensen 2000), so that labor markets are not fully integrated across regions or countries. One explanation, according to this model, has to do with the informational frictions involved in the movement of producers. This outcome is certainly not restricted to the one-sector, two-factor setting, and it would apply equally well to the two-sector HO model. The finding of “factor price insensitivity” discussed in the HO model (see chapter 1) relies on countries being able to freely

²⁴ Notice that $\underline{z}(w^*/w)$ is decreasing in (w^*/w) , which reduces the lower limit of integration so the value of the integral on the left-hand side of (12.26) is increasing in (w^*/w) . Likewise, P in (12.23) is increasing in (w^*/w) , and then so is $1/(1-P)$, which appears on the left-hand side of (12.26).

export the goods that use intensively the factors that are growing in abundance. If there are informational frictions in the ability to match buyers and sellers in international trade, then factor price insensitivity would no longer hold: a growing labor force, for example, would necessitate increased exports of the labor-intensive goods, but if buyers are hard to find, then we would expect a terms of trade decline for those goods and therefore falling wages.

The idea that selling goods on international market requires a “match” between buyers and sellers is explored empirically by Rauch and Trindade (2002). Earlier work by Gould (1994) for the United States and by Head and Ries (1998) for Canada confirms the importance of immigrant networks in promoting trade with their home countries. Rauch and Trindade (2002) expand this to consider global bilateral trade, and show the importance of *Chinese ethnic networks* in enhancing trade. These networks are measured by the level or fraction of the population in a country that is of Chinese ethnic background, which is entered along with other typical regressors in a gravity equation.

Following earlier work by Rauch (1999), the trade flows in the gravity equation are distinguished depending on whether the products are homogeneous, with published prices; differentiated, with prices that are not available; or a middle category of “reference priced” products. It is found that the variable measuring Chinese networks enters most significantly in the bilateral trade equations for *differentiated* goods. This supports the idea that the informational difficulty in assessing these goods is more severe than in the case of homogeneous goods, and that having traders of the same ethnic background can overcome this difficulty to some extent.

Other empirical results that indicate the informational difficulties in international trade come from studies of the large traders in Hong Kong, such as Li and Fung (Slater and Amaha 1999). Their activities include matching firms or consumers who are the buyers of goods with the most appropriate producers of the goods, located in China or elsewhere in Asia. The value of these informational activities to the buyers and sellers involved will depend on the type of good (homogeneous versus differentiated), uncertainty in prices and quality, market “thickness”, and income levels of the buyers. Feenstra and Hanson (2001) find that these regressors are significant explanations of the markups charged on goods reexported through Hong Kong. This further indicates the importance of information in allowing for successful “matches” in international trade.

CONCLUSIONS

In this chapter we have reviewed a large literature on the role of foreign direct investment (FDI) and multinationals. The early contributions of Mundell (1957) and MacDougall (1960) dealt with capital flows between economies, though FDI can instead involve a change in ownership and location of firms, without any capital flows. The later contributions of Helpman (1984c) and Markusen (1984) focused on the decision to establish vertical or horizontal foreign subsidiaries. The fact that the parent company has invested in headquarters services gives it an asset that could be transferred to the subsidiary. It was assumed by Helpman and Markusen that this asset was kept internal to the firm, but of course, this need not be the case: a great deal of international trade occurs between unaffiliated firms. This circumstance creates both an empirical and theoretical challenge of explaining cross-firm versus within-firm transactions.

At the theoretical level, transactions are kept internal to a firm when the cost of pursuing them between firms is too great. This is the starting point for the *transactions cost* explanation for the organization of the firm. This idea has been elaborated in the industrial organization models and has also been incorporated into international trade models that we have reviewed. This marriage of the techniques in industrial organization and international trade is not just for theoretical convenience. Rather, the question of organization of production is motivated by actual trade patterns in our increasingly globalized world. International trade today is only rarely about sending a finished product from one country to another, as David Ricardo imagined between England and Portugal. Rather, it is more likely to involve the production of a good over sites in several countries, as companies engage in outsourcing or offshoring. By its very nature, outsourcing the production of a good involves a negotiation between two parties, in a market where competitive conditions will typically not hold. Especially when outsourcing occurs across country borders, it may also be difficult to specify all the payments and contingencies of production in a contract. So the recognition of outsourcing as a dominant feature of international trade leads us very quickly to explore the organization of how this production actually occurs, as we have explored in this chapter.

One objective of this research is to identify types of trading relationships that lie in between the two extremes of market transactions and vertical integration, such as when firms engage in a joint venture or other ongoing relationship. We can refer to such trade as a “link” between the two firms, and the set of links between buyers and sellers is referred to as a “network.”²⁵ As an extension of our study of multinationals, we are naturally led to a discussion of *networks* in international trade. These are relationships that represent more than market transactions, but less than full integration within a firm. There are many types of economic and social networks that can be considered (Rauch 2001; Rauch and Casella 2001), and we have discussed just one, arising from the “matching” of agents across international borders. Since firms will want to be linked with good partners abroad, we can expect that agents will find it profitable to undertake this matching activity. A model along these lines is developed by Rauch and Watson (2002), where individuals with expertise choose to become “international trade intermediaries.” This also creates a role for governments to support such intermediation activities, through trade fairs, associations, and so on. Evidently, these ideas are bringing the *trader* back into international trade, where he or she should have been all along!

PROBLEMS

Our analysis of capital flows in the HO model compared the equilibrium before and after the tariff, with a complete flow of capital. It is of interest to ask what happens to the welfare of the country imposing the tariff *during the transition*. That is, as the capital inflow moves the countries from point V^i to \bar{V}^i in figure 12.2, how does the welfare of the capital-importing country change? This question is addressed by Brecher and Diaz Alejandro (1977). To answer it we can work in reverse and ask what happens if capital *leaves* the country starting at the point \bar{V}^i , with the tariff on the capital-intensive good applied by country 1.

²⁵ This terminology follows Kranton and Minehart (2001).

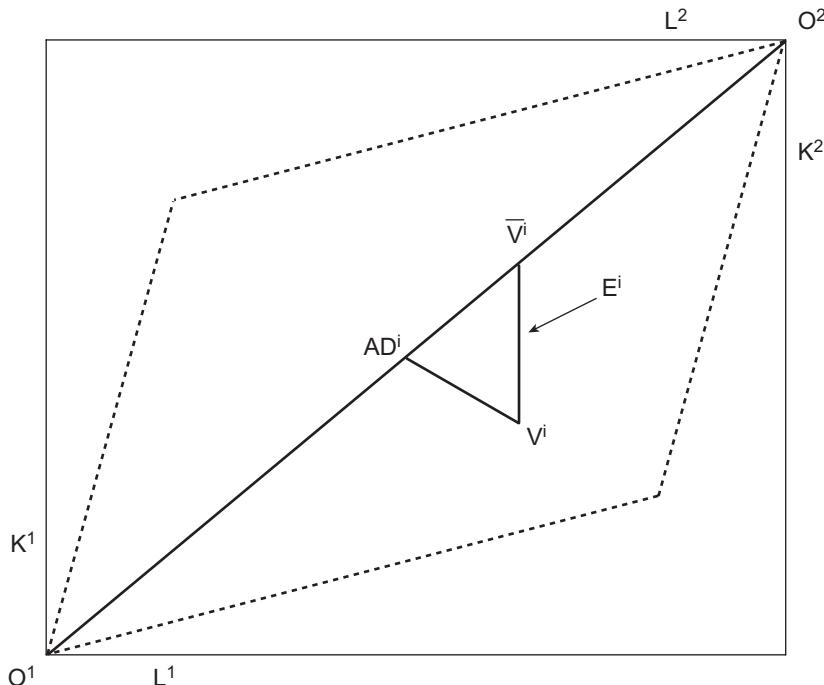


Figure 12.9

The outflow of capital from the point \bar{V}^i will reduce production of the capital-intensive good in country 1, and increase production of the labor-intensive good. Normally, this would lead to the capital-intensive good being imported. However, due to the tariff in that industry, there will be *some range* of capital outflow that *does not* lead to any imports; instead, country 1 remains in autarky but with a change in relative prices. This is illustrated in figure 12.9 by the point E^i , lying on the line between V^i and \bar{V}^i , such that for the range of endowments between \bar{V}^i and E^i , both countries remain in autarky due to the tariff applied by country 1. Now answer problems 12.1–12.3:

- 12.1 As capital enters country 1, going from \bar{V}^i to E^i , treat international prices as *fixed*. The welfare of a representative consumer in that country can be written as $V(p, I)$, where $p = p^* + t$ is the domestic relative price of imports and t is the tariff. As in earlier chapters, we assume that the exportable is additively separable in utility, so that consumption of the import good is $\partial V / \partial p = -d(p)$ and $\partial V / \partial I = 1$. Total national income equals GDP plus tariff revenue minus payments to foreign capital, so that $I = w(p)L + r(p)(K + K^*) + tm(p) - r(p)K^*$, where K^* is the foreign capital, $m(p) = d(p)y(p, L, K + K^*)$ are imports, and $y(p, L, K + K^*)$ is production of the import good, which depends on the labor and capital endowments with the usual Rybczynski effects. Notice that national income simplifies to $I = w(p)L + r(p)K + t[d(p) - y(p, L, K + K^*)]$.

Then differentiate consumer utility with respect to the capital inflow K^* , and sign and interpret this result, assuming that the import good is capital intensive.

- 12.2 The results in problem 12.1 apply for inflows of foreign capital between points \bar{V}^i and the point E^i . At E^i , country 1 is now in autarky. Suppose that capital *continues* to flow in:
- What is the effect of the capital inflow on equilibrium prices for goods, and on the rental?
 - What is the effect of the capital inflow on consumer welfare in country 1? (Hint: illustrate the effect on the rental as in figure 12.1).
- 12.3 Summarize your results from problems 12.1 and 12.2 by graphing the relationship between the capital inflow and welfare of the representative consumer in country 1. How does welfare at \bar{V}^i compare to the free trade welfare at \bar{V}^i ?
- 12.4 Suppose that $f(L_{ij}, z_{ij})$ take on the Cobb-Douglas form $f(L_{ij}, z_{ij}) = L_{ij}^\alpha z_{ij}^{(1-\alpha)}$, $0 < \alpha < 1$. Then solve equation (12.20), and show that $\pi(w) \equiv g(w, 1)$ can be written in the iso-elastic form $\pi(w) = bw^{-\gamma}$, with $b, \gamma > 0$.
- 12.5 In the one-sector model with perfect capital mobility, assume that the technologies are identical across countries and homogeneous of degree one.
- If the rental rates are equalized across countries, what does this imply about the labor-capital ratios used in production?
 - Therefore, how do wages compare across countries?
- 12.6 Rather than capital flows, let us instead consider inflows of labor to a country through immigration. We will contrast the results of two models:
- If there is only one sector in the economy, with fixed capital and fixed output price, then illustrate the effect of immigration on the home wages, and on welfare.
 - Suppose instead that we are in a Ricardian model, as discussed in chapter 1. Let labor flow from the foreign to the home country, and illustrate the impact on the welfare of workers at home (not including the new migrants). This Ricardian case is discussed in greater detail by Davis and Weinstein (2002b).
- 12.7 (a) Using the Pareto distribution from (6.15), show that the share of exports relative to foreign sales by FDI in Helpman, Melitz, and Yeaple (2004) equals

$$\frac{s_x}{s_I} = \tau^{1-\sigma} \left[\left(\frac{\varphi_x}{\varphi_I} \right)^{\frac{1}{\sigma-1}} - 1 \right].$$

(b) Then use (6.12) and (12.10) to show that

$$\frac{s_x}{s_I} = \tau^{1-\sigma} \left[\left(\frac{f_I - f_x}{f_x} \frac{1}{\tau^{\sigma-1} - 1} \right)^{\frac{1}{\sigma-1}} - 1 \right].$$

EMPIRICAL EXERCISES

- 12.1 In this exercise, you are asked to reproduce the empirical results shown in table 2 of Chen and Feenstra (2008), which provides a replication of the Antràs (2003) regression, with a broader sample of years, from 1989 to 2004. To replicate it you should download the dataset “Reg_intrafirm.dta”. This data file

should be stored in the directory c:\Empirical_Exercise\Chapter_12\. After the data file is in this folder, download STATA program “Problem_12_1.do” to the same folder and run it to replicate table 2 from Feenstra and Chen (2008).

- a) Interpret the coefficient at capital-labor ratio in the last specification.
 - b) Do you expect coefficients at capital-labor ratio and human capital-labor ratio to have the same sign? Explain.
 - c) Explain why openness index is endogenous in the last specification. Explain the choice of instruments. What theoretical foundations do the instruments have?
- 12.2 In this exercise, you are asked to reproduce the empirical results shown in the benchmark regression (table IV) of Nunn (2007). To replicate it you should download the file “qje_contracts_final1.zip” from the following link: http://scholar.harvard.edu/files/nunn/files/qje_contracts_final1.zip and unpack it to the directory c:\Empirical_Exercise\Chapter_12\. After the data file is in this folder, download STATA program “Problem_12_2.do” to the same folder and run it to replicate table IV from Nunn (2007), which is based on the following specification

$$\ln(x_{ic}) = \alpha_i + \alpha_c + \beta_1 z_i Q_c + \beta_2 h_i H_c + \beta_3 k_i K_c + \varepsilon_{ic}$$

Then answer the following questions:

- a) Do the positive and significant coefficient estimates imply the impact of interaction terms on the volume of exports, or the pattern of trade? Justify the usage of the interaction of an industry characteristic with a country characteristic.
- b) Notice that table IV reports standardized beta coefficients: what is the advantage of this choice? What does $\beta_1 = 0.33$ imply economically?

APPENDIX A

Price, Productivity, and Terms of Trade Indexes

At various points throughout this book we have relied on index numbers to measure prices or total factor productivity (TFP). There is a good deal of theory, dating back more than a hundred years, that we can rely on to guide us in the choice of index number formulas. Like econometrics, an understanding of index numbers can be very helpful in empirical research. Unlike econometrics, the use of index numbers is seldom taught. So in this appendix we present a self-contained treatment of index numbers, focusing on price indexes.¹

We begin with some examples of price indexes. Let $p = (p_1, \dots, p_N) > 0$ denote a vector of prices, and $q = (q_1, \dots, q_N) \geq 0$ denote the corresponding vector of quantities. We let p^i and q^i denote these vectors in situation $i = 0, 1$. We can think of $i = 0, 1$ as measuring two time periods, or later in the appendix, as measuring two countries. A bilateral index number formula $P(p^0, p^1, q^0, q^1)$ uses just this information on prices and quantities to measure the “average” change in prices from situation 0 to 1. For example, the Laspeyres price index is

$$P_L(p^0, p^1, q^0, q^1) \equiv \frac{\sum_{i=1}^N q_i^0 p_i^1}{\sum_{i=1}^N q_i^0 p_i^0}. \quad (\text{A.1})$$

Thus, the Laspeyres weights the prices in each period by the *quantities in the base period*.

Conversely, if we use the *current period quantities* to weight the prices, we obtain the Paasche price index:

$$P_P(p^0, p^1, q^0, q^1) \equiv \frac{\sum_{i=1}^N q_i^1 p_i^1}{\sum_{i=1}^N q_i^1 p_i^0}. \quad (\text{A.2})$$

We might expect the Paasche price index to give a *lower* value than the Laspeyres index when applied to a consumer choice problem, because by using current period quantities, the Paasche index incorporates consumer substitution *away* from goods with the highest price increase. We will show below that this is indeed the case. A third commonly used formula is Irving Fisher’s (1922) ideal price index, P_F :

¹This appendix draws extensively from Alterman, Diewert, and Feenstra (1999, chapter 2).

$$P_F(p^0, p^1, q^0, q^1) \equiv [P_L(p^0, p^1, q^0, q^1) P_P(p^0, p^1, q^0, q^1)]^{1/2}. \quad (\text{A.3})$$

which is the geometric mean of the Laspeyres and Paasche indexes.

Another important formula is the Törnqvist (1936) price index P_T , defined as a geometric mean of the price ratios (p_i^1/p_i^0) , using as weights the average of the expenditure shares $s_i^0 \equiv p_i^0 q_i^0 / \sum_{i=1}^N p_i^0 q_i^0$ and $s_i^1 \equiv p_i^1 q_i^1 / \sum_{i=1}^N p_i^1 q_i^1$ in the two periods:

$$P_T(p^0, p^1, q^0, q^1) \equiv \prod_{i=1}^N (p_i^1/p_i^0)^{(s_i^0 + s_i^1)/2}, \quad (\text{A.4})$$

or in logs,

$$\ln P_T(p^0, p^1, q^0, q^1) = \sum_{i=1}^N (1/2)(s_i^0 + s_i^1) \ln(p_i^1/p_i^0). \quad (\text{A.4}')$$

How can we choose between these various formulas? The economic approach to index numbers is to ask how well each formula reflects the behaviour of consumers or firms. On the consumer side, we would need to take into account that tastes may be nonhomothetic (see Diewert 1974 and Pollack 1989). We will focus here on the problem for firms, where we consider either an index of the *input prices* faced by a firm or the *output prices*. In either case, we will also want to develop an index of *productivity change* for the firm, and with an open economy, an index of the *terms of trade*.

INPUT PRICE INDEXES

Let $p = (p_1, \dots, p_N)$ denote a positive vector of input prices that producers face in period t , and let y denote an M dimensional vector of nonnegative outputs that the production sector might be asked to produce in period t . Then the *cost function* using period t technology is defined as

$$C^t(p, y) \equiv \min_q \{p'q : (q, y) \text{ belongs to } S^t\}, \quad (\text{A.5})$$

where S^t denotes the period t technology set. Thus $C^t(p, y)$ is the minimum cost that the economy can achieve in period t , given that the vector of outputs y must be produced. Productivity improvements are reflected in an expansion of the technology set S^t , and this is why the cost function $C^t(p, y)$ is also indexed by the period $t = 0, 1$.

Generally, the goal of price and productivity indexes will be to measure changes in the value of costs C^t over time: costs increase as input prices rise, and decrease as productivity improves. The difficulty is that we do not observe the *separate* effect of each, but rather, just observe the total change in costs. We would like to decompose the total change in costs into components that can be attributed to input prices, productivity, and changes in output.

Toward this end, we can use the function C^t to define the economy's period t technology *theoretical input price index* P^t between any two periods, say period 0 and period 1, as follows:

$$P^t(p^0, p^1, y) = C^t(p^1, y) / C^t(p^0, y), \quad (\text{A.6})$$

where p^t is the vector of output prices that the economy faces in period t , $t = 0, 1$, and y is a reference vector of outputs. We refer to this as a "theoretical" index because it

cannot be computed from information on p^0, p^1, q^0 , and q^1 alone, but also depends on the function C^t . Note that this theoretical index has the advantage that it is not influenced by technological change; rather, it is measured relative to a given period t technology. Our immediate goal is to see whether, for specific functional forms for the cost function in (A.5), there will correspond some observable price index formula (depending only on prices and quantities p^0, p^1, q^0 , and q^1) that will equal (A.6), or at least provide upper and lower-bounds to it.

We are interested in two special cases of the theoretical input price index (A.6): (a) $P^0(p^0, p^1, y^0)$, which uses the period 0 technology and the output vector y^0 that was actually used in period 0, and (b) $P^1(p^0, p^1, y^1)$, which uses the period 1 technology and the output vector y^1 . Let q^0 and q^1 be the observed input quantities used in periods 0 and 1, respectively. We assume that these are *optimally chosen* to minimize costs in (A.5), so that

$$C^0(p^0, y) = p^{0'} q^0 \quad \text{and} \quad C^1(p^1, y^1) = p^{1'} q^1. \quad (\text{A.7})$$

Under these cost-minimizing assumptions, we can adapt the arguments of Fisher and Shell (1972, 57–58) and Archibald (1977, 66) to show that the two theoretical indexes, $P^0(p^0, p^1, y^0)$ and $P^1(p^0, p^1, y^1)$, satisfy the following inequalities:

$$\begin{aligned} P^0(p^0, p^1, y^0) &\equiv C^0(p^1, y^0)/C^0(p^0, y^0) \quad \text{using definition (A.6)} \\ &= C^0(p^1, y^0)/p^{0'} q^0 \quad \text{using (A.7)} \\ &\leq p^{1'} q^0/p^{0'} q^0 \quad \text{since } q^0 \text{ is feasible for the minimization problem} \\ &\quad \text{that defines } C^0(p^1, y^0), \text{ and so } C^0(p^1, y^0) \leq p^{1'} q^0 \\ &\equiv P_L(p^0, p^1, q^0, q^1), \end{aligned}$$

where P_L is the Laspeyres price index. Similarly, we have

$$\begin{aligned} P^1(p^0, p^1, y^1) &\equiv C^1(p^1, y^1)/C^1(p^0, y^1) \quad \text{using definition (A.6)} \\ &= p^{1'} q^1/C^1(p^0, y^1) \quad \text{using (A.7)} \\ &\geq p^{1'} q^1/p^{0'} q^1 \quad \text{since } q^1 \text{ is feasible for the minimization problem} \\ &\quad \text{that defines } C^1(p^0, y^1) \text{ and so } C^1(p^0, y^1) \leq p^{0'} q^1 \\ &\equiv P_P(p^0, p^1, q^0, q^1), \end{aligned}$$

where P_P is the Paasche price index. So we find that the Laspeyres index of input prices P_L is an *upper bound* to the theoretical price index $P^0(p^0, p^1, y^0)$, and that the Paasche index of input prices P_P is a *lower bound* to the theoretical price index $P^1(p^0, p^1, y^1)$. To further explore these bounds, we separately discuss the cases of constant and changing technology.

Constant Technology

If technology is constant, then the Paasche and Laspeyres price indexes provide bounds to the “true” price index. That is, if technology is constant, the cost function can be written as $C(p, y)$, so the bounds above immediately imply that

$$P_P(p^0, p^1, q^0, q^1) \leq C(p^1, y^1)/C(p^0, y^1), \text{ and } C(p^1, y^0)/C(p^0, y^0) \leq P_L(p^0, p^1, q^0, q^1). \quad (\text{A.8})$$

These expressions can be further simplified by assuming that there is a single output, and that the technology has constant returns to scale, so that $C(p, y) = yc(p)$, where $c(p)$ is the unit-cost function. Then the inequalities in (A.8) simplify to

$$P_p(p^0, p^1, q^0, q^1) \leq c(p^1)/c(p^0) \leq P_L(p^0, p^1, q^0, q^1). \quad (\text{A.9})$$

Thus, we see that the Paasche and Laspeyres price indexes provide lower and upper bounds to the ratio of unit-costs for the firm, which is the “true” price index. The same would apply to a consumer, where we would use the ratio of unit-expenditure functions in the two periods rather than the ratio of unit-cost. This confirms our assertion above that the Paasche price index is less than the Laspeyres price index for a consumer choice problem, provided that the quantities used are optimal for the consumer.

These observations lead us to ask whether it is possible to construct a price index that *exactly equals* the ratio of unit-costs, and that is the goal of the economic approach to index numbers. The strategy is to choose some specific functional form for the cost function, and then see whether there is a price index formula such that $P(p^0, p^1, q^0, q^1) = c(p^1)/c(p^0)$. For certain choices of the functional form for costs, it turns out that there is indeed a price index formula that will satisfy this equality. In this case, Diewert (1976) calls the price index *exact*.

For example, suppose that we choose the quadratic production function $y = (q' A q)^{1/2}$, where $A = A'$ is a symmetric matrix of parameters. It is clear that this production function is homogeneous of degree one. Diewert (1976, 130) shows that this quadratic production function can approximate any linearly homogeneous function $f(q)$ to the second order around a point.² For this reason, Diewert calls the quadratic function $(q' A q)^{1/2}$ a “flexible” functional form. If A is invertible,³ then it is not difficult to show that the cost function is $C(p, y) = y(p' B p)^{1/2}$, where $B = A^{-1}$, so the unit-cost function becomes $c(p) = (p' B p)^{1/2}$, which is again a “flexible” functional form.

Then corresponding to the quadratic production and unit-cost functions, we have the following theorem.

THEOREM 1 (DIEWERT 1976)

If the production and unit-cost functions are quadratic and linearly homogeneous, and the observed quantities $q^t > 0$ are the cost-minimizing, then

$$\frac{c(p^1)}{c(p^0)} = P_F(p^0, p^1, q^0, q^1). \quad (\text{A.10})$$

The price index on the right of (A.10) is the Fisher ideal formula, defined by (A.3), which is therefore exact for the quadratic unit-cost function. Since the quadratic function is itself flexible, then Diewert (1976) refers to the Fisher ideal price index as *superlative*. The remarkable feature of this theorem is that we achieve equality between the Fisher ideal index and the ratio of unit-cost function for all values of the parameters

²That is, the parameters A can be chosen such that $(q'' A q^*)^{1/2} = f(q^*)$, $Aq^* = f_q(q^*)$, and $A = f_{qq}(q^*)$, for any choice of a linearly homogeneous function $f(q)$.

³The solution for the unit-expenditure function when A is not invertible is discussed in (Diewert 1976, 134).

A and *B*. Intuitively, as these parameters change, so would the optimal choices q , and so these choices are “revealing” what the underlying parameters must be. Using the data on the optimal choices q^t , combined with their prices p^t , we are able to measure the ratio of the unit-cost functions in (A.10) without knowing *B*. This result shows the power of the economic approach to index numbers.

Changing Technology

With technology changing over time, the situation is more difficult. We are in the awkward position that the Paasche and Laspeyres indexes provide bounds to a theoretical price index $P^t(p^0, p^1, y)$ that itself depends on the period $t = 0, 1$. Fortunately, this issue can be resolved, and it is possible to define a theoretical price index that falls *between* the observable Paasche and Laspeyres price indexes. To do this, first we define a hypothetical cost function, $C(p, \alpha)$, that corresponds to the use of an α -weighted average of the technology sets S^0 and S^1 for periods 0 and 1 as the reference technology, and that uses an α -weighted average of the period 0 and period 1 net output vectors y^0 and y^1 :

$$C(p, \alpha) \equiv \min_q \{p'q : [q, (1-\alpha)y^0 + \alpha y^1] \text{ belongs to } (1-\alpha)S^0 + \alpha S^1\}. \quad (\text{A.11})$$

We can now use this cost function to define the following family of theoretical price indexes:

$$P(p^0, p^1, \alpha) \equiv C(p^1, \alpha)/C(p^0, \alpha). \quad (\text{A.12})$$

We can adapt the proof of Diewert (1983, 1060–61) and show that there exists an α such that the theoretical price index defined by (A.12) lies between the Paasche and Laspeyres output indexes, P_p and P_L :⁴

THEOREM 2 (DIEWERT 1983)

There exists α between 0 and 1 such that:

$$P_L \equiv p^1 q^0 / p^0 q^0 \leq P(p^0, p^1, \alpha) \leq p^1 q^1 / p^0 q^1 \equiv P_p \quad \text{or} \quad P_p \leq P(p^0, p^1, \alpha) \leq P_L. \quad (\text{A.13})$$

If the Paasche and Laspeyres indexes are numerically close to each other, then (A.13) tells us that a “true” economic input price index is fairly well determined and we can find a reasonably close approximation to the “true” index by taking an average of P_L and P_p , such as Fisher’s ideal price index P_F defined earlier by (A.3). This provides a justification for using the Fisher ideal formula even when the technology is changing.

Further results can be obtained by making more specific assumptions on the nature of technological change. In particular, suppose that the cost function takes on the *translog functional form* for $t = 0, 1$:⁵

⁴The proofs for all theorems can be found in the sources cited.

⁵This functional form was introduced and named by Christensen, Jorgenson, and Lau (1971). It was adapted to the expenditure or cost function context by Diewert (1974).

$$\begin{aligned} \ln C^t(p, y) = & \alpha_0^t + \sum_{i=1}^N \alpha_i^t \ln p_i + \sum_{k=1}^M \beta_k^t \ln y_k + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij}^t \ln p_i \ln p_j \\ & + \frac{1}{2} \sum_{k=1}^M \sum_{\ell=1}^M \delta_{k\ell}^t \ln y_k \ln y_\ell + \sum_{i=1}^N \sum_{k=1}^M \varphi_{ik}^t \ln p_i \ln y_k, \end{aligned} \quad (\text{A.14})$$

where the parameters satisfy the restrictions $\gamma_{ij}^t = \gamma_{ji}^t$ and where

$$\sum_{i=1}^N \alpha_i^t = 1 \quad \text{and} \quad \sum_{i=1}^N \gamma_{ij}^t = \sum_{i=1}^N \varphi_{ik}^t = 0. \quad (\text{A.15})$$

These restrictions ensure that $C^t(p, y)$ is linearly homogeneous in the input price vector p (which is a property that any cost function must satisfy). Note that we are allowing the parameters that characterize the technology (the α 's, β 's, γ 's, δ 's, and φ 's) to be completely different in each period, reflecting technological change.

Now let us add the restriction that the quadratic price parameters in (A.14) are *equal* across the two periods, that is, $\gamma_{ij}^0 = \gamma_{ij}^1$, though we still allow the other parameters to change over time. Then we can adapt the result in Caves, Christensen, and Diewert (1982, 1410) to show that the geometric mean of the theoretical price indexes in periods 0 and 1, $P^0(p^0, p^1, y^0)$ and $P^1(p^0, p^1, y^1)$, is equal to the Törnqvist price index P_T defined by (A.4):

THEOREM 3 (CAVES, CHRISTENSEN, AND DIEWERT 1982)

Assume that $\gamma_{ij}^0 = \gamma_{ij}^1$, and that the cost-minimizing input quantities q^t are strictly positive in both periods, $t = 0, 1$. Then

$$[P^0(p^0, p^1, y^0) P^1(p^0, p^1, y^1)]^{1/2} = P_T(p^0, p^1, q^0, q^1). \quad (\text{A.16})$$

The assumptions required for this result seem rather weak; in particular, there is no requirement that the technologies exhibit constant returns to scale in either period and our assumptions are consistent with technological progress occurring between the two periods being compared. Because the index number formula P_T is *exactly* equal to the geometric mean of two theoretical price indexes, which corresponds to the translog, and is itself a “flexible” functional form, this confirms that Törnqvist index number formula is superlative.

To see the usefulness of this result, let us decompose the total change in costs between period 0 and period 1 in two alternative ways:

$$\frac{C^0(p^0, y^0)}{C^1(p^1, y^1)} = \frac{C^0(p^0, y^0)}{C^0(p^1, y^0)} \frac{C^0(p^1, y^0)}{C^0(p^1, y^1)} \frac{C^0(p^1, y^1)}{C^1(p^1, y^1)}, \quad (\text{A.17})$$

and,

$$\frac{C^0(p^0, y^0)}{C^1(p^1, y^1)} = \frac{C^1(p^0, y^1)}{C^1(p^1, y^1)} \frac{C^1(p^0, y^0)}{C^1(p^0, y^1)} \frac{C^0(p^0, y^0)}{C^1(p^0, y^0)}. \quad (\text{A.18})$$

In each of these decompositions, the first ratio on the right-hand side reflects the changing input prices; the second ratio reflects the changing output vector; and the

third ratio reflects pure technological change. Let us define the measure of technological change, or *total factor productivity* (TFP), as a geometric mean of the final ratios in (A.17) and (A.18):

$$\text{TFP} \equiv \left[\frac{C^0(p^1, y^1)}{C^1(p^1, y^1)} \frac{C^0(p^0, y^0)}{C^1(p^0, y^0)} \right]^{1/2}. \quad (\text{A.19})$$

This index reflects the fall in costs from period 0 to period 1 due to pure technological change, but as defined, it is not observable: neither the period 0 costs with period 1 prices and outputs, nor the period 1 costs with period 0 prices and outputs, are known. However, we can make use of theorem 3 to indirectly measure TFP. In particular, taking the geometric mean of (A.17) and (A.18), and using (A.19), we obtain

$$\frac{C^0(p^0, y^0)}{C^1(p^1, y^1)} = P_T(p^0, p^1, q^0, q^1)^{-1} \times \text{TFP} \times \left[\frac{C^1(p^1, y^0)}{C^1(p^1, y^1)} \frac{C^0(p^0, y^0)}{C^0(p^0, y^1)} \right]^{1/2}. \quad (\text{A.20})$$

Thus, the total change in costs from period 0 to period 1 is decomposed into the inverse of the Törnqvist input price index, TFP, and the final term in (A.20), which reflects the change in outputs. We will see some techniques later in the appendix that allow us to measure the final term, but note here a particularly simple result.

If there is a single output, so that y^t is a scalar, and the cost function is linearly homogeneous in this output, then it can be written as $C^t(p, y) = yc^t(p)$ where $c^t(p)$ is the unit-cost function. In that case, the final bracketed term on the right-hand side of (A.20) is simply y^0/y^1 . This ratio of outputs also appears on the left-hand side of (A.20), which equals $(y^0/y^1)[c^0(p^0)/c^1(p^1)]$, so it cancels from both sides. Then we can rearrange the terms in (A.20) to obtain total factor productivity as

$$\ln(\text{TFP}) = \ln P_T(p^0, p^1, q^0, q^1) - \ln [c^1(p^1)/c^0(p^0)]. \quad (\text{A.21})$$

If there is perfect competition in the industry, unit-costs equal price, so the ratio of unit-costs on the right of (A.21) is replaced by the ratio of output prices in the two periods. Then (A.21) becomes the *dual Törnqvist measure of total factor productivity* that we used in chapter 4, which is a weighted average of the change in input prices minus the change in output prices. The above results provide a rigorous justification for the use of this index (or its primal counterpart, the change in log output minus the log Törnqvist index of input quantities), which is widely thought to be the “best” measure of productivity change.

GDP PRICE INDEX

Our results above can be readily extended to the *GDP price index* for an economy. We suppose that the feasible set of outputs $q = (q_1, \dots, q_N)$ and factor endowments $v = (v_1, \dots, v_M) \geq 0$ during period t is a set S^t , a subset of $N + M$ dimensional space. We make the following conventions on outputs: if commodity i in the list of net outputs is an output in period t , then its corresponding quantity q_i^t is positive; however, if commodity i is an input in period t , then its corresponding quantity q_i^t is negative. The latter would include imported intermediate inputs, for example. Then the sum of price times quantity over all net outputs, $\sum_{i=1}^N p_i^t q_i^t$, equals net revenue or GDP for the economy.

Let $p = (p_1, \dots, p_N)$ denote a positive vector of net output prices that producers face in period t . Then the *revenue or GDP function* using period t technology is defined as

$$G^t(p, v) \equiv \max_q \{p'q : (q, v) \text{ belongs to } S^t\}. \quad (\text{A.22})$$

Thus $G^t(p, v)$ is the maximum value of (net) output that the economy can produce, given that the vector of net inputs v is available for use, using the period t technology.

The GDP function can be used to define the economy's period t technology *theoretical GDP price index* P^t between any two periods, say period 0 and period 1, as follows:

$$P^t(p^0, p^1, v) = G^t(p^1, v) / G^t(p^0, v), \quad (\text{A.23})$$

where p^t is the vector of net output prices that the economy faces in period t , $t = 0, 1$, and v is a reference vector of factor endowments.

Note that there are a wide variety of price indexes of the form (A.23) depending on which (t, v) reference technology and input vector v that we choose. Again, we are interested in two special cases: (a) $P^0(p^0, p^1, v^0)$, which uses the period 0 technology and the input vector v^0 that was actually used in period 0, and (b) $P^1(p^0, p^1, v^1)$, which uses the period 1 technology and the input vector v^1 . Let q^0 and q^1 be the observed net output vectors in periods 0 and 1 respectively. If there is competitive behaviour on the part of producers in periods 0 and 1, then observed GDP in periods 0 and 1 should be equal to $G^0(p^0, v^0)$ and $G^1(p^1, v^1)$ respectively; that is,

$$G^0(p^0, v^0) = p^0' q^0 \quad \text{and} \quad G^1(p^1, v^1) = p^1' q^1. \quad (\text{A.24})$$

Then we can use these results of Fisher and Shell (1972, 57–58) and Archibald (1977, 66) to again develop a series of inequalities on the theoretical price indexes:

$$\begin{aligned} P^0(p^0, p^1, v^0) &\equiv G^0(p^1, v^0) / G^0(p^0, v^0) \quad \text{using definition (A.23)} \\ &= G^0(p^1, v^0) / p^0' q^0 \quad \text{using (A.24)} \\ &\geq p^1' q^0 / p^0' q^0 \quad \text{since } q^0 \text{ is feasible for the maximization problem} \\ &\quad \text{that defines } G^0(p^1, v^0), \text{ and so } G^0(p^1, v^0) \geq p^1' q^0 \\ &\equiv P_L(p^0, p^1, q^0, q^1), \end{aligned}$$

where P_L is the Laspeyres price index. Similarly, we have

$$\begin{aligned} P^1(p^0, p^1, v^1) &\equiv G^1(p^1, v^1) / G^1(p^0, v^1) \quad \text{using definition (A.23)} \\ &= p^1' q^1 / G^1(p^0, v^1) \quad \text{using (A.24)} \\ &\leq p^1' q^1 / p^0' q^1 \quad \text{since } q^1 \text{ is feasible for the maximization problem} \\ &\quad \text{that defines } G^1(p^0, v^1) \text{ and so } G^1(p^0, v^1) \geq p^0' q^1 \\ &\equiv P_P(p^0, p^1, q^0, q^1), \end{aligned}$$

where P_P is the Paasche price index. Note that these inequalities are in the opposite direction compared to that of their counterparts for the input index, which reflects the fact that revenue is maximized in (A.22), whereas costs are minimized in (A.5).

As with the input index, it is possible to define a theoretical GDP price index that falls *between* the observable Paasche and Laspeyres price indexes. To do this, we

first define a hypothetical GDP function, $G(p, \alpha)$, that corresponds to the use of an α -weighted average of the technology sets S^0 and S^1 for periods 0 and 1 as the reference technology, and that uses an α - weighted average of the period 0 and period 1 input vectors v^0 and v^1 :

$$G(p, \alpha) \equiv \max_q \{p'q : [q, (1 - \alpha)v^0 + \alpha v^1] \text{ belongs to } (1 - \alpha)S^0 + \alpha S^1\}. \quad (\text{A.25})$$

Then we can define the following family (indexed by α) of theoretical GDP price indexes:

$$P(p^0, p^1, \alpha) \equiv G(p^1, \alpha) / G(p^0, \alpha). \quad (\text{A.26})$$

It turns out that for the theoretical index in (A.26), theorem 2 *continues to hold as stated*. That is, there exists α between 0 and 1 such that the theoretical GDP index in (A.26) lies between the observable Paasche and Laspeyres indexes. There is an important qualification to this result, however: as stated in (A.13), we do not know in general whether $P_L \geq P_p$ or $P_L \leq P_p$. For the input index, cost-minimizing behavior *with a constant technology* will cause the Laspeyres price index to exceed the Paasche price index, since the former uses base-period quantities that do not reflect the substitution by firms *away* from the inputs whose prices increase the most. Conversely, for the GDP price index, firms will substitute *toward* the goods with the greatest increase in price, and *with a constant technology* this will lead the Paasche price index to exceed the Laspeyres price index. In practice, however, even for GDP price indexes we typically find that the *Laspeyres price index exceeds the Paasche price index*. The same holds for quantity indexes: the Laspeyres quantity index typically exceeds the Paasche quantity index.⁶ This empirical regularity is called the “Gerschenkron effect,” after the work of Alexander Gerschenkron (1951).⁷

The case where the Gerschenkron effect *does not* hold is illustrated in figure A.1, where we hold technology constant along a given production possibilities frontier (PPF). We choose good 2 as the numeraire and evaluate GDP by the intercepts on the vertical axis. The initial prices lead to outputs at point A , with GDP measured by $0A$ of good 2 along the vertical axis. At the new prices outputs occur at point B , with GDP is measured by $0B$ of good 2. So the ratio $0B/0A$ is the theoretical GDP price index. The Laspeyres price index would evaluate the price change at the initial quantities to obtain the ratio $0A'/0A < 0B/0A$. On the other hand, the Paasche price index would evaluate the price change at the new quantities, to obtain $0B/0B' > 0B/0A > 0A'/0A$, which exceeds that obtained from the Laspeyres price index. This confirms that the Paasche price index exceeds the Laspeyres index when technology does not change, due to output substitution by producers.

In figure A.2, by contrast, we suppose that the PPF is shifting out due to technological change. At unchanged prices, GDP would increase from point A to point B (i.e., GDP would increase from the intercept of AA' and BB' with the vertical axis). Suppose, however, that the relative price of good 1 falls due to its increased supply, so that the

⁶Using the formula (A.1) and (A.2), the Laspeyres and Paasche quantity indexes are defined as $Q_L(p^0, p^1, q^0, q^1) \equiv P_L(q^0, q^1, p^0, p^1)$ and $Q_p(p^0, p^1, q^0, q^1) \equiv P_p(q^0, q^1, p^0, p^1)$, that is, just reversing the role of prices and quantities in the price indexes. It is readily confirmed that $Q_L(p^0, p^1, q^0, q^1) P_p(p^0, p^1, q^0, q^1) = Q_p(p^0, p^1, q^0, q^1) P_L(p^0, p^1, q^0, q^1) = p^1' q^1 / p^0' q^0$, which is the ratio of total spending on the goods in the two periods. Then having $P_L \geq P_p$ for the price indexes immediately implies $Q_L \geq Q_p$ for the quantity indexes as well.

⁷Recent examples of the Gerschenkron effect can be found in Van Ark, Monnikhof, and Timmer (1999) for various OECD and transition economies, Dikhanov (1999) for the Soviet Union, and Alterman, Diewert, and Feenstra (1999) for the U.S. export price index.

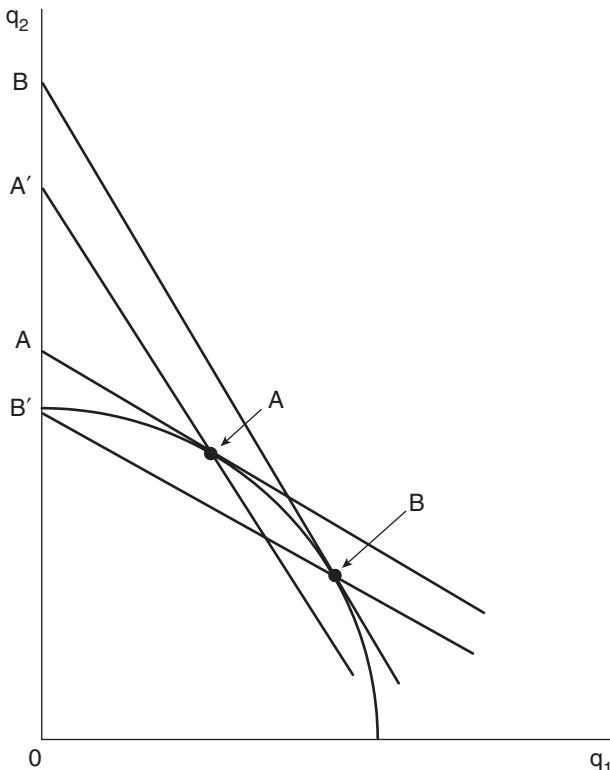


Figure A.1

slope of the line CC shows the new prices. Production now occurs at point B' rather than point B . We have drawn the case where the budget lines AA and CC are both tangent to an indifference curve U , indicating that the change in equilibrium prices are due to changing supply with constant utility.

In the case we have illustrated, the production points A and B' lie on the same ray from the origin, so that the *relative outputs* of the two goods are unchanged. This means that the Laspeyres and Paasche price indexes, using the outputs at points A or B' , would be *identical*. In fact, evaluating the Laspeyres price index using the initial outputs at point A , we would obtain GDP of $0A$ along the vertical axis at the old prices and $0C'$ at the new prices, so the price index is $0C'/0A < 1$, indicating a substantial fall in the price of good 1. The Paasche price index, evaluated at the outputs at point B' , would give exactly the same result. If point B' occurred along the PPF *slightly to the right* of the ray through the origin and A , however, then the higher supply of good 1, whose relative price has fallen, would cause the Paasche price index to be less than the Laspeyres price index. This would illustrate the Gerschenkron effect.

Finding that the Laspeyres price index exceeds the Paasche is indirect evidence of changing technology (or factor endowments). The theoretical GDP price index $P(p^0, p^1, \alpha)$, defined in (A.26), holds technology and factor endowments constant, and is therefore a suitable measure of price change. Theorem 2 says that this theoretical index can be measured by an average of the Paasche and Laspeyres price indexes, which is

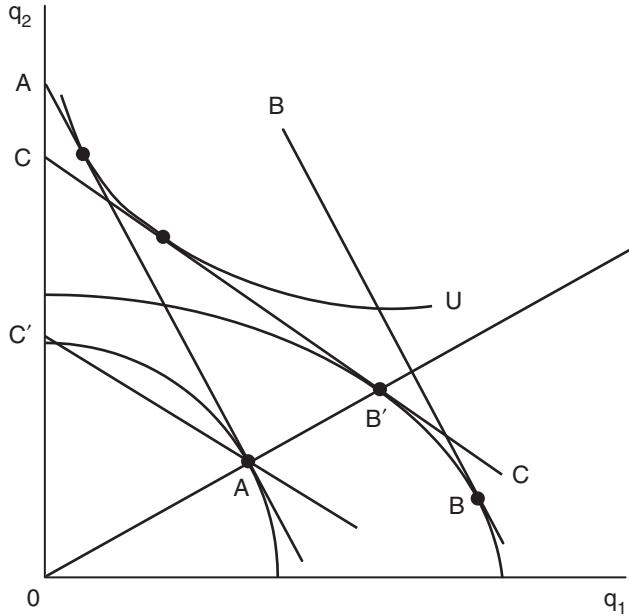


Figure A.2

one justification for using the Fisher ideal index for the output price index, as we also found for the input price index.

Further results can be obtained by using the translog form for the GDP function, where we allow its parameters to vary over time due to technological change:

$$\begin{aligned} \ln G^t(p, v) = & \alpha_0^t + \sum_{i=1}^N \alpha_i^t \ln p_i + \sum_{k=1}^M \beta_k^t \ln v_k + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij}^t \ln p_i \ln p_j \\ & + \frac{1}{2} \sum_{k=1}^M \sum_{\ell=1}^M \delta_{k\ell}^t \ln v_k \ln v_\ell + \sum_{i=1}^N \sum_{k=1}^M \varphi_{ik}^t \ln p_i \ln v_k. \end{aligned} \quad (\text{A.27})$$

In order to ensure that the translog GDP function is homogeneous of degree one in prices, we impose symmetry $\gamma_{ij}^t = \gamma_{ji}^t$ and the requirements

$$\sum_{i=1}^N \alpha_i^t = 1 \quad \text{and} \quad \sum_{i=1}^N \gamma_{ij}^t = \sum_{i=1}^N \varphi_{ik}^t = 0. \quad (\text{A.28})$$

In addition, to ensure that the GDP function is homogeneous of degree one in endowments, we impose symmetry $\delta_{k\ell}^t = \delta_{\ell k}^t$ and the requirements

$$\sum_{k=1}^M \beta_k^t = 1 \quad \text{and} \quad \sum_{k=1}^M \delta_{k\ell}^t = \sum_{k=1}^M \varphi_{ik}^t = 0. \quad (\text{A.29})$$

The next result is completely analogous to theorem 3 for the translog cost function, where again we require that the second-order parameters on prices γ_{ij}^t remain constant. This result follows by adapting the arguments in Caves, Christensen, and

Diewert (1982). It does not require the GDP function to be homogeneous of degree one in factor endowments, though we will make use of (A.29) later.

THEOREM 4 (CAVES, CHRISTENSEN, AND DIEWERT 1982)

Assuming that $\gamma_{ij}^0 = \gamma_{ij}^1$, and that the output quantities q^t are revenue-maximizing at prices $p^t, t = 0, 1$, then

$$[P^0(p^0, p^1, v^0) P^1(p^0, p^1, v^1)]^{1/2} = P_T(p^0, p^1, q^0, q^1). \quad (\text{A.30})$$

Thus, a geometric mean of the theoretical GDP price indexes $P^0(p^0, p^1, v^0)$ and $P^1(p^0, p^1, v^1)$ equals the translog price index defined by (A.4). Unlike with theorem 3, we no longer need to assume that the quantities q^t are strictly positive: we are now letting $q_i^t > 0$ denote an output, and $q_i^t < 0$ denote an input, and a commodity can switch from one to the other across the two periods. This raises an important question of interpretation, however. Since $q_i^t < 0$ can reflect imported intermediate inputs, while $q_i^t > 0$ can reflect exported commodities, it appears that the output price index is mixing together terms of trade effects with other price movements. This is indeed the case, and in the next section we use the results of Diewert and Morrison (1986) and Kohli (1990b, 2004) to separately measure the terms of trade impact and productivity.

TERMS OF TRADE AND PRODUCTIVITY

Let us return to figure A.2, where we illustrated a productivity improvement and a fall in the relative price of goods 1. We have already argued that the Paasche and Laspeyres output price indexes would be equal in the example illustrated, and would indicate a very substantial decline in prices, of $0C'/0A < 1$. Using good 2 as the numeraire, the resulting value of GDP (measured off the vertical axis) indicates a much smaller decline, from $0A$ to $0C$, with the ratio $0C/0A$. If we divide this change in GDP by the price index, we obtain a measure of the change in *real* GDP, which is just $(0C/0A)/(0C'/0A) = 0C/0C' > 1$. Thus, we find that real GDP has risen, which is entirely appropriate since production has indeed increased from point A to point B' . Indeed, the ratio $0C/0C'$ along the vertical axis is identical to the ratio $0B'/0A$ along the ray from the origin, which precisely measures the increase in outputs and real GDP. We conclude that our GDP price index, and the corresponding measure of real GDP, accurately reflects the outward shift in the PPF due to technological change or factor accumulation.

Now suppose we pose a different question, and ask what has happened to the cost of living for the representative consumer, with indifference curve shown by U in figure A.2. An exact cost of living index for the consumer would compare the initial expenditure $0A$ in figure A.2 with the new expenditure $0C$, both measured with good 2 as the numeraire and giving constant utility of U . Thus, an exact cost of living index would equal $0C/0A < 1$. Notice that this index is constructed using exactly the same *prices* as the GDP price index, but now using different *quantities* as weights: the cost of living index uses the quantities demanded, whereas the GDP price index uses the quantities supplied. For this reason we can expect the indexes to differ, and in our example, the cost of living index $0C/0A$ exceeds the GDP price index, which was $0C'/0A$. Indeed,

the difference between the two indexes is entirely due to the *terms of trade change* for the economy. The fall in the relative price of good 1 is reflected *more* in the GDP price index than in the cost of living index, because this export good has production exceeding consumption.

The above argument suggests the following definition of the *terms of trade index*:

$$(\text{Terms of trade index}) \equiv (\text{GDP price index}) / (\text{Cost of Living index}), \quad (\text{A.31})$$

where all of these indexes are computed over the set of final goods sold in an economy. To further motivate this definition, consider the Törnqvist index in (A4) or (A4'). Let $P_T(p^0, p^1, q^0, q^1)$ denote the cost of living index, computed with prices p^t and quantities consumed q^t , $t = 0, 1$. In addition, let $P_T(p^0, p^1, y^0, y^1)$ denote the GDP price index, computed with prices p^t and quantities produced y^t , $t = 0, 1$. Net exports of final goods are simply $x^t = y^t - q^t$, and let us assume for simplicity that trade is balanced, so that $p^t x^t = p^t y^t - p^t q^t = 0$. Then it follows immediately from (A4') that

$$\ln P_T(p^0, p^1, y^0, y^1) - \ln P_T(p^0, p^1, q^0, q^1) = \sum_{i=1}^N \left(\frac{1}{2} \right) \left(\frac{p_i^0 x_i^0}{p_i^0 y^0} + \frac{p_i^1 x_i^1}{p_i^1 y^1} \right) \ln(p_i^1/p_i^0). \quad (\text{A.32})$$

The left-hand side of (A.32) is the difference between the log GDP price index and the log cost of living index, while the right-hand side is a price index constructed using as weights the average value of exports relative to GDP in the two periods. These weights are negative for imports, so the right-hand side of (A.32) is indeed interpreted as a terms of trade index.

These results can be extended to the case where imports include intermediate inputs, as in Diewert and Morrison (1986). To achieve this extension, let us identify three groups of commodities: those for final domestic demand (quantity $q_{di}^t > 0$ and price $p_{di}^t > 0$, for $i = 1, \dots, N_d$); those for export ($q_{xi}^t > 0$ and $p_{xi}^t > 0$, for $i = 1, \dots, N_x$); and imported intermediate inputs ($q_{mi}^t < 0$ and $p_{mi}^t > 0$, for $i = 1, \dots, N_m$). The quantity vector for all commodities is $q^t = (q_{d*}^t, q_x^t, q_m^t)$, with prices $p^t = (p_{d*}^t, p_x^t, p_m^t)$. Then the GDP function for the economy is still defined as in (A.22), where q^t is the revenue-maximizing choices by firms. We will suppose that GDP takes on the translog form as in (A.27)–(A.29).

For this GDP function, let us define the *theoretical productivity index*

$$R(p, v) = G^1(p, v) / G^0(p, v). \quad (\text{A.33})$$

This theoretical index holds prices and factor endowments fixed, and considers only the change in productivity between periods 0 and 1. We are interested in two special cases: (a) the Laspeyres productivity index, $R_L \equiv R(p^0, v^0) = G^1(p^0, v^0) / G^0(p^0, v^0)$, which uses prices and endowments from period 0; (b) the Paasche productivity index, $R_p \equiv R(p^1, v^1) = G^1(p^1, v^1) / G^0(p^1, v^1)$, which uses prices and endowments from period 1. Each of these are unobservable, but as in our earlier discussion of productivity in (A.19)–(A.21), we can measure their geometric mean:

THEOREM 5 (DIEWERT AND MORRISON 1986)

Assuming that the GDP function takes on the translog form in (A.27)–(A.29), and the revenue-maximizing quantities q^t are nonzero, then

$$(R_L R_p)^{1/2} = \left(\frac{p^{1'} q^1}{p^0 q^0} \right) / [P_T(p^0, p^1, q^0, q^1) Q_T(w^0, w^1, v^0, v^1)] \quad (\text{A.34})$$

The first term on the right-hand side of (A.34) is the ratio of GDP in the two periods, or the growth in nominal GDP. This is deflated by the two other terms on the right side: the first is the Törnqvist price index of GDP using the formula in (A.4); and the second is a Törnqvist quantity index of *factor endowments*, defined by

$$\ln Q_T(w^0, w^1, v^0, v^1) \equiv \sum_{i=1}^N \left(\frac{1}{2} \right) \left(\frac{w_i^0 v_i^0}{w^0 v^0} + \frac{w_i^1 v_i^1}{w^1 v^1} \right) \ln(v_i^1/v_i^0). \quad (\text{A.35})$$

Thus, this is a factor share-weighted average of the growth in primary inputs for the economy. Notice that under our assumption (A.29) that the GDP function is linearly homogeneous in v , $p^{t'} q^t = w^{t'} v^t$, so the shares in the Törnqvist price index and the Törnqvist quantity index are both measured relative to GDP. Theorem 5 generalizes our earlier derivation of TFP in (A.19)–(A.21), to allow now for multiple inputs and outputs.

Let us now take the Törnqvist price index of GDP that appears on the right side of (A.34), and decompose it into the portion dealing with domestic goods, exports, and imports:

$$\begin{aligned} \ln P_T(p^0, p^1, q^0, q^1) &= \sum_{i=1}^{N_d} \left(\frac{1}{2} \right) \left(\frac{p_{di}^0 q_{di}^0}{p^0 q^0} + \frac{p_{di}^1 q_{di}^1}{p^1 q^1} \right) \ln(p_{di}^1/p_{di}^0) \\ &\quad + \sum_{i=1}^{N_x} \left(\frac{1}{2} \right) \left(\frac{p_{xi}^0 q_{xi}^0}{p^0 q^0} + \frac{p_{xi}^1 q_{xi}^1}{p^1 q^1} \right) \ln(p_{xi}^1/p_{xi}^0) + \sum_{i=1}^{N_m} \left(\frac{1}{2} \right) \left(\frac{p_{mi}^0 q_{mi}^0}{p^0 q^0} + \frac{p_{mi}^1 q_{mi}^1}{p^1 q^1} \right) \ln(p_{mi}^1/p_{mi}^0). \end{aligned} \quad (\text{A.36})$$

We will refer to the first summation on the right-hand side of (A.36) as the cost of living index, because it uses the quantities and prices of final goods demanded by the consumer. The exponent of this first summation is denoted by the Törnqvist price index $P_T(p_d^0, p_d^1, q_d^0, q_d^1)$, where the shares are computed relative to total GDP $p^{t'} q^t$. The second and third terms are the difference between the export and import price indexes, respectively, (recall that $q_{mi}^t < 0$). Let us denote the exponent of these two summations by the Törnqvist index $P_T(p_{xm}^0, p_{xm}^1, q_{xm}^0, q_{xm}^1)$, where again, the shares are computed relative to total GDP $p^{t'} q^t$.

Then we can rewrite the result in (A.34) in the following form:

$$\left(\frac{p^{1'} q^1}{p^0 q^0} \right) = (R_L R_p)^{1/2} P_T(p_d^0, p_d^1, q_d^0, q_d^1) P_T(p_{xm}^0, p_{xm}^1, q_{xm}^0, q_{xm}^1) Q_T(w^0, w^1, v^0, v^1). \quad (\text{A.37})$$

Thus, the ratio of nominal GDP is decomposed as a total factor productivity index, a cost of living price index, a terms of trade index, and a quantity index of endowment growth. Kohli (2004) argues that if we wish to deflate the growth in GDP to obtain a measure of welfare change for the representative consumer, then we should use the cost of living index rather than the GDP price index as a deflator. That is, a measure of welfare change would be

$$\left(\frac{p^{1'} q^1}{p^0 q^0} \right) / P_T(p_d^0, p_d^1, q_d^0, q_d^1) = (R_L R_p)^{1/2} P_T(p_{xm}^0, p_{xm}^1, q_{xm}^0, q_{xm}^1) Q_T(w^0, w^1, v^0, v^1). \quad (\text{A.38})$$

This decomposition allows productivity growth, improvements in the terms of trade, and endowment growth in factors to all contribute positively to welfare. Notice that this measure of welfare is quite different from the normal method of calculating growth in real GDP, which is to deflate the ratio of nominal GDP by the GDP price index, obtaining

$$\left(\frac{p^1 q^1}{p^0 q^0} \right) / P_T(p^0, p^1, q^0, q^1) = (R_L R_p)^{1/2} Q_T(w^0, w^1, v^0, v^1). \quad (\text{A.39})$$

The log of (A.39) is reported as real GDP growth in national income statistics, but it *leaves out* changes in the terms of trade as a source of welfare gain. In (A.38), by contrast, we deflate nominal GDP by the price index computed for domestic goods only, that is, by the price index for domestic absorption, $C + I + G$. By applying this cost of living index to all components of GDP, $C + I + G + X - M$, including exports and imports, we obtain a measure of “real welfare” that is more appropriate for assessing changes in the representative consumer. Kohli (1990b, 2004) shows that deflating nominal GDP by the cost of living index leads to a considerably higher growth in “real welfare” for some countries (such as Switzerland) than real GDP growth. This is explained by the terms of trade improvements for such countries in recent years.

MEASUREMENT OF REAL GDP ACROSS COUNTRIES

Our discussion in this appendix has dealt with the measurement of price indexes *over time*, including the GDP price index and terms of trade price index. It is also important, however, to consider the measurement of real GDP *across countries*. In that case, we can make a distinction between real GDP on the output side and real GDP on the expenditure side. This distinction can be illustrated by the simple diagram in figure A.2. Rather than supposing that the production possibilities frontier (PPF) shifts out due to technological change, we instead suppose that the two PPF's shown correspond to two different countries. At initial prices, equal in the two countries, production in the country *A* occurs at point *A* and production in country *B* occurs at point *B*. Suppose, however, that the relative price of good 1 in country *B* is lower than in country *A*, due to the higher relative supply of good 1 there, so that the prices in country *B* are shown by the slope of the line *CC*. Production in country *B* then occurs at point *B'* rather than *B*. We have drawn the case where the budget lines *AA* and *CC* are both tangent to an indifference curve *U*, indicating that the utility of the representative consumer is *the same* in the two countries.

In the case we have illustrated, the production points *A* and *B'* lie on the same ray from the origin so that the relative outputs of the two goods are unchanged. This means that any index of real output in country *B* relative to country *A* should equal $OB'/OA > 1$, which is the proportional increase in both outputs in country *B*. This is the real GDP of country *B* relative to country *A* as measured on the output side, or real GDP^o in country *B* relative to country *A*. In practice, real GDP across the two countries could be measured by deflating the ratio of nominal GDP by an exact GDP price index, such as the Fisher Ideal price index. That GDP price index would use the production quantities of the two countries in its construction.

Now suppose we pose a different question, and ask how the welfare of the representative consumer compares in the two countries. The representative consumer has

equal utility using the indifference curve shown by U in figure A.2. It follows that an exact measure of consumer welfare, or real GDP^e measured on the expenditure side, would be unity in one country relative to the other. In practice, real GDP^e in one country relative to the other would be measured by deflating relative nominal expenditure in the two countries by an exact consumer price index, which would be constructed using *consumption quantities* rather than *production quantities* within the index. Another way to think about the difference between the GDP price index and the consumer price index is that the GDP price index would include the prices of exports and imports (including imported intermediate inputs), whereas the consumer price index would only include the price of final goods (including consumption goods, investment goods, and government expenditure, but not including exports or imports).

We conclude that real GDP across countries can be measured in two ways: either including the prices of exports and imports, resulting in an output-side or production measure, or excluding trade goods, resulting in an expenditure-side or welfare measure. The measurement of real GDP is the goal of the Penn World Table (PWT). So we might ask, Which concept does the PWT use as “real GDP”—the output-side measure real GDP^o, or the expenditure-side measure real GDP^e? It turns out that the measurement of real GDP in PWT up to version 7 is closest to the expenditure-side measure real GDP^e. That is especially true in a *benchmark year*, where real GDP across countries is constructed using the prices of final goods collected by the International Comparisons Project (ICP). The ICP does not include the prices of exports or imports, but only consumption goods, investment goods, and government expenditures, so that an expenditure-side measure of real GDP across countries is obtained. Each version of PWT up to version 7 uses the most recent benchmark years of the ICP, and then PWT extrapolates real GDP forward and backward using the growth rate of real GDP from the national accounts of each country.

Beginning with PWT version 8, improvements have been made to allow for the measurement of both real GDP^e and real GDP^o, and also to allow for the use of ICP prices from multiple benchmark years. The formulas used to obtain these two measures of real GDP as well as results for a limited range of countries are discussed in Feenstra, Heston, Timmer, and Deng (2009). In order to obtain real GDP^o, those authors use prices for exports and imports that are in fact unit-values in trade, that is, the customs values of narrowly defined categories of goods divided by their quantities. It is well known that such unit-values suffer from including differences in the qualities of narrowly defined goods exported or imported by different countries. As discussed in chapter 9, it is possible to correct the unit-values for imputed quality of exports and imports using the methods of Feenstra and Romalis (2014). That correction has been incorporated into PWT version 8, and later versions, as described in Feenstra, Inklaar, and Timmer (2015). In addition to using the quality-adjusted prices of exports and imports to obtain real GDP^o, PWT version 8 also includes different measures of real GDP over time, depending on whether only the most recent benchmark year of ICP data are used, or whether multiple years of ICP data are used. The reader is referred to Feenstra, Inklaar, and Timmer (2015) for a detailed explanation of this “next generation” of the PWT.

APPENDIX B

Discrete Choice Models

Discrete choice models are commonly taught in econometrics, but the economic generality of these models is sometimes unclear. This is especially true when we consider models that incorporate not just demand but also the optimal prices charged by firms under imperfect competition. In that case, we want to derive demand and prices in a mutually consistent fashion, as will be discussed in this appendix.

We begin with an aggregation theorem due to McFadden (1978, 1981). He has shown that when individual have random utility over discrete alternatives, the decisions that they make can sometimes be aggregated up to a representative consumer. McFadden focuses on the case where individuals have utility functions that are linear in income, and where the goods over which a discrete choice is being made are purchased in the quantity 0 or 1. Under certain conditions, it is possible to generalize this to the case where utility is nonlinear in income, and where the discrete-choice good is purchased in continuous quantities. This is a principal result of Anderson, de Palma, and Thisse (1992, chapters 2–5) who show, for example, that the representative consumer can have a CES utility function, as we also demonstrate below.

At the same time as presenting these aggregation results, we will discuss the empirical strategy to estimate discrete choice models due to Berry (1994). In this approach, we suppose that the individual-level data on consumer choices is *not* available. Rather, we just observe the market-level data on the quantity purchased of each product, as well as their prices and characteristics. The goal is to infer the individual taste parameters from this market-level data on quantities and prices, as well as parameters of marginal cost, so that price-cost margins can be estimated.

DISCRETE CHOICE MODEL

Suppose there is a population of consumers h , and we normalize its size to unity. Each consumer must decide which of a discrete number of alternatives $j=1,\dots,N$ to purchase, and receives the following utility from consuming one unit of product j :

$$V_j^h = u_j + \varepsilon_j^h, \quad (\text{B.1})$$

where u_j is the utility obtained from product j by *every* consumer, while ε_j^h is an additional part of utility obtained by consumer h . We will treat ε_j^h as a random

variable with cumulative density function $F(\varepsilon)$, where each consumer obtains a different draw of $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$. Given this draw, the consumer chooses the product with highest utility. The probability that a consumer will choose product j is

$$\begin{aligned} P_j &= \text{Prob}[V_j \geq V_k \text{ for all } k=1, \dots, N] \\ &= \text{Prob}[u_j + \varepsilon_j \geq u_k + \varepsilon_k \text{ for all } k=1, \dots, N] \\ &= \text{Prob}[\varepsilon_j - \varepsilon_k \geq u_k - u_j \text{ for all } k=1, \dots, N]. \end{aligned} \quad (\text{B.2})$$

We can think of (B.2) as the probability that any consumer choose alternative j , or equivalently, as the expected fraction of the population that chooses that alternative. The solution for these choice probabilities depends on the distribution function $F(\varepsilon)$, as well as on the specification of utility u_j . A simplified version of utility used by Berry (1994) is

$$u_j = y + \beta' z_j - \alpha p_j + \xi_j \quad \alpha > 0. \quad (\text{B.3})$$

Thus, utility is linear in individual income y and decreasing in the price p_j , and also depends on the characteristics z_j of the product.⁸ The term ξ_j is another random element in utility, but unlike ε_j^h , it *does not* vary across consumers. We could think of ξ_j as some unmeasured characteristics of product j , which are random across products but not consumers.

In general, computing the choice probabilities from (B.2) is computationally intensive since it requires integration over various subsets of the domain of $F(\varepsilon)$. However, McFadden has shown that for a class of distribution functions $F(\varepsilon)$ known as the *generalized extreme value*, the expected demands can be *equivalently* obtained by simple differentiation.

THEOREM 1 (MCFADDEN 1978, 1981)

Let H be a nonnegative function defined over R_+^N that satisfies the following properties:

- (a) H is homogeneous of degree one.
- (b) $H \rightarrow \infty$ as any of its argument approaches infinity.
- (c) The mixed partial derivatives of H with respect to k variables exist and are continuous, nonnegative if k is odd and nonpositive if k is even, $k=1, \dots, N$.

Define the generalized extreme value distribution function,

$$F(\varepsilon_1, \dots, \varepsilon_N) \equiv \exp[-H(e^{-\varepsilon_1}, e^{-\varepsilon_2}, \dots, e^{-\varepsilon_N})]. \quad (\text{B.4})$$

Then the expected value of consumer utility (up to a constant) is given by the aggregate utility function,

$$G(u_1, \dots, u_N) \equiv \ln H(e^{u_1}, e^{u_2}, \dots, e^{u_N}), \quad (\text{B.5})$$

⁸This utility function is not homogeneous of degree zero in income y and price p_j . To obtain this property, we should measure y and p_j relative to a numeraire price p_0 , so that utility becomes $u_j = yp_0 - \alpha p_j/p_0 + \beta' z_j + \xi_j$, which is homogeneous of degree zero in (y, p_j, p_0) .

and the choice probabilities P_j in (B.2) can be obtained as

$$P_j = \partial G / \partial u_j. \quad (\text{B.6})$$

To interpret this theorem, consider the linear utility function in (B.3). Substituting this into (B.5) and using the linear homogeneity of H , we find that the aggregate utility function is

$$\tilde{G}(p_1, z_1, \dots, p_N, z_N, y) = y + \ln H(e^{\beta z_1 - \alpha p_1 + \xi_1}, \dots, e^{\beta z_N - \alpha p_N + \xi_N}). \quad (\text{B.5}')$$

Notice that $\partial \tilde{G} / \partial y = 1$, and consider the case where $\alpha = 1$. Then price is inversely related to utility ($du_j = -dp_j$), so the choice probabilities in (B.6) can be computed as $P_j = \partial G / \partial u_j = -(\partial \tilde{G} / \partial p_j) / (\partial \tilde{G} / \partial y)$. Thus, the result that $P_j = \partial G / \partial u_j$ in (B.6) of the theorem can be interpreted as saying that Roy's Identity holds at the aggregate level.

The conditions on H in (a)–(c) are technical properties needed to ensure that the $F(\varepsilon)$ defined in (B.4) is indeed a cumulative distribution function. There will be many choices of H satisfying these conditions, and for each, we obtain the aggregate utility function in (B.5), whose demands are identical to those aggregated from the individual choice problem in (B.1)–(B.2). Furthermore, this aggregate utility function can be used to make welfare statements about the individual consumers, since it reflects expected individual utility.

To see the usefulness of this result, we consider the well-known example where the random utility ε_j in (B.1) from consuming each product is independently distributed across products, with the extreme value distribution.

Example 1: Logit demand system

Let us choose the function H as linear in its arguments:

$$H(e^{-\varepsilon_1}, e^{-\varepsilon_2}, \dots, e^{-\varepsilon_N}) = \sum_{j=1}^N e^{-\varepsilon_j}. \quad (\text{B.7})$$

Substituting (B.7) into (B.4), the distribution function is

$$F(\varepsilon_1, \dots, \varepsilon_N) = \prod_{j=1}^N \exp(-e^{-\varepsilon_j}). \quad (\text{B.8})$$

This cumulative distribution function is therefore the product of N iid “double-exponential” or extreme value distributions, which apply to the random utility terms in (B.1). Therefore, the random term in utility is distributed as iid extreme value.⁹

Computing the choice probabilities as in (B.6) using (B.7) and (B.3), we obtain

$$P_j = \frac{e^{u_j}}{\sum_{k=1}^N e^{u_k}} = \frac{e^{\beta z_j - \alpha p_j + \xi_j}}{\sum_{k=1}^N e^{\beta z_k - \alpha p_k + \xi_k}}, \quad (\text{B.9})$$

which are the choice probabilities under the logit system.

⁹A discussion of the “double exponential” or extreme value distribution is provided by Anderson, de Palma, and Thisse (1992, 58–62).

Berry (1994) discusses how estimates of α and β can be obtained even if we do not have data on the purchases by each individual, but just observe the quantity share s_j of each product in demand, as well as their prices and characteristics. Then the probabilities in (B.9) would be measured by these quantity shares s_j . Suppose in addition there is some outside option $j = 0$, which gives utility normalized to zero, $u_0 = 0$. Then setting $s_j = P_j$, and taking logs of the ratio of (B.9) to s_0 , we obtain

$$\ln s_j - \ln s_0 = \beta' z_j - \alpha p_j + \xi_j. \quad (\text{B.10})$$

In addition, we follow Berry to solve for the optimal prices of the firm, where we assume for simplicity that each firm produces only one product. Denoting the marginal costs of producing good j by $g_j(z_j)$, and ignoring fixed costs, we find that the profits from producing model j are

$$\pi_j = [p_j - g_j(z_j)] s_j. \quad (\text{B.11})$$

Maximizing (B.11) over the choice of p_j , treating the prices of all other products as fixed, we obtain

$$p_j = g_j(z_j) - s_j / (\partial \ln s_j / \partial p_j) = g_j(z_j) - (\partial \ln s_j / \partial p_j)^{-1}. \quad (\text{B.12})$$

For the special case of the logit system (B.9) with $s_j = P_j$ denoting the quantity shares, we compute that $\partial \ln s_j / \partial p_j = -\alpha(1 - s_j)$. If we also specify marginal costs as $g_j(z_j) = \gamma' z_j + \omega_j$, where ω_j is a random error, then from (B.12) the optimal prices are

$$p_j = \gamma' z_j + \frac{1}{\alpha(1 - s_j)} + \omega_j, \quad (\text{B.13})$$

which can be estimated jointly with (B.10). It is apparent, though, that the random error ξ_j influences the market shares in (B.10), and therefore from (B.13) will be correlated with prices p_j . Accordingly, the joint estimation of (B.10) and (B.13) should be done with instrumental variables.¹⁰

The problem with this simple logit example is that the demand elasticities obtained are implausible. From the market shares $s_j = P_j$ in (B.9) we readily see that s_j/s_k is independent of the price or characteristics of any third product i . This property is known as the “*independence of irrelevant alternatives*” in the discrete choice literature, and it implies that the cross-elasticity of demand between products j and k and the third product i must be equal. To improve on this, we consider the nested logit system.

Example 2: Nested Logit demand system

Now suppose that the consumers have a choice between two levels of the differentiated product. First, an individual decides whether to purchase a product in each of $g = 1, \dots, G$ groups (e.g., small cars and big cars), and second, the individual decides which of the products in that group to purchase. Suppose that the products available in each group g are denoted by the set $J_g \subset \{1, \dots, N\}$, while J_0 denotes the outside

¹⁰Berry (1994, 249) suggests that appropriate instruments for prices in (B.10) are characteristics of other models z_k , as well as variables that affect the costs of producing product j .

option. Utility from consumer h is still given by (B.1), where the errors ε_j are distributed as extreme value but are *not* independent. Instead, we suppose that if consumer h has a high value of ε_j^h for some product $j \in J_g$, then that person will *also* tend to have high values of ε_k^h for all other products $k \in J_g$, so ε_j is positively correlated across products in each group. For example, an individual who has a preference for some small car also tends to like other small cars.

To achieve this correlation between the random errors in (B.1), McFadden (1981, 228) chooses the function H as

$$H(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_N}) = \sum_{g=0}^G \left[\sum_{j \in J_g} e^{-\varepsilon_j/(1-\rho_g)} \right]^{(1-\rho_g)}. \quad (\text{B.14})$$

To satisfy property (c) of theorem 1, we need to specify that $0 \leq \rho_g < 1$. Using this choice of H , we obtain a distribution function $F(\varepsilon)$ from (B.4),

$$F(\varepsilon_1, \dots, \varepsilon_N) = \exp \left(\sum_{g=0}^G \left[\sum_{j \in J_g} -e^{-\varepsilon_j/(1-\rho_g)} \right]^{(1-\rho_g)} \right), \quad (\text{B.15})$$

where ρ_g roughly measures the correlation between random terms ε_j within a group.¹¹ Computing the choice probabilities as in (B.6) using (B.14), we obtain

$$P_j = \frac{e^{u_j/(1-\rho_g)}}{D_g} \frac{D_g^{(1-\rho_g)}}{\left[\sum_{g=0}^G D_g^{(1-\rho_g)} \right]}, \quad \text{for } j \in J_g, \quad (\text{B.16})$$

where the term $D_g \equiv \sum_{k \in J_g} e^{u_k/(1-\rho_g)}$ appearing in (B.16) is called an *inclusive value*, since it summarizes the utility obtained from all products in the group g .

Berry (1994, 252) motivates this nested logit case somewhat differently. He rewrites the random errors ε_j^h as

$$\varepsilon_j^h = \zeta_g^h + (1-\rho_g) e_j^h, \quad \text{for } j \in J_g, \quad (\text{B.17})$$

where the errors e_j^h are iid extreme value. The random variable ζ_g^h in (B.17) is *common* to all products within group g , and therefore induces a correlation between the random utilities for those products. Cardell (1997) shows that there exists a distribution for ζ_g^h (depending on ρ_g) such that when e_j^h are iid extreme value, then ε_j^h are *also* distributed extreme value but are *not* independent. Notice that as the parameter ρ_g approaches unity, ε_j^h are perfectly correlated within the group g (since they equal ζ_g^h), whereas when ρ_g approaches zero (in which case ζ_g^h also approaches zero) then ε_j^h become independent. Using the errors in (B.17) gives the same choice probabilities as shown in (B.16).

Using the nested logit choice probabilities in (B.16), we can rederive the estimating equations for market share and optimal prices, as does Berry (1994, 252–53). The first term on the right-hand side of (B.16) is the probability that an individual will choose product $j \in J_g$ *conditional on* having already chosen the group g . Let us denote this conditional probability by $s_{j|g}$. The second term on the right-hand side of (B.16) is the probability of choosing any product from group g , which we write as \bar{s}_g .

¹¹Johnson and Kotz (1972, 256) report that the parameters $(1-\rho_g)$ equal $\sqrt{1 - \text{corr}(\varepsilon_i, \varepsilon_k)}$, for $i, k \in J_g$ and $i \neq k$, so that $\text{corr}(\varepsilon_i, \varepsilon_k) > 0$ for $\rho_g > 0$.

So replacing P_j in (B.16) by the market shares s_j , we write this choice probability as $s_j = s_{j|g} \bar{s}_g$. In addition, we suppose that the outside good has $u_0 = 0$ and inclusive value $D_0 = 1$, so that from (B.16), $s_0 = P_0 = [\sum_{g=0}^G D_g^{(1-\rho_g)}]^{-1}$. Taking logs of the ratio $s_j/s_0 = s_{j|g} \bar{s}_g/s_0$ and using the linear utility u_j in (B.3), we therefore have

$$\ln s_j - \ln s_0 = \frac{(\beta' z_j - \alpha p_j + \xi_j)}{(1 - \rho_g)} - \rho_g \ln D_g. \quad (\text{B.18})$$

To solve for the inclusive value D_g , recall that \bar{s}_g equals the second term on the right of (B.16). Therefore, $\bar{s}_g/s_0 = D_g^{(1-\rho_g)}$ and so $\ln \bar{s}_g - \ln s_0 = (1 - \rho_g) \ln D_g$. Using this in (B.18) and simplifying with $s_j = s_{j|g} \bar{s}_g$, we obtain

$$\ln s_j - \ln s_0 = \beta' z_j - \alpha p_j + \rho_g \ln s_{j|g} + \xi_j, \quad j \in J_g. \quad (\text{B.19})$$

This gives us a regression to estimate the parameters (α, β) , where the final term $\ln s_{j|g}$ measures the market share of product j within the group g , and is endogenous. Once again, we can use instrumental variables to estimate (B.19).¹²

In addition, we follow Berry (1994, 255) to solve for the optimal prices of the firm, assuming for simplicity that it produces only one product j . Profits are still given by (B.11), and the first-order condition by (B.12). For the nested logit system with $s_j = P_j$ in (B.16), the derivative of the log market shares is $\partial \ln s_j / \partial p_j = -\frac{\alpha}{(1-\rho_g)} [1 - \rho_g \bar{s}_{j|g} - (1 - \rho_g) s_j]$. With marginal costs as $g_j(z_j) = \gamma' z_j + \omega_j$, where ω_j is a random error, from (B.11) the optimal prices are

$$p_j = \gamma' z_j + \frac{(1 - \rho_g)}{\alpha [1 - \rho_g \bar{s}_{j|g} - (1 - \rho_g) s_j]} + \omega_j, \quad (\text{B.20})$$

which can be estimated simultaneously with (B.19) using instrumental variables. This pricing equation was generalized in chapter 8 to allow for multiproduct firms, as used by Irwin and Pavcnik (2004) in their study of export subsidies to commercial aircraft.

The nested logit allows for more general substitution between products than does the simple logit model (and does not suffer from “independence of irrelevant alternatives”). Goldberg (1995), for example, has used the nested logit in her study of automobile demand in the United States and the impact of the voluntary export restraint (VER) with Japan. Even more flexibility in the pattern of demand across autos can be achieved by introducing greater consumer heterogeneity, as done by Berry, Levinsohn, and Pakes (1995, 1999) in their work on autos and the VER with Japan. They suppose that the utilities in (B.3) instead appear as

$$u_j^h = \beta^h' z_j - \alpha p_j + \xi_j, \quad \alpha > 0, \quad (\text{B.3}')$$

where now the parameters β^h depend on the individual h , reflecting both demographic characteristics and income. Since individual income is included in β^h , without loss of generality we omit this variable from appearing explicitly in (B.3').

If we observed both individual characteristics and also their discrete choices, then it would be possible to estimate the other parameters appearing in (B.3') using standard econometric programs for discrete choice. However, when we observe only the

¹²Berry (1994, 254) suggests that appropriate instruments will include the characteristics of other products or firms in group g .

market-level demands and overall distribution of individual characteristics, then estimating the parameters of (B.3') becomes much more difficult. To see this, let us suppose that the individual characteristics β^h are distributed as $\bar{\beta} + \eta^h$, with mean value $\bar{\beta}$ over the population and η^h , a random variable that captures individual tastes for characteristics. Substituting this into (B.3') and then into (B.1), we obtain utility,

$$V_j^h = \bar{\beta}' z_j - \alpha p_j + \xi_j + (\eta^h' z_j + \varepsilon_j^h), \quad (\text{B.1}')$$

where now the random error includes the interaction term $\eta^h' z_j$ between individual tastes and product characteristics. The presence of this interaction term means that theorem 1 *cannot* be used: the cumulative distribution of the combined error $(\eta^h' z_j + \varepsilon_j^h)$ is more general than allowed for in theorem 1. It follows that there is no closed-form solution such as (B.6) for the choice probabilities.

Instead, the choice probabilities need to be calculated numerically from (B.2). That is, given the data (p_j, z_j) across products, we can simulate the distribution of $(\eta^h' z_j + \varepsilon_j^h)$. For each draw from this distribution, and for given parameters $(\alpha, \bar{\beta}, \xi_j)$, we can calculate utility V_j^h for each product. The choice probabilities P_j in (B.2) are computed as the proportion of draws for which product j gives the highest utility. Then the parameters $(\alpha, \bar{\beta}, \xi_j)$ are chosen so that these choice probabilities match the observed market shares s_j as closely as possible. Likewise, the predicted *change* in the market shares due to prices, $(\partial s_j / \partial p_j)$ in (B.12), would be calculated numerically by seeing how the choice probability P_j varies with prices p_j . This change also depends on the parameters $(\alpha, \bar{\beta}, \xi_j)$, which are then chosen so that the pricing equations fit as closely as possible. Thus, the market share and pricing equations are simultaneously used to estimate the underlying taste and costs parameters, using simulated distributions of the random term in utility. This is the idea behind the work of Berry, Levinsohn, and Pakes (1995, 1999), which readers should consult for more details.¹³

It is also useful to compare the above approaches with a *third* approach to discrete choice models: rather than using a representative consumer, or individuals with random utility, we could instead use individuals that differ in terms of their *preferred* characteristics, in what is sometimes called the “ideal variety” approach. Anderson, de Palma, and Thisse (1992) derive an equivalence between all three approaches. Bresnahan (1981) was the first to estimate a discrete choice model for autos in which consumers differ in their ideal varieties, and found that higher-priced models tended to have higher percentage markups. Feenstra and Levinsohn (1995) generalized the model of Bresnahan by allowing characteristics to differ multidimensionally, rather than along a line as in Bresnahan, and showed that the optimal prices of firms will vary with the distance to their neighbors.

DISCRETE CHOICE WITH CONTINUOUS QUANTITIES

Setting aside the issue of consumer heterogeneity, there is another direction in which we can attempt to generalize theorem 1, while still allowing the aggregate utility function to exist so that demands can be computed using Roy’s Identity. We again suppose

¹³Estimating a discrete choice model by simulating the taste parameters is referred to as “mixed logit.” See, for example, McFadden and Train (1997) and Revelt and Train (1999). A distance learning course on discrete choice methods, including software to estimate mixed logit models, is available from Kenneth Train at <http://elsa.berkeley.edu/~train/distant.html>.

that consumers are still choosing over products $j=1,\dots,N$, and now have the utility function

$$V_j^h = \ln y - \ln \varphi(p_j z_j) + \varepsilon_j^h, \quad j=1,\dots,N, \quad (\text{B.21})$$

where p_j is the price of product j , z_j is a vector of characteristics, y is the consumers' income, and ε_j is a random term that reflects the additional utility received by consumer h from product j . Thus, income no longer enters utility in a linear fashion, and we will allow individuals to consume *continuous* quantities (not restricted to 0 or 1) of their preferred product j .

Each consumer still chooses his or her preferred product j with probability

$$P_j = \text{Prob}[V_j \geq V_k] \quad \text{for all } k=1,\dots,N. \quad (\text{B.22})$$

If product j is chosen, then the quantity consumed is determined from the indirect utility function in (B.21), using Roy's Identity:

$$c_j = -\frac{\partial V_j^h / \partial p_j}{\partial V_j^h / \partial y} = y(\partial \ln \varphi / \partial p_j). \quad (\text{B.23})$$

It follows that expected demand for each product is

$$X_j = c_j P_j. \quad (\text{B.24})$$

This formulation of the consumer's problem is somewhat more general than that which we considered earlier, because now we are allowing the consumer to make a *continuous* choice of the quantity purchased. This falls into the category of so-called continuous/discrete models (see Train 1986, chapter 5). Our formulation of the problem is simplified, however, because there is no uncertainty over the *quantity* of purchases in (B.23); the random term in utility affects only the choice of product in (B.22).¹⁴ It turns out that in this setting, the aggregation results of McFadden (1978, 1981) apply equally well, and we can extend theorem 1 as theorem 2.¹⁵

THEOREM 2

Let H be a nonnegative function defined over R_+^N that satisfies conditions (a)–(c) of theorem 1. Define the cumulative distribution function F as in (B.3), and define an aggregate indirect utility function,

$$G(p_1, z_1, \dots, p_N, z_N, y) = \ln y + \ln H[\varphi(p_1 z_1)^{-1}, \dots, \varphi(p_N z_N)^{-1}]. \quad (\text{B.25})$$

Then

- (a) Expected demand computed from (B.24) equals $-(\partial G / \partial p_j) / (\partial G / \partial y)$.
- (b) G equals expected individual utility in (B.21) (up to a constant).

¹⁴In contrast, Dubin and McFadden (1984) consider an application where there is uncertainty in both the discrete choice of the product and the continuous amount to consume.

¹⁵The proof of theorem 2, based on the results of McFadden, is provided in Feenstra (1995, Prop. 1).

Thus, we can still compute demand using Roy's Identity as in part (a), and use the aggregate utility function to infer welfare as in part (b). Notice that expected utility can be rewritten in a more familiar form by taking the monotonic transformation $\tilde{G} = \exp(G)$, so that

$$\tilde{G}(p_1, z_1, \dots, p_N, z_N, y) = yH[\varphi(p_1, z_1)^{-1}, \dots, \varphi(p_N, z_N)^{-1}]. \quad (\text{B.25}')$$

This can be interpreted as an indirect utility function for the representative consumer.

To see the usefulness of this result, we consider again a random utility ε_j in (B.21) that is independently distributed across products, with the extreme value distribution, and derive a CES aggregate utility function.

Example 3: CES demand system

Let us specify the individual utility functions as¹⁶

$$V_j^h = \ln y - \alpha \ln [p_j/f(z_j)] + \varepsilon_j^h, \quad \alpha > 0, \quad (\text{B.26})$$

where we are measuring prices relative to consumers' perceived "quality" of products $f(z_j)$.

We will choose the function H as linear in its arguments, as in (B.7). It follows that the cumulative distribution function in (B.8) is the product of N iid extreme value distributions, so the errors in (B.26) are distributed as iid extreme value.

Then using (B.7), (B.25), and (B.26), we obtain the aggregate utility function as

$$G(p_1, z_1, \dots, p_N, z_N, y) = y \sum_{j=1}^N [p_j/f(z_j)]^{-\alpha}, \quad (\text{B.27})$$

so that expected aggregate demand is

$$-\frac{\partial G/\partial p_j}{\partial G/\partial y} = y \left(\frac{\alpha p_j^{-\alpha-1}/f(z_j)^{-\alpha}}{\sum_{k=1}^N [p_k/f(z_k)]^{-\alpha}} \right). \quad (\text{B.28})$$

Thus, we obtain a CES indirect utility function in (B.27) for the aggregate consumer, with elasticity of substitution with $\sigma = 1 + \alpha$, and the associated CES demand in (B.28). This is an alternative demonstration of the result in Anderson, de Palma, and Thisse (1989; 1992, 85–90), that the CES utility function arrives from a discrete choice model with iid extreme value errors in random utility. Notice that this CES demand system is obtained with exactly the same assumptions on the random term in utility as the logit system (both used iid extreme value distributions), but differs from the logit system in allowing for *continuous* quantities demanded of the discrete good.

Example 4: Nested CES demand system

As in our above discussion of the nested logit model, suppose that the consumers have a choice over two levels of the differentiated product. First, an individual decides

¹⁶This utility function is not homogeneous of degree zero in income y and price p_j . To obtain this property, we should measure y and p_j relative to a numeraire price p_0 , so that utility becomes $\ln(y/p_0) - \alpha \ln[(p_j/p_0)/f(z_j)]$, which is homogeneous of degree zero in (y, p_j, p_0) .

whether to purchase a product in each of $g = 1, \dots, G$ groups, and second, the individual decides which of the products in that group to purchase. Let the products available in each group g be denoted by the set $J_g \subset \{1, \dots, N\}$, while J_0 denotes the outside option. Utility for consumer h is still given by (B.26), where the errors ε_j are distributed as extreme value but are *not* independent. We will specify the function H as in (B.14), with the cumulative density function F as in (B.15).

Then using (B.14), (B.25), and (B.26), we obtain the aggregate utility function

$$G(p_1, z_1, \dots, p_N, z_N, y) = y \sum_{g=0}^G \left(\sum_{j \in J_g} [p_j/f(z_j)]^{-\alpha/(1-\rho_g)} \right)^{(1-\rho_g)}. \quad (\text{B.29})$$

Applying Roy's Identity, we can readily obtain the expected demand for product j ,

$$X_j = -\frac{\partial G / \partial p_j}{\partial G / \partial y} = \left(\frac{\alpha y}{p_j} \right) \frac{[p_j/f(z_j)]^{-\alpha/(1-\rho_g)}}{D_g} \frac{D_g^{(1-\rho_g)}}{\left[\sum_{g=1}^G D_g^{(1-\rho_g)} \right]}, \quad \text{for } j \in J_g, \quad (\text{B.30})$$

where the term $D_g \equiv \sum_{j \in J_g} [p_j/f(z_j)]^{-\alpha/(1-\rho_g)}$ in (B.30) is again the “inclusive value,” analogous to (B.16). Notice that expected demand on the right-hand side of (B.30) is composed of three terms: the first term, $(\alpha y/p_j)$, is a conventional Cobb-Douglas demand function, reflecting the continuous demand for the product given the utility function in (B.26); the second term, $[p_j/f(z_j)]^{-\alpha/(1-\rho_g)}/D_g$, is the share of product j in the demand for group g ; and the third term, $D_g^{(1-\rho_g)}/[\sum_{g=1}^G D_g^{(1-\rho_g)}]$, is the share of group g in the total demand for the product. The latter two terms both appear in the nested logit system (B.16), so the new feature of (B.30) is the continuous demand of $(\alpha y/p_j)$.

We can easily relate the parameters α and ρ_g to the elasticity of substitution between products. Notice that for two products $i, j \in J_g$, we obtain relative demand from (B.30),

$$\frac{X_i}{X_j} = \left(\frac{f(z_i)}{f(z_j)} \right)^{\alpha/(1-\rho_g)} \left(\frac{p_i}{p_j} \right)^{-[\alpha/(1-\rho_g)]-1}, \quad i, j \in J_g. \quad (\text{B.31})$$

Thus, the elasticity of substitution between two products in the same group is $1 + [\alpha/(1-\rho_g)] > 1$.

In addition, we can compute from (B.29) that the ratio of expenditure on products from two different groups g and h :

$$\frac{\sum_{i \in J_g} p_i X_i}{\sum_{j \in J_h} p_j X_j} = \frac{D_g^{(1-\rho_g)}}{D_h^{(1-\rho_h)}}, \quad i \in J_g \text{ and } j \in J_h. \quad (\text{B.32})$$

Let us define the price index for the products in group g as

$$P_g \equiv \left[\sum_{j \in J_g} [p_j/f(z_j)]^{-\alpha/(1-\rho_g)} \right]^{-(1-\rho_g)/\alpha} = D_g^{-(1-\rho_g)/\alpha}.$$

Then we can rewrite the ratio of expenditures in (B.32) as

$$\frac{(\sum_{i \in J_g} p_i X_i) / P_g}{(\sum_{j \in J_h} p_j X_j) / P_h} = \left(\frac{P_g}{P_h} \right)^{-\alpha-1}, \quad i \in J_g \text{ and } j \in J_h. \quad (\text{B.33})$$

The left side of (B.33) is ratio of expenditures on groups g and h , deflated by their price indexes, so it can be interpreted as a ratio of real expenditures. We see that this depends on the ratio of price indexes for the two groups, with the elasticity $1 + \alpha > 1$. Feenstra, Hanson, and Lin (2004) use a slightly more general version of this nested CES structure to estimate the gains from having Hong Kong traders act as intermediaries for firms outsourcing with China.

REFERENCES

- Abramowitz, Moses. 1956. "Resource and Output Trends in the U.S. Since 1870." *American Economic Review Papers and Proceedings*, 46(2): 5–23.
- Acemoğlu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2014. "Import Competition and the Great U.S. Employment Sag of the 2000s." NBER Working Paper No. 20395
- Acemoğlu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–401.
- Acemoğlu, Daron, and Jaume Ventura. 2002. "The World Income Distribution." *Quarterly Journal of Economics* 117(2): 659–94.
- Aghion, Phillippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105(1): 1–29.
- Aho, C. Michael, and Thomas O. Bayard. 1984. "Costs and Benefits of Trade Adjustment Assistance." Pages 153–92 in Robert E. Baldwin and Anne O. Krueger, eds., *The Structure and Evolution of Recent U.S. Trade Policy*. Chicago: University of Chicago and NBER.
- Aitken, Brian, and Ann E. Harrison. 1999. "Do Domestic Firms Benefit from Foreign Investment? Evidence from Venezuela." *American Economic Review* 89: 605–18.
- Aitken, Brian, Ann E. Harrison, and Robert Lipsey. 1996. "Wages and Foreign Ownership: A Comparative Study of Mexico, Venezuela, and the United States." *Journal of International Economics* 40(3/4): 345–71.
- Akerlof, George, Andrew Rose, Janet Yellen, and Helga Hessenius. 1991. "East Germany in from the Cold: The Economic Aftermath of Currency Union." *Brookings Papers on Economic Activity* 1: 1–87.
- Alchian, Armen A., and William R. Allen. 1964. *University Economics*. Belmont, Calif.: Wadsworth.
- Alesina, Alberto, and Dani Rodrik. 1994. "Distribute Politics and Economic Growth." *Quarterly Journal of Economics* 109(2): 465–90.
- Alessandria, George. 2004. "International Deviations from the Law of One Price: The Role of Search Frictions and Market Share." *International Economics Review* 45(4): 1263–91.
- Allanson, Paul, and Catia Montagna. 2005. "Multiproduct Firms and Market Structure: An Explorative Application to the Product Life Cycle." *International Journal of Industrial Organization* 23(7–8): 587–97.
- Alterman, William F., W. Erwin Diewert, and Robert C. Feenstra. 1999. *International Trade Price Indexes and Seasonal Commodities*. U.S. Department of Labor, Bureau of Labor Statistics, Washington, D.C.
- Alvarez, Fernando, and Robert Lucas, Jr. 2007. "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade." *Journal of Monetary Economics* 54(6): 1726–68.
- Amiti, Mary. 2014. "Do China's Productivity Gains Benefit US Consumers?" Federal Reserve Bank of New York.
- Amiti, Mary, and Donald R. Davis. 2012. "Trade, Firms, and Wages: Theory and Evidence." *Review of Economic Studies* 79(1): 1–36.
- Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2012. "Importers, Exporters, and Exchange Rate Disconnect." NBER Working Paper No. 18615.

- Amiti, Mary, and Shang-Jin Wei. 2005a. "Service Offshoring, Productivity, and Employment: Evidence from the United States." IMF Working Paper 05/238, International Monetary Fund, Washington, D.C.
- . 2005b. "Fear of Service Outsourcing: Is It Justified?" *Economic Policy* 20(42): 308–47.
- . 2009. "Service Offshoring and Productivity: Evidence from the United States." *The World Economy* 32(2): 203–20.
- Anderson, James A. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review* 69: 106–16.
- . 1985. "The Relative Inefficiency of Quotas: The Cheese Case." *American Economic Review* 75(1): 178–90.
- . 1988. *The Relative Inefficiency of Quotas*. Cambridge: MIT Press.
- Anderson, James A., and J. Peter Neary. 1996. "A New Approach to Evaluating Trade Policy." *Review of Economic Studies* 63: 107–25.
- Anderson, James A., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93(1): 170–92.
- Anderson, Simon P., Andre de Palma and Jacques-François Thisse. 1989. "Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach." *Review of Economic Studies* 56: 21–35.
- . 1992. *Discrete Choice Theory of Product Differentiation*, Cambridge: MIT Press.
- Anderton, Bob, and Paul Brenton. 1999. "Outsourcing and Low-Skilled Workers in the UK." *Bulletin of Economic Research* 51(4): 267–85.
- Antràs, Pol. 2003. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics* 118: 1375–418.
- . 2005. "Incomplete Contracts and the Product Cycle." *American Economic Review* 95(4): 1054–73.
- . 2015. *Global Production: Firms, Contracts, and Trade Structure*. Princeton: Princeton University Press.
- Antràs, Pol, and Davin Chor. 2013. "Organizing the Global Value Chain." *Econometrica* 81(6): 2127–204.
- Antràs, Pol, David Chor, Thibault Fally, and Russell Hillberry. 2012. "Measuring the Upstreamness of Production and Trade Flows." NBER Working Paper No. 17819.
- Antràs, Pol, and Arnaud Costinot. 2011. "Intermediated Trade." *Quarterly Journal of Economics* 126(3): 1319–74.
- Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2006a. "Offshoring in a Knowledge Economy." *The Quarterly Journal of Economics*: 121(1): 31–77.
- . 2006b. "Organizing Offshoring: Middle Managers and Communication Costs." NBER Working Paper No. 12196.
- Antràs, Pol, and Elhanan Helpman. 2004. "Global Sourcing." *Journal of Political Economy* 112(3): 552–80.
- Antràs, Pol, and Esteban Rossi-Hansberg. 2009. "Organizations and Trade." *Annual Review of Economics, Annual Reviews* 1(1): 43–64.
- Antràs, Pol, and Robert W. Staiger. 2012. "Offshoring and the Role of Trade Agreements." *American Economic Review* 102(7): 3140–83.
- Antweiler, Werner, and Daniel Trefler. 2002. "Increasing Returns and All That: A View from Trade." *American Economic Review* 92(1): 93–119.
- Archibald, Robert B. 1977. "On the Theory of Industrial Price Measurement: Output Price Indexes." *Annals of Economic and Social Measurement* 6(1): 57–72.
- Arkolakis, Costas. 2010. "Market Penetration Costs and the New Consumers Margin in International Trade." *Journal of Political Economy* 118: 1151–99.
- Arkolakis, Costas, Arnaud Costinot, David Donaldson, and Andrés Rodríguez-Clare. 2012. "'The Elusive Pro-Competitive Effect of Trade,' March 16. <http://www.econ.yale.edu/~ka265/research/ACDR/ACDRdraft.pdf> (accessed 10/16/2014).
- Arkolakis, Costas, Arnaud Costinot, and Andrés Rodriguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102(1): 94–130.

- Arkolakis, Costinot, Sharat Ganapati, and Marc-Andreas Muendler. 2014. "The Extensive Margin of Exporting Products: A Firm-Level Analysis." http://www.econ.yale.edu/~ka265/research/MultiProduct/Arkolakis_Ganapati_Muendler_products.pdf (accessed 11/17/2014).
- Armington, Paul S. 1969. "A Theory of Demand for Products Distinguished by Place and Production." *IMF Staff Papers* 16: 159–78.
- Arndt, Sven. 1997. "Globalization and the Open Economy." *North American Journal of Economics and Finance* 8: 71–79.
- . 1998a. "Globalization and the Gains from Trade." In K. Jaeger and K.-J. Koch, eds., *Trade, Growth, and Economic Policy in Open Economies*. New York: Springer-Verlag.
- . 1998b. "Super-specialization and the Gains from Trade." *Contemporary Economic Policy* 16: 480–85.
- Arndt, Sven, and Henryk Kierzkowski, eds. 2001. *Fragmentation: New Production and Trade Patterns in the World Economy*. Oxford: Oxford University Press.
- Atkeson, Andrew, and Ariel Tomás Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98(5): 1998–2031.
- . 2010. "Innovation, Firm Dynamics, and International Trade." *Journal of Political Economy* 118(3): 433–84.
- Atkin, David, and Dave Donaldson. 2014. "Who's Getting Globalized? The Size and Implications of Intranational Trade Costs." https://dl.dropboxusercontent.com/u/2734209/Atkin_Donaldson_WGG_paper.pdf (accessed 10/22/2014).
- Auer, Raphael, and Raphael Schoenle. 2012. "Market Structure and Exchange Rate Pass-Through." Federal Reserve Bank of Dallas, Working Paper No. 130.
- Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.
- Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics* 90(2): 300–323.
- Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113(4): 1169–213.
- Aw, Bee Yan. 1983. "The Interpretation of Cross-Section Regression Tests of the Heckscher-Ohlin Theorem with Many Goods and Factors." *Journal of International Economics* 14(1/2): 163–67.
- Bagwell, Kyle. 1991. "Optimal Export Policy for a New-Product Monopoly." *American Economic Review* 81(5): 1156–69.
- Bagwell, Kyle, and Robert W. Staiger. 1999. "An Economic Theory of GATT." *American Economic Review* 89(1): 215–48.
- . 2002. *The Economics of the World Trading System*. Cambridge: MIT Press.
- . 2011. "What Do Trade Negotiators Negotiate About? Empirical Evidence from the World Trade Organization." *American Economic Review* 101(4): 1238–73.
- Baier, Scott, and Jeffrey H. Bergstrand. 2001. "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity." *Journal of International Economics* 53(1): 1–27.
- . 2003. "Endogenous Free Trade Agreements and the Gravity Equation." University of Notre Dame, manuscript.
- . 2009. "Bonus vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation." *Journal of International Economics* 77(1): 77–85.
- Baldwin, Richard E. 1995. "A Domino Theory of Regionalism." Pages 25–48 in Richard E. Baldwin, P. Haaparanta, and J. Kiander, eds., *Expanding Membership in the European Union*. Cambridge: Cambridge University Press.
- Baldwin, Richard E., and Rikard Forslid. 2010. "Trade Liberalization with Heterogeneous Firms." *Review of Development Economics* 14(2): 161–76.
- Baldwin, Richard E., and Tadashi Ito. 2011. "Quality Competition versus Price Competition Goods: An Empirical Classification." *Journal of Economic Integration* 26: 110–35.
- Baldwin, Richard E., and Paul R. Krugman. 1988a. "Market Access and International Competition: A Simulation Study of 16K Random Access Memories." In Robert C. Feenstra, ed.,

- Empirical Methods for International Trade.* Cambridge: MIT Press. Reprinted in Gene M. Grossman, ed., *Imperfect Competition and International Trade.* Cambridge: MIT Press, 1992.
- . 1988b. “Industrial Policy and International Competition in Wide-Bodied Aircraft.” Pages 45–71 in Robert E. Baldwin, ed., *Trade Policy Issues and Empirical Analysis.* Chicago: University of Chicago Press and NBER.
- Baldwin, Richard E., and Frédéric Robert-Nicoud. 2008. “Trade and Growth with Heterogeneous Firms.” *Journal of International Economics* 74(1): 21–34.
- . 2014. “Trade-in-goods and Trade-in-tasks: An Integrating Framework.” *Journal of International Economics* 92(1): 51–62.
- Baldwin, Richard E., and Anthony J. Venables. 2013. “Spiders and Snakes: Offshoring and Agglomeration in the Global Economy.” *Journal of International Economics* 90(2): 245–54.
- Baldwin, Robert E. 1969. “The Case against Infant-Industry Protection.” *Journal of Political Economy* 77: 295–305.
- . 1971. “Determinants of the Commodity Structure of U.S. Trade.” *American Economic Review* 61: 126–46.
- Baldwin, Robert E., and G. G. Cain. 2000. “Shifts in Relative U.S. Wages: The Role of Trade, Technology and Factor Endowments.” *Review of Economics and Statistics* 82: 580–95.
- Baldwin, Robert E., and R. S. Hilton. 1984. “A Technique for Indicating Comparative Costs and Predicting Changes in Trade Ratios.” *Review of Economics and Statistics*, 64: 105–10.
- Ballistreri, Edward J., Russell H. Hillberry, and Thomas F. Rutherford. 2010. “Trade and Welfare: Does Industrial Organization Matter?” *Economics Letters* 109(2): 85–87.
- Barro, Robert J. 1991. “Economic Growth in a Cross-Section of Countries.” *Quarterly Journal of Economics* 106: 407–43.
- Barro, Robert J., and Xavier Sala-i-Martin. 1991. “Convergence Across States and Regions.” *Brookings Papers on Economic Activity*, 1: 107–58.
- . 1992. “Convergence.” *Journal of Political Economy* 100(2): 223–51.
- Bartelsman, Eric J., and Wayne Gray. 1996. “The NBER Manufacturing Productivity Database.” NBER Technical Working Paper No. 205.
- Bas, Maria, Thierry Mayer, and Mathias Thoenig. 2014. “From Micro to Macro: Demand and Supply-side Determinants of the Trade Elasticity.” <https://sites.google.com/site/thierry-mayer/research> (accessed 11/7/2014).
- Basco, Sergi, and Martí Mestieri. 2013. “Heterogeneous Trade Costs and Wage Inequality: A Model of Two Globalizations.” *Journal of International Economics* 89(2): 393–406.
- Behrens, Kristian, Andreas Lamorgese, Gianmarco I. P. Ottaviano, and Takatoshi Tabuchi. 2004. “Testing the ‘Home Market Effect’ in a Multi-Country World” A Theory-Based Approach.” CEPR discussion paper no. 4468.
- Behrens, Kristian, and Yasusada Murata. 2007. “General Equilibrium Models of Monopolistic Competition: A New Approach.” *Journal of Economic Theory* 13: 776–87.
- . 2012. “Globalization and Individual Gains from Trade.” *Journal of Monetary Economics* 59: 703–20.
- Ben-David, Dan. 1993. “Trade Liberalization and Income Convergence.” *Quarterly Journal of Economics* 108(3): 653–79.
- . 1998. “Convergence Clubs and Subsistence Economies.” *Journal of Development Economics* 55(1): 155–71.
- . 2001. “Trade Liberalization and Income Convergence: A Comment.” *Journal of International Economics* 55(1): 229–34.
- Benkard, C. Lanier. 2000. “Learning and Forgetting: The Dynamics of Aircraft Production.” *American Economic Review* 90: 1034–54.
- . 2004. “A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft.” *Review of Economic Studies* 71: 581–611.
- Bergin, Paul R., and Robert C. Feenstra. 2000. “Staggered Price Setting and Endogenous Persistence.” *Journal of Monetary Economics* 45: 657–80.
- . 2001. “Pricing to Market, Staggered Contracts, and Real Exchange Rate Persistence.” *Journal of International Economics* 54(2): 333–59.

- . 2009. "Pass-through of Exchange Rates and Competition Between Fixers and Floaters." *Journal of Money, Credit, and Banking* 41(s1): 35–70.
- Bergsten, C. Fred, Kimberly Ann Elliott, Jeffrey J. Schott, and Wendy E. Takacs. 1987. "Auction Quotas and United States Trade Policy." Peterson Institute for International Economics, Washington, D.C.
- Bergstrand, Jeffrey H. 1985. "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade." *Review of Economics and Statistics* 71(1): 143–53.
- Bergstrand, Jeffrey H. 1989. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Review of Economics and Statistics* 67(3): 474–81.
- Berman, Eli, John Bound, and Zvi Griliches. 1994. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics* 104: 367–98.
- Berman, Eli, John Bound, and Stephen Machin. 1998. "Implications of Skill-Biased Technological Change: International Evidence." *Quarterly Journal of Economics* 113: 1245–80.
- Bernanke, Ben S., and Refet S. Gürkaynak. 2001. "Is Growth Exogenous? Taking Mankiw, Romer and Weil Seriously." Pages 11–57 in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual, 2001*. Cambridge: MIT Press.
- Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *American Economic Review* 93(4): 1268–90.
- Bernard, Andrew B., and J. Bradford Jensen. 1997. "Exporters, Skill Upgrading, and the Wage Gap." *Journal of International Economics* 42(1/2): 3–32.
- . 2000. "Understanding Increasing and Decreasing Wage Inequality." Pages 227–68 in Robert C. Feenstra, ed., *The Impact of International Trade on Wages*. Chicago: University of Chicago Press, and NBER.
- Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007. "Firms in International Trade." *Journal of Economic Perspectives* 21(3): 105–30.
- Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2007. "Comparative Advantage and Heterogeneous Firms." *Review of Economic Studies* 74: 31–66.
- . 2011. "Multi-Product Firms and Trade Liberalization." *Quarterly Journal of Economics* 126(3): 1271–318.
- Bernheim, B. Douglas, and Michael D. Whinston. 1986. "Menu Auctions, Resource Allocation, and Economic Influence." *Quarterly Journal of Economics* 101: 1–31.
- Bernhofen, Daniel M., and John C. Brown. 2004a. "A Direct Test of the Theory of Comparative Advantage: The Case of Japan." *Journal of Political Economy* 112(1): 48–67.
- . 2004b. "An Empirical Assessment of the Comparative Advantage Gains from Trade: Evidence from Japan." *The American Economic Review* 95(1): 208–25.
- Bernstein, Jeffrey R., and David E. Weinstein. 2002. "Do Endowments Predict the Location of Production? Evidence from National and International Data." *Journal of International Economics* 56(1): 55–76.
- Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *Rand Journal of Economics* 25(2): 242–46.
- Berry, Steven T., James A. Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63: 841–90.
- . 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review* 89(3): 400–430.
- Bhagwati, Jagdish N. 1958. "Immiserizing Growth: A Geometrical Note." *Review of Economic Studies* 25: 201–5.
- . 1965. "On the Equivalence of Tariffs and Quotas." Pages 53–67 in Robert E. Baldwin et al., eds., *Trade Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler*. Chicago: Rand-McNally.
- . 1972. "The Heckscher-Ohlin Theorem in the Multi-Commodity Case." *Journal of Political Economy* 80(5): 1052–55.

- . 1987. "Quid Pro Quo DFI and VIEs: A Political-Economy-Theoretic Analysis." *International Economic Journal* 1: 1–14.
- . 1993. Regionalism and Multilateralism: An Overview." In Jaime De Melo and Arvind Panagariya, eds., *New Dimensions in Regional Integration*. Cambridge: Cambridge University Press.
- . 2002. *Free Trade Today*. Princeton: Princeton University Press.
- Bhagwati, Jagdish N., Richard A. Brecher, Elias Dinopoulos, and T. N. Srinivasan. 1987. "Quid Pro Quo Foreign Investment and Welfare: A Political-Economy-Theoretic Model." *Journal of Development Economics* 27: 127–38.
- Bhagwati, Jagdish, and Marvin H. Kosters, eds. 1994. *Trade and Wages: Leveling Wages Down?* Washington, D.C.: American Enterprise Institute.
- Bhagwati, Jagdish N., and T. N. Srinivasan. 1980. "Revenue Seeking: A Generalization of the Theory of Tariffs." *Journal of Political Economy* 88(6): 1069–87.
- Bilbiie, Florin, Fabio Ghironi, and Marc J. Melitz. 2012. "Endogenous Entry, Product Variety, and Business Cycles." *Journal of Political Economy* 120: 304–45.
- Blomstrom, Magnus. 1986. "Foreign Investment and Productive Efficiency: The Case of Mexico." *Journal of Industrial Economics* 35: 97–110.
- Blomstrom, Magnus, and A. Kokko. 1998. "Multinational Corporations and Spillovers." *Journal of Economic Surveys*, 12(2): 247–77.
- Blomstrom, Magnus, and H. Persson. 1983. "Foreign Investment and Spillover Efficiency in an Underdeveloped Economy: Evidence from Mexican Manufacturing Industry." *World Development* 11(6): 493–501.
- Blomstrom, Magnus, and F. Sjoholm. 1999. "Technology Transfer and Spillovers: Does Local Participation with Multinationals Matter." *European Economic Review* 43: 915–23.
- Blonigen, Bruce. 1997. "Firm-Specific Assets and the Link between Exchange Rates and Foreign Direct Investment." *American Economic Review* 87(3): 447–66.
- . 2001. "In Search of Substitution between Foreign Production and Exports." *Journal of International Economics* 53(1): 81–104.
- . 2002. "Tariff Jumping Antidumping Duties." *Journal of International Economics* 57(1): 31–50.
- Blonigen, Bruce A., Ronald B. Davies, and Keith Head. 2003. "Estimating the Knowledge-Capital Model of the Multinational Enterprise: Comment." *American Economic Review* 93(3): 980–94.
- Blonigen, Bruce A., and Stephen E. Haynes. 2002. "Antidumping Investigations and the Pass-through of Antidumping Duties and Exchange Rates." *American Economic Review* 92(4): 1044–61.
- Blonigen, Bruce A., and KaSaundra Tomlin. 2001. "Size and Growth of Japanese Plants in the United States." *International Journal of Industrial Organization* 19: 931–52.
- Blum, Bernardo S., Sebastian Claro, and Ignatius Horstmann. 2010. "Intermediation and the Nature of Trade Costs: Theory and Evidence." University of Toronto, manuscript.
- Borcherding, Thomas E., and Eugene Silberberg. 1978. "Shipping the Good Apples Out: The Alchian and Allen Theorem Reconsidered." *Journal of Political Economy* 86(1): 131–38.
- Borjas, George J., Richard B. Freeman, and Lawrence F. Katz. 1997. "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity* 1: 1–90.
- Bowen, Harry P., Edward E. Leamer, and Leo Sveikauskas. 1987. "Multicountry, Multifactor Tests of the Factor Abundance Theory." *American Economic Review* 77(5): 791–809.
- Bowen, Harry P., and Leo Sveikauskas. 1992. "Judging Factor Abundance." *Quarterly Journal of Economics* 107(2): 599–620.
- Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade." *American Economic Review* 87(4): 520–44.
- Brainard, S. Lael, and David Martimort. 1997. "Strategic Trade Policy with Incompletely Informed Policymakers." *Journal of International Economics* 42(1–2): 33–65.

- Brander, James A. 1981. "Intra-industry Trade in Identical Commodities." *Journal of International Economics* 11: 1–14.
- Brander, James A., and Paul R. Krugman. 1983. "A Reciprocal Dumping Model of International Trade." *Journal of International Economics* 15: 313–23. Reprinted in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- Brander, James A., and Barbara Spencer. 1984a. "Trade Warfare: Tariffs and Cartels." *Journal of International Economics* 16: 227–42.
- . 1984b. "Tariff Protection and Imperfect Competition." In Henryk Kierzkowski, ed., *Monopolistic Competition and International Trade*. Oxford: Oxford University Press. Reprinted in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- . 1985. "Export Subsidies and International Market Share Rivalry." *Journal of International Economics* 16: 83–100.
- Branstetter, Lee. 2000. "Is Foreign Direct Investment a Channel of Knowledge Spillovers? Evidence from Japan's FDI in the United States." NBER Working Paper No. 8015.
- . 2001. "Are Knowledge Spillovers International or Intranational in Scope? Microeconometric Evidence from Japan and the United States." *Journal of International Economics* 53(1): 53–79.
- Branstetter, Lee, and Robert Feenstra. 2002. "Trade and Foreign Direct Investment in China: A Political Economy Approach." *Journal of International Economics* 58(2): 335–58.
- Brecher, Richard A., and Ehsan U. Choudhri. 1982a. "The Leontief Paradox, Continued." *Journal of Political Economy* 90(4): 820–23.
- . 1982b. "The Factor Content of International Trade without Factor Price Equalization." *Journal of International Economics* 12(3/4): 276–83.
- Brecher, Richard, and Carlos Diaz-Alejandro. 1977. "Tariffs, Foreign Capital and Immiserizing Growth." *Journal of International Economics* 7: 317–22.
- Bresnahan, Timothy F. 1981. "Departures from Marginal Cost Pricing in the American Automobile Industry." *Journal of Econometrics* 17: 201–27.
- Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics* 121(2): 541–85.
- Broda, Christian, David E. Weinstein, and Nuno Limao. 2008. "Optimal Tariffs and Market Power: The Evidence." *American Economic Review* 98(5): 2032–65.
- Burgess, David F. 1974a. "A Cost Minimization Approach to Import Demand Equations." *Review of Economics and Statistics* 56: 225–34.
- . 1974b. "Production Theory and the Derived Demand for Imports." *Journal of International Economics* 4: 103–17.
- . 1976. "Tariffs and Income Distribution: Some Empirical Evidence for the United States." *Journal of Political Economy* 84(1): 17–45.
- Burstein, Ariel, and Gita Gopinath. 2014. "International Price and Exchange Rates." Pages 391–451 in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Trade, Volume 4*. Amsterdam: Elsevier.
- Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82(1): 1–44.
- Caliendo, Lorenzo, and Esteban Rossi-Hansberg. 2012. "The Impact of Trade on Organization and Productivity." *Quarterly Journal of Economics* 127(3): 1393–467.
- Caliendo, Lorenzo, Esteban Rossi-Hansberg, and Ferdinando Monte. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy* 123(4): 809–52.
- Campa, José, and Linda S. Goldberg. 2005. "Exchange Rate Pass-Through into Import Prices." *Review of Economics and Statistics* 87(4): 679–90.
- Card, David, and John E. DiNardo. 2002. "Skill-biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20(4): 733–83.
- Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13: 185–213.

- Carmichael, Calum M. 1987. "The Control of Export Credit Subsidies and Its Welfare Consequences." *Journal of International Economics* 23(1/2): 1–19.
- Caron, Justin, Thibault Fally, and James R. Markusen. 2014. "International Trade Puzzles: A Solution Linking Production and Preferences." *Quarterly Journal of Economics* 129(3): 1501–52.
- Carr, David L., James R. Markusen, and Keith E. Maskus. 2001. "Estimating the Knowledge-Capital Model of the Multinational Enterprise." *American Economic Review* 91(3): 693–708.
- Casella, Alessandra, and James E. Rauch. 2002. "Anonymous Market and Group Ties in International Trade." *Journal of International Economics* 58(1): 19–48.
- Caves, D. W., L. R. Christensen, and W. E. Diewert. 1982. "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity." *Econometrica* 50: 1393–414.
- Chamberlin, Edward. 1936. *The Theory of Monopolistic Competition: A Reorientation of the Theory of Value*. Cambridge: Harvard University Press.
- Chaney, Thomas. 2008. "Distorted Gravity: the Intensive and Extensive Margins of International Trade." *American Economic Review* 98(4): 1707–21.
- . 2014. "The Network Structure of International Trade." *American Economic Review* 104(11): 3600–3634.
- Chaney, Thomas, and Ralph Ossa. 2012. "Market Size, Division of Labor and Firm Productivity." *Journal of International Economics* 90: 177–80.
- Chang, Won, and L. Alan Winters. 2002. "How Regional Blocs Affect Excluded Countries: The Price Effects of MERCOSUR." *American Economic Review* 92(4): 889–904.
- Chen, Yongmin, and Robert C. Feenstra. 2008. "Buyer Investment, Export Variety, and Intrafirm Trade." *European Economic Review* 52(8): 1313–37.
- Cheng, Leonard K., and Henryk Kierzkowski, eds. 2001. *Global Production and Trade in East Asia*, Kluwer Academic Publishers, Boston, 2001.
- Chipman, John S. 1987. "Compensation Principle." Pages 524–31 in John Eatwell, Murray Milgate, and Peter Newman, eds. *The New Palgrave: A Dictionary of Economics*, vol. 1. London: MacMillan, and New York: Stockton.
- Choi, Yong-Seok, and Pravin Krishna. 2004. "The Factor Content of Bilateral Trade: An Empirical Test." *Journal of Political Economy* 112(4): 887–914.
- Christensen, L. R., D. W. Jorgenson, and L. J. Lau. 1971. "Conjugate Duality and the Transcendental Logarithmic Production Function." *Econometrica* 39: 255–56.
- Clark, Gregory, and Robert C. Feenstra. 2003. "Technology in the Great Divergence." Pages 277–322 in Michael D. Bordo, Alan M. Taylor, and Jeffrey G. Williamson, eds., *Globalization in Historical Perspective*. University of Chicago Press and NBER.
- Clausing, Kimberley A. 2001. "Trade Creation and Trade Diversion in the Canada-U.S. Free Trade Agreement." *Canadian Journal of Economics* 34(3): 677–96.
- Coase, Ronald. 1937. "The Nature of the Firm." *Economica* (November): 386–405.
- Coe, David T., and Elhanan Helpman. 1995. "International R&D Spillovers." *European Economic Review* 39(5): 859–87.
- Coe, David T., Elhanan Helpman, and Alexander W. Hoffmaister. 1997. "North-South R&D Spillovers." *Economic Journal* 107: 134–49.
- Collins, Susan M., ed. 1998. *Imports, Exports, and the American Worker*. Washington, D.C.: Brookings Institution Press.
- Corden, W. Max. 1974. *Trade Policy and Economic Welfare*. Oxford: Clarendon Press.
- . 1984. "Normative Theory of International Trade." Pages 63–130 in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, vol. 1. Amsterdam: Elsevier.
- Corden, W. Max, and J. Peter Neary. 1982. "Booming Sector and Deindustrialization in a Small Open Economy." *Economic Journal* 82: 825–48.
- Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer. 2012. "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies* 79: 581–608.
- Costinot, Arnaud, and Andrés Rodríguez-Clare, 2014 "Trade Theory with Numbers: Quantifying the Consequences of Globalization." Pages 197–262 in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, vol. 4. Amsterdam: Elsevier.

- Costinot, Arnaud, and Jonathan Vogel. 2010. "Matching and Inequality in the World Economy." *Journal of Political Economy* 118(4): 747–86.
- . 2015. "Beyond Ricardo: Assignment Models in International Trade." *Annual Review of Economics* 7: 31–62.
- Costinot, Arnaud, Jonathan Vogel, and Su Wang. 2013. "An Elementary Theory of Global Supply Chains." *Review of Economic Studies* 80: 109–44.
- Cox, David, and Richard G. Harris. 1985. "Trade Liberalization and Industrial Organization: Some Estimates for Canada." *Journal of Political Economy* 93(1): 115–45.
- . 1986. "A Quantitative Assessment of the Economic Impact on Canada of Sectoral Free Trade with the United States." *Canadian Journal of Economics* 19(3): 377–94.
- Cragg, M., and M. Epelbaum. 1996. "Why Has Wage Dispersion Grown in Mexico? Is it the Incidence of Reforms or the Growing Demand for Skills?" *Journal of Development Economics* 51: 99–116.
- Crinò, Rosario. 2008. "Service Offshoring and Productivity in Western Europe." *Economics Bulletin* 6(35): 1–8.
- . 2010. "Service Offshoring and White-Collar Employment." *Review of Economic Studies* 77(2): 595–632.
- . 2013. "Service Offshoring and Labor Demand in Europe." Pages 41–97 in Ashok Bhadran, Dwight M. Jaffee, and Cynthia A. Kroll, eds., *Oxford Handbook of Global Employment and Offshoring*. Oxford: Oxford University Press.
- Crozet, Matthieu, Keith Head, and Thierry Mayer. 2012. "Quality Sorting and Trade: Firm-level Evidence for French Wine." *Review of Economic Studies* 79: 609–44.
- Cuñat, Alejandro, and Marco Maffezzoli. 2004. "Neoclassical Growth and Commodity Trade." *Review of Economic Dynamics* 7(3): 707–36.
- Das, Satya P., and Shabtai Donnenfeld. 1987. "Trade Policy and Its Impact on the Quality of Imports: A Welfare Analysis." *Journal of International Economics* 23(1/2): 77–96.
- . 1989. "Oligopolistic Competition and International Trade: Quantity and Quality Restrictions." *Journal of International Economics* 27(3/4): 299–318.
- Dasgupta, P., and Joseph Stiglitz. 1988. "Learning-by-Doing, Market Structure, and Industrial and Trade Policies." *Oxford Economic Papers* 40: 246–68.
- Davidson, Carl, and Steven J. Matusz. 2006. "Trade Liberalization and Compensation." *International Economic Review* 47(3): 723–47.
- Davis, Donald R. 1995. "Intra-Industry Trade: A Heckscher-Ohlin-Ricardo Approach." 39(3/4): 201–26.
- . 1998. "The Home Market, Trade and Industrial Structure." *American Economic Review* 88(5): 1264–76.
- Davis, Donald R., and James Harrigan. 2011. "Good Jobs, Bad Jobs, and Trade Liberalization." *Journal of International Economics* 84(1): 26–36.
- Davis, Donald R., and David Weinstein. 1996. "Does Economic Geography Matter for International Specialization?" NBER Working Paper No. 5706.
- . 1999. "Economic Geography and Regional Production Structure: An Empirical Investigation." *European Economic Review* 43: 379–407.
- . 2001. "An Account of Global Factor Trade, *American Economic Review* 91(5): 1423–53.
- . 2002a. "Bones, Bombs and Break Points: The Geography of Economic Activity." *American Economic Review* 92(5): 1269–89.
- . 2002b. "Technological Superiority and the Losses from Migration." NBER Working Paper No. 8971.
- . 2003. "The Factor Content of Trade." In Kwan Choi and James Harrigan, eds., *Handbook of International Trade, Volume 1*. Oxford: Blackwell, 119–46.
- Deardorff, Alan V. 1979. "Weak Links in the Chain of Comparative Advantage." *Journal of International Economics* 9(2): 197–209.
- . 1984a, "Testing Trade Theories and Predicting Trade Flows." Pages 467–517 in Ronald Jones and Peter Kenen, eds., *Handbook of International Economics*, vol. 1. Amsterdam and New York: North Holland.

- . 1984b. "The General Validity of the Law of Comparative Advantage" *Journal of Political Economy* 88(5): 941–57. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 1987. "Safeguards Policy and the Conservative Social Welfare Function." Pages 22–40 in Henryk Kierzkowski, ed., *Protection and Competition in International Trade: Essays in Honor of W.M. Corden*. Oxford: Blackwell
- . 2000. "Factor Prices and the Factor Content of Trade Revisited: What's the Use?" *Journal of International Economics* 50: 73–90.
- Debaere, Peter. 2003. "Relative Factor Abundance and Trade." *Journal of Political Economy* 111(3): 589–610.
- Debaere, Peter, and Ufuk Demiroğlu. 2003. "On the Similarity of Country Endowments." *Journal of International Economics* 59(1): 101–36.
- Debaere, Peter, and Hongshik Lee. 2002. "The Real-Side Determinants of Countries' Terms of Trade: A Panel Data Analysis." University of Texas at Austin, manuscript.
- . 2003. "The Real-Side Determinants of Countries' Terms of Trade: A Panel Data Analysis." <http://faculty.darden.virginia.edu/debaere/pdf/termsWP.pdf> (accessed 11/30/2014).
- De Blas, Beatriz, and Kathryn Niles Russ. 2015. "Understanding Markups in the Open Economy." *American Economic Journals—Macroeconomics* 7(2): 157–80.
- Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global Rebalancing with Gravity: Measuring the Burden of Adjustment." *IMF Staff Papers* 55(3): 511–40.
- Dell'mour, R., P. Egger, K. Gugler, and M. Pfaffermayr. 2000. "Outsourcing of Austrian Manufacturing to Eastern European Countries: Effects on Productivity and the Labor Market." In Sven Arndt, H. Handler, and D. Salvatore, eds., *Fragmentation of the Value Added Chain*. Vienna: Austrian Ministry for Economic Affairs and Labour.
- De Loecker, Jan. 2009. "Recovering Markups from Production Data." *International Journal of Industrial Organization* 29: 350–55.
- . 2011. "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity." *Econometrica* 79(5): 1407–51.
- De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2012, revised 2014. "Prices, Markups and Trade Reform." NBER Working Paper No. 17925.
- De Loecker, Jan, and Frank Warzynski. 2012. "Markups and Firm-level Export Status." *American Economic Review* 102(6): 2437–71.
- Destler, I. M. 2005. *American Trade Politics*. Washington, D.C.: Peterson Institute for International Economics.
- Dhingra, Swati. 2013. "Trading Away Wide Brands for Cheap Brands." *American Economic Review* 103(6): 2554–84.
- Dhingra, Swati, and John Morrow. 2013. "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity." London School of Economics.
- Dicken, Peter, Philip F. Kelley, Kris Olds, and Henry Wai-Chung Yeung. 2001. "Chains and Networks, Territories and Scales: Towards a Relational Framework for Analyzing the Global Economy." *Global Networks*, 1(2): 99–123.
- Diewert, W. Erwin. 1974. "Applications of Duality Theory." Pages 106–71 in M. Intriligator and D. Kendrick, eds., *Frontiers of Quantitative Economics*, Volume II. Amsterdam: North-Holland.
- . 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4: 115–46.
- . 1983. "The Theory of the Output Price Index and the Measurement of Real Output Change." Pages 1049–113 in W. E. Diewert and C. Montmarquette, eds., *Price Level Measurement*. Ottawa: Statistics Canada.
- Diewert, W. Erwin, and Catherine J. Morrison. 1986. "Adjusting Outputs and Productivity Indexes for Changes in the Terms of Trade." *Economic Journal* 96: 659–79.
- Diewert, W. Erwin, Arja H. Turunen-Red, and Alan D. Woodland. 1989. "Productivity and Pareto Improving Changes in Taxes and Tariffs." *Review of Economic Studies* 56: 199–216.
- Dikhanov, Yuri. 1999. "A Critique of CIA Estimates of Soviet Performance from the Gerschenkron Perspective." Pages 271–76 in Alan Heston and Robert E. Lipsey, eds., *International and*

- Interarea Comparisons of Income, Output, and Prices.* NBER Studies in Income and Wealth, no. 61. Chicago: University of Chicago Press and NBER.
- Dinopoulos, Elias. 1989. "Quid Pro Quo Foreign Investment." *Economics and Politics*, 1: 145–60.
- . 1992. "Quid Pro Quo Foreign Investment and VERs: A Nash Bargaining Approach." *Economics and Politics* 4: 43–60.
- Dinopoulos, Elias, and Mordechai E. Kreinin. 1988. "Effects on the U.S.-Japan Auto VER on European Prices and on U.S. Welfare." *Review of Economics and Statistics*, 70(3): 484–91.
- Dinopoulos, Elias, and Paul Segerstrom. 1999a. "The Dynamic Effects of Contingent Tariffs." *Journal of International Economics* 47: 191–222.
- . 1999b. "A Schumpeterian Model of Protection and Real Wages." *American Economic Review* 89(3): 450–72.
- Dinopoulos, Elias, and Constantinos Syropoulos. 2001. "Globalization and Scale-Invariant Growth." University of Florida and Florida International University, manuscript.
- Dixit, Avinash. 1986. "Gains from Trade without Lump Sum Compensation." *Journal of International Economics* 21: 111–22.
- Dixit, Avinash K., and A. S. Kyle. 1985. "The Use of Protection and Subsidies for Entry Promotion and Deterrence." *American Economic Review* 75: 139–52.
- Dixit, Avinash, and Victor Norman. 1980. *Theory of International Trade*. Cambridge: Cambridge University Press.
- Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67(3): 297–308.
- Dollar, David. 1992. "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976–1985." *Economic Development and Cultural Change*, 40(3): 523–44.
- Dollar, David, and Aart Kraay. 2002. "Growth Is Good for the Poor." *Journal of Economic Growth* 7: 195–225.
- . 2004. "Trade, Growth, and Poverty." *The Economic Journal* 114(493): F29–F49.
- Dollar, David, Edward N. Wolff, and William Baumol. 1988. "The Factor Price Equalization Model and Industry Labor Productivity: An Empirical Test." Pages 23–47 in Robert C. Feenstra, ed., *Empirical Methods for International Trade*, Cambridge: MIT Press.
- Doms, Mark E., and J. Bradford Jensen. 1998. "Comparing Wages, Skills, and Productivity between Domestic and Foreign-Owned Manufacturing Establishments in the United States." In Robert E. Baldwin, Robert E. Lipsey, and J. David Richardson, eds., *Geography and Ownership as Bases for Accounting*. Chicago: University of Chicago Press and NBER.
- Dornbusch, Rüdiger. 1987. "Exchange Rates and Prices." *American Economic Review* 77(1): 93–106.
- Dornbusch, Rüdiger, Stanley Fischer, and Paul A. Samuelson. 1977. "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods." *American Economic Review* 67(5): 823–39.
- . 1980. "Heckscher-Ohlin Trade Theory with a Continuum of Goods." *Quarterly Journal of Economics* 95(2): 203–24.
- Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52(2): 345–62.
- Dunning, John H. 1977. "Trade, Location of Economic Activity and the MNE: A Search for an Eclectic Approach." Pages 395–418 in B. Ohlin, P. O. Hesselborn, and P. M. Wijkman, eds., *The International Allocation of Economic Activity*. London: Macmillan.
- . 1981. *International Production and the Multinational Enterprise*. London: George Allen and Unwin.
- Dutt, Pushan, and Devashish Mitra. 2002. "Endogenous Trade Policy through Majority Voting." *Journal of International Economics* 58(1): 107–34.
- Easterly, William, and Ross Levine. 2001. "What Have We Learned From a Decade of Empirical Research on Growth? It's Not Factor Accumulation: Stylized Facts and Growth Models." *World Bank Economic Review* 15(2): 177–219.

- Eastman, H. C., and S. Stykolt. 1967. *The Tariff and Competition in Canada*. Toronto: Macmillan.
- Eaton, Jonathan, and Gene M. Grossman. 1986. "Optimal Trade and Industrial Policy under Oligopoly." *Quarterly Journal of Economics* 101(2): 383–406. Reprinted in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography and Trade." *Econometrica* 70(5): 1741–780.
- Eaton, Jonathan, Samuel S. Kortum, and Sebastian Sotelo. 2012. "International Trade: Linking Micro and Macro." NBER Working Paper No. 17864.
- Ebenstein, Avraham, Ann Harrison, Margaret McMillan, and Shannon Phillips. 2014. "Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Survey." *Review of Economics and Statistics* 96(4): 581–97.
- Eckel, Carsten, and J. Peter Neary. 2009. "Multi-product Firms and Flexible Manufacturing in the Global Economy" *Review of Economic Studies* 77(1): 188–217.
- Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2012. "Competition, Markups, and the Gains from International Trade." NBER Working Paper No. 18041.
- Edwards, Sebastian. 1998. "Openness, Productivity and Growth: What Do We Really Know?" *Economic Journal* 108: 383–98.
- Engel, Charles, and John H. Rogers. 1996. "How Wide Is the Border?" *American Economic Review* 86(5): 1112–25. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- Estevadeordal, Antoni, and Alan M. Taylor. 2002a. "A Century of Missing Trade?" *American Economic Review* 92(1): 383–93.
- . 2002b. "Testing Trade Theory in Ohlin's Time." In Ronald Findlay, Lars Jonung, and Mats Lundahl, eds., *Bertil Ohlin: A Centennial Celebration, 1899–1999*. Cambridge: MIT Press.
- Estevadeordal, Antoni, Brian Frantz, and Alan M. Taylor. 2003. "The Rise and Fall of World Trade, 1870–1939." *Quarterly Journal of Economics* 118(2): 359–407.
- Ethier, Wilfred. 1974. "Some of the Theorems of International Trade with Many Good and Factors." *Journal of International Economics* 4: 199–206.
- . 1979. "Internationally Decreasing Costs and World Trade." *Journal of International Economics* 9: 1–24.
- . 1982a. "Decreasing Costs in International Trade and Frank Graham's Argument for Protection." *Econometrica* 50(5): 1243–68.
- . 1982b. "Dumping." *Journal of Political Economy* 90(3): 487–506.
- . 1984. "Higher Dimensional Issues in Trade Theory." Pages 131–84 in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, vol. 1. Amsterdam and New York: North Holland.
- . 1986. "The Multinational Firm." *Quarterly Journal of Economics* 101: 805–33.
- . 1998. "Regionalism in a Multilateral World." *Journal of Political Economy* 106: 1214–45.
- Ethier, Wilfred, and James R. Markusen. 1996. "Multinational Firms, Technology Diffusion, and Trade." *Journal of International Economics* 41: 1–28.
- Fabinger, Michal, and Glen E. Weyl. 2014. "A Tractable Approach to Pass-Through Patterns with Applications to International Trade." <http://ssrn.com/abstract=2194855> or <http://dx.doi.org/10.2139/ssrn.2194855> (accessed 10/22/2014).
- Facchini, Giovanni, and Gerald Willmann. 1999. "The Gains from Duty Free Zones." *Journal of International Economics* 49(2): 403–12.
- Fajgelbaum, Pablo D., and Amit K. Khandelwal. 2014. "Measuring the Unequal Gains from Trade." NBER Working Paper No. 20331.
- Fajgelbaum, Pablo D., Gene M. Grossman, and Elhanan Helpman. 2011a. "Income Distribution, Product Quality, and International Trade." *Journal of Political Economy* 119(4): 721–65
- . 2011b. "A Linder Hypothesis for Foreign Direct Investment." NBER Working Paper No. 17550.
- Fally, Thibault. 2014. "Structural Gravity and Fixed Effects." University of California, Berkeley. http://are.berkeley.edu/~fally/Papers/Poisson_JIE_Sept_2014.pdf (accessed 10/16/2014).

- Fally, Thibault, and Russell Hillberry. 2014. "A Coasian Model of International Production Chains." University of California-Berkeley, Department of Agricultural and Resource Economics, and World Bank, manuscript.
- Feenstra, Robert C. 1984. "Voluntary Export Restraints in U.S. Autos, 1980–81: Quality, Employment and Welfare Effects" Pages 35–59 in R. E. Baldwin and A. O. Krueger, eds., *The Structure and Evolution of Recent U.S. Trade Policy*. Chicago: University of Chicago Press, and NBER.
- . 1985. "Automobile Prices and Protection: The U.S.-Japan Trade Restraint." *Journal of Policy Modeling*, 7(1): 49–68.
- . 1986. "Trade Policy with Several Goods and 'Market Linkages.'" *Journal of International Economics* 20: 249–67.
- . 1988a "Quality Change under Trade Restraints in Japanese Autos." *Quarterly Journal of Economics* 103(1): 131–46.
- . 1988b. "Gains from Trade in Differentiated Products: Japanese Compact Trucks." Pages 119–136 in Robert C. Feenstra, ed., *Empirical Methods for International Trade*. Cambridge: MIT Press.
- . 1989. "Symmetric Pass-Through of Tariffs and Exchange Rates under Imperfect Competition: An Empirical Test." *Journal of International Economics* 27(1/2): 25–45.
- . 1992. "How Costly Is Protectionism?" *Journal of Economic Perspectives* 6: 159–78.
- . 1994. "New Product Varieties and the Measurement of International Prices." *American Economic Review* 84(A1): 157–77.
- . 1995. "Exact Hedonic Price Indexes." *Review of Economics and Statistics* 78(4): 634–53.
- . 1996. "Trade and Uneven Growth." *Journal of Development Economics* 49: 229–56.
- . 1998. "Integration and Disintegration in the Global Economy." *Journal of Economic Perspectives* (Fall): 31–50.
- . 1999. "Facts and Fallacies about Foreign Direct Investment." Pages 331–50 in Martin Feldstein, ed., *International Capital Flows*. Chicago: University of Chicago Press, and NBER.
- , ed. 2000. *The Impact of International Trade on Wages*. Chicago: University of Chicago Press.
- . 2002. "Border Effects and the Gravity Equation: Consistent Methods for Estimation." *Scottish Journal of Political Economy* 49(5): 491–506.
- . 2003. "A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity." *Economic Letters* 78(1): 79–86.
- . 2010a. *Offshoring in the Global Economy: Theory and Evidence*. Cambridge: MIT Press.
- . 2010b. *Product Variety and the Gains from International Trade*. Cambridge: MIT Press.
- . 2010c. "Measuring the Gains from Trade under Monopolistic Competition." *Canadian Journal of Economics* 43(1): 1–28
- . 2014. "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity." NBER Working Paper No. 19833.
- Feenstra, Robert C., Joseph Gagnon, and Michael Knetter. 1996. "Market Share and Exchange Rate Pass-through in World Automobile Trade." *Journal of International Economics* 40(1/2): 187–208.
- Feenstra, Robert C., Gary G. Hamilton, and Deng-Shing Huang. 2001. "The Organization of the Taiwanese and South Korean Economies: A Comparative Equilibrium Analysis." Pages 86–142 in Alessandra Casella and James E. Rauch, eds., *Networks and Markets*. New York: Russell Sage.
- Feenstra, Robert C., and Gary G. Hamilton. 2006. *Emergent Economies, Divergent Paths: Business Groups and Economic Organization in South Korea and Taiwan*. Cambridge: Cambridge University Press.
- Feenstra, Robert C., and Gordon H. Hanson. 1996. "Foreign Investment, Outsourcing and Relative Wages." Pages 89–127 in R. C. Feenstra, G. M. Grossman, and D. A. Irwin, eds. *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*. Cambridge: MIT Press.
- . 1997. "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras." *Journal of International Economics* 42(3/4): 371–93.

- . 1999. "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the U.S., 1979–1990." *Quarterly Journal of Economics* 114(3): 907–40.
- . 2000. "Aggregation Bias in the Factor Content of Trade" Evidence from U.S. Manufacturing." *American Economic Review Papers and Proceedings* 90(2): 155–60.
- . 2001. "Intermediaries in Entrepôt Trade: Hong Kong Re-Exports of Chinese Goods." NBER Working Paper No. 8088.
- . 2003. "Global Production Sharing and Rising Inequality: A Survey of Trade and Wages." Pages 146–18 in Kwan Choi and James Harrigan, eds., *Handbook of International Trade, Volume 1*. Oxford: Blackwell
- . 2005. "Ownership and Control in Outsourcing to China: Estimating the Property-Rights Theory of the Firm." *Quarterly Journal of Economics* 120(2): 729–62.
- Feenstra, Robert C., Gordon H. Hanson, and Songhua Lin. 2004. "The Value of Information in International Trade: Gains to Outsourcing through Hong Kong." *The B.E. Journals in Economic Analysis and Policy, Advances in Economic Analysis and Policy* 4(1), Article 1. <http://www.bepress.com/bejap/advances/vol4/iss1/> (accessed 11/7/2014).
- Feenstra, Robert C., Alan Heston, Marcel P. Timmer, and Haiyan Deng. 2009. "Estimating Real Production and Expenditures Across Countries: A Proposal for Improving the Penn World Tables." *Review of Economics and Statistics* 91(1): 201–12.
- Feenstra, Robert C., Deng-Shing Huang, and Gary G. Hamilton. 2003. "A Market-Power Based Model of Business Groups." *Journal of Economic Behavior and Organization* 51(4): 459–85.
- Feenstra, Robert C., Robert Inklaar, and Marcel Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105(10): forthcoming.
- Feenstra, Robert C., and J. Bradford Jensen. 2012. "Evaluating Estimates of Materials Offshoring from U.S. Manufacturing." *Economic Letters* 117: 170–73.
- Feenstra, Robert C., and James Levinsohn. 1995. "Estimating Markups and Market Conduct with Multidimensional Product Attributes." *Review of Economics Studies* 62: 19–52.
- Feenstra, Robert C., and Tracy R. Lewis. 1991a. "Distributing the Gains from Trade with Incomplete Information." *Economics and Politics* 3(1): 21–39.
- . 1991b. "Negotiated Trade Restrictions with Private Political Pressure." *Quarterly Journal of Economics* 106: 1287–307.
- . 1994. "Trade Adjustment Assistance and Pareto Gains from Trade." *Journal of International Economics* 36: 201–22.
- Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Kathryn Russ. 2014. "In Search of the Armington Elasticity." NBER Working Paper No. 20063.
- Feenstra, Robert C., and Hong Ma. 2008. "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition." In Elhanan Helpman, Dali Marin, and Thierry Verdier, eds. *The Organization of Firms in a Global Economy*. Cambridge: Harvard University Press, 173–99.
- Feenstra, Robert C., and James R. Markusen. 1995. "Accounting for Growth with New Intermediate Inputs." *International Economic Review* 35(A2): 429–47.
- Feenstra, Robert C., James R. Markusen, and Andrew K. Rose. 1998. "Understanding the Home Market Effect and the Gravity Equation: The Role of Differentiating Goods." NBER Working Paper No. 6804.
- . 2001. "Using the Gravity Equation to Differentiate among Alternative Theories of Trade." *Canadian Journal of Economics* 34(2): 430–47.
- Feenstra, Robert C., John McMillan, and Tracy R. Lewis. 1990. "Designing Policies to Open Trade." *Economics and Politics*, 2(3): 223–40.
- Feenstra, Robert C., and John Romalis. 2014. "International Prices and Endogenous Quality." *Quarterly Journal of Economics* 129(2): 477–528.
- Feenstra, Robert C., and Clinton R. Shiells. 1997. "Bias in U.S. Import Prices and Demand." Pages 249–73 in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods*, NBER Studies in Income and Wealth No. 58. Chicago: University of Chicago Press.
- Feenstra, Robert C., and Alan M. Taylor. 2014. *International Economics*. 3rd edition. New York: Worth.

- Feenstra, Robert C., Tzu-Han Yang, and Gary Hamilton. 1999. "Business Groups and Product Variety in Trade: Evidence from South Korea, Taiwan, and Japan." *Journal of International Economics* 48: 71–100.
- Feenstra, Robert C., and David Weinstein. 2010. "Globalization, Competition, and U.S. Welfare." NBER Working Paper No. 15749.
- Fernandes, Ana, and Heiwei Tang. 2012. "Determinants of Vertical Integration in Export Processing: Theory and Evidence from China." *Journal of Development Economics* 99(2): 396–414.
- Fernandez, Raquel, and Dani Rodrik. 1991. "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty." *American Economic Review* 81(5): 1146–55.
- Fieler, Ana Cecilia. 2011. "Non-Homotheticity and Bilateral Trade: Evidence and a Quantitative Explanation." *Econometrica* 79(4): 1069–101.
- Findlay, Ronald. 1984. "Growth and Development in Trade Models." Pages 325–65 in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, vol. 1. Amsterdam and New York: North Holland.
- . 1996. "Modeling Global Interdependence: Centers. Peripheries, and Frontiers." *American Economic Review: Papers and Proceedings*, 86(2): 47–51.
- Findlay, Ronald, and Harry Grubert. 1959. "Factor Intensities, Technological Progress, and the Terms of Trade." *Oxford Economic Papers*, 1: 111–21.
- Findlay, Ronald, and Stanislaw Wellisz. 1982. "Endogenous Tariffs, the Political Economy of Trade Restrictions and Welfare." Pages 223–34 in Jagdish N. Bhagwati, ed., *Import Competition and Response*. Chicago: University of Chicago Press, and NBER.
- Fisher, Franklin M., and Karl Shell. 1972. *The Economic Theory of Price Indexes*. New York: Academic Press.
- Fisher, Irving. 1922. *The Making of Index Numbers*. Boston: Houghton Mifflin.
- François, Patrick, and Tanguy van Ypersele. 2002. "On the Protection of Cultural Goods." *Journal of International Economics* 56(2): 359–70.
- Frankel, Jeffrey A. 1997. *Regional Trade Blocs in the World Economic System*. Washington, D.C.: Institute for International Economics.
- Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3): 379–99.
- Freeman, Richard B. 1995. "Are Your Wages Set in Beijing?" *Journal of Economic Perspectives*, 9: 15–32.
- Freeman, Richard, and Lawrence Katz. 1994. "Rising Wage Inequality: The United States vs. Other Advanced Countries." In Richard Freeman, ed., *Working under Different Rules*. New York: Russell Sage Foundation.
- Froot, Kenneth A., and Jeremy C. Stein. 1991. "Exchange Rates and Foreign Direct Investment: An Imperfect Capital Markets Approach." *Quarterly Journal of Economics* 106(4): 1191–217.
- Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. *The Spatial Economy*. Cambridge: MIT Press.
- Funk, Mark. 2001. "Trade and International R&D Spillovers among OECD Countries." *Southern Economic Journal* 67(3): 725–36.
- Gabaix, Xavier. 1999. "The Factor Content of Trade: A Rejection of the Heckscher-Ohlin-Leontief Hypothesis." Department of Economics, MIT, manuscript.
- Gallaway, Michael P., Bruce A. Blonigen, and Joseph E. Flynn. 1999. "Welfare Costs of the U.S. Antidumping and Countervailing Duty Laws." *Journal of International Economics* 49(2): 211–44.
- Garett, Stefania. 2012. "Firms' Heterogeneity and Incomplete Pass-Through." http://people.bu.edu/garettos/Garett_pt_12.pdf (accessed 10/25/2014).
- . 2013. "Input Sourcing and Multinational Production." *American Economic Journal: Macroeconomics* 5(2): 118–51.
- Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*: 108 (5): 874–904
- Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics*: 121(4): 1383–435.
- Gawande, Kishore. 1995. "Are U.S. Nontariff Barriers Retaliatory? An Application of Extreme Bounds Analysis in the Tobit Model." *Review of Economics and Statistics* 77: 677–88.

- Gawande, Kishore, and Usree Bandyopadhyay. 2000. "Is Protection for Sale? A Test of the Grossman-Helpman Theory of Endogenous Protection." *Review of Economics and Statistics* (February): 139–52.
- Gawande, Kishore, and Pravin Krishna. 2003. "The Political Economy of Trade Policy: Empirical Approaches." Pages 213–50 in Kwan Choi and James Harrigan, eds., *Handbook of International Trade, Volume 1*. Oxford: Blackwell.
- Geishecker, Ingo. 2004. "Outsourcing and the Relative Demand for Low-Skilled Labour: Exemplary Evidence from German Manufacturing Industries." In Danièle Meulders, Robert Plasman, and Francois Rycx, eds., *Minimum Wages, Low Pay and Unemployment*. New York: Palgrave Macmillan.
- Gerefki, Gary, and Miguel Korzeniewicz, eds. 1994. *Commodity Chains and Global Capitalism*. Westport, Conn.: Praeger.
- Gerschenkron, Alexander. 1951. "A Dollar Index of Soviet Machinery Output, 1927–28 to 1937." Report R-197. Santa Monica, Calif.: Rand.
- Goldberg, Linda S., and Michael W. Klein. 2000. "International Trade and Factor Mobility: An Empirical Investigation." In Guillermo Calvo, Rudiger Dornbusch, and Maurice Obstfeld, eds., *Festschrift in Honor of Robert Mundell*. Cambridge: MIT Press.
- Goldberg, Pinelopi Koujianou. 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry." *Econometrica*, 63: 891–951.
- , ed. 2015. *Trade and Inequality*. The International Library of Critical Writings in Economics. Cheltenham: Edward Elgar.
- Goldberg, Pinelopi Koujianou, Amit Khandelwal, Nina Pavcnik, and Petia Topalova. 2010. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *Quarterly Journal of Economics* 125(4): 1727–67.
- Goldberg, Pinelopi Koujianou, and Michel M. Knetter. 1997. "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature* 35: 1244–72.
- Goldberg, Pinelopi, and Giovanni Maggi. 1999. "Protection for Sale: An Empirical Investigation." *American Economic Review* 89(5): 1135–55.
- Gopinath, Gita, and Brent Neiman. 2014. "Trade Adjustment and Productivity in Large Crises." *American Economic Review* 104(3): 793–831.
- Görg, Holger, A. Hijzen, and R. C. Hine. 2003. "International Fragmentation and Relative Wages in the U.K." IZA Discussion Paper No. 717.
- . 2005. "International Outsourcing and the Skill Structure of Labour Demand in the United Kingdom." *Economic Journal* 115(506): 860–78.
- Gould, David M. 1994. "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows." *Review of Economics and Statistics* 76: 302–16.
- Graham, Frank. 1923. "Some Aspects of Protection Further Considered." *Quarterly Journal of Economics* 37: 199–227.
- Grether, Jean-Marie, Jaime de Melo, and Marcelo Olarrega. 2001. "Who Determines Mexican Trade Policy?" *Journal of Development Economics* 64(2): 343–70.
- Grinols, Earl L. 1981. "An Extension of the Kemp-Wan Theorem on the Formation of Customs Unions." *Journal of International Economics* 11: 259–66.
- . 1984. "The Thorn in the Lion's Paw: Has Britain Paid Too Much for Common Market Membership?" *Journal of International Economics* 16: 271–93.
- . 1991. "Increasing Returns and the Gains from Trade." *International Economics Review* 32(4): 973–84.
- Grinols, Earl L., and Peri Silva. 2007. "An Enhancement of Modern Free Trade Area Theory." *Oxford Economic Papers* 59: 219–25
- Grinols, Earl L., and Kar-yiu Wong. 1991. "An Exact Measure of Welfare Change." *Canadian Journal of Economics* 24(2): 429–49.
- Groizard, José L., Priya Ranjan, and Antonio Rodríguez-López. 2014. "Offshoring and Jobs: The Myriad Channels of Influence." http://www.socsci.uci.edu/~jantonio/Papers/offshoring_and_jobs.pdf (accessed 10/5/2014).

- Gron, Anne, and Deborah Swenson. 1996. "Incomplete Exchange-Rate Pass-Through and Imperfect Competition: The Effect of Local Production," *American Economic Review* 86(2): 71–76.
- Gross, Daniel. 1987. "A Note on Optimal Tariff Retaliation and the Welfare Loss from Trade Wars in a Framework with Intra-Industry Trade." *Journal of International Economics* 23: 357–67.
- Grossman, Gene M. 2004. "The Distribution of Talent and the Pattern and Consequences of International Trade." *Journal of Political Economy*: 112(1) 209–39.
- Grossman, Gene M., and Elhanan Helpman. 1989. "Product Development and International Trade." *Journal of Political Economy* 97(6): 1261–83.
- _____. 1990. "Comparative Advantage and Long Run Growth." *American Economic Review* 80: 796–815.
- _____. 1991. *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.
- _____. 1994. "Protection for Sale." *American Economic Review* 84: 833–50.
- _____. 1995a. "Trade Wars and Trade Talks." *Journal of Political Economy* 103(4): 675–708.
- _____. 1995b. "The Politics of Free-Trade Agreements." *American Economic Review* 85(4): 667–90.
- _____. 1996. "Foreign Investment with Endogenous Protection." Pages 199–224 in Robert C. Feenstra, Gene M. Grossman, and Douglas A. Irwin, eds., *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*. Cambridge: MIT Press.
- _____. 2001. *Special Interest Politics*. Cambridge: MIT Press.
- _____. 2002. "Integration versus Outsourcing in Industry Equilibrium." *Quarterly Journal of Economics* 117(1): 85–120.
- _____. 2004. "Managerial Incentives and the International Organization of Production." *Journal of International Economics* 63(2): 237–62.
- _____. 2005. "Outsourcing in a Global Economy." *Review of Economic Studies* 72(250): 135–59.
- _____. 2008. "Fair Wages and Foreign Sourcing." Pages 273–310 in E. Helpman, D. Marin, and T. Verdier eds., *The Organization of Firms in a Global Economy*. Cambridge: Harvard University Press.
- Grossman, Gene M., and Giovanni Maggi. 2000. "Diversity and Trade." *American Economic Review* 90(5): 1255–75.
- Grossman, Gene M., and Esteban Rossi-Hansberg. 2006. "The Rise of Offshoring: It's Not Wine for Cloth Anymore." In *The New Economic Geography: Effects and Policy Implications*. Federal Reserve Bank of Kansas City. <http://www.kansascityfed.org/Publicat/sympos/2006/PDF/8GrossmanandRossi-Hansberg.pdf> (accessed 10/16/2014).
- _____. 2008. "A Simple Theory of Offshoring." *American Economic Review* 98(5): 1978–97.
- _____. 2012. "Task Trade between Similar Countries." *Econometrica* 80(2): 593–629.
- Grossman, Sanford J., and Oliver D. Hart. 1986. "Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94(4): 691–719.
- Gustafsson, Peter, and Paul Segerstrom. 2010. "Trade Liberalization and Productivity Growth." *Review of International Economics* 18(2): 207–28.
- Haddad, Mona, and Ann E. Harrison. 1993. "Are There Positive Spillovers from Foreign Direct Investment? Evidence from Panel Data from Morocco." *Journal of Development Economics* 42: 51–74.
- Hall, Robert. 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96: 921–947.
- Hallak, Juan Carlos. 2006. "Product Quality and the Direction of Trade." *Journal of International Economics* 68(1): 238–65.
- Hallak, Juan Carlos, and Peter K. Schott. 2011. "Estimating Cross-Country Differences in Product Quality." *Quarterly Journal of Economics* 126(1): 417–74.
- Hamilton, Alexander. 1791. *Report on Manufactures*. Reprinted in U.S. Senate Documents, vol. 22, no. 172. Washington, D.C.: U.S. Congress, 1913.
- Hansen, Wendy, and Thomas J. Prusa. 1995. "The Road Most Taken: The Rise of Title VII Protection." *The World Economy* 18(2): 295–313.

- Hanson, Gordon H. 1997. "Increasing Returns, Trade and the Regional Structure of Wages." *Economic Journal* 107: 113–33.
- . 1998. "Market Potential, Increasing Returns, and Geographic Concentration." NBER Working Paper No.6429.
- Hanson, Gordon, and Ann E. Harrison. 1999. "Trade, Technology, and Wage Inequality." *Industrial and Labor Relations Review* 52(2): 271–88.
- Hanson, Gordon H., Raymond J. Mataloni, Jr., and Matthew J. Slaughter. 2001. "Expansion Strategies of U.S. Multinational Firms." In Dani Rodrik and Susan Collins, eds., *Brookings Trade Forum 2001*. Washington, D.C.: Brookings.
- Hanson, Gordon H., and Chong Xiang. 2004. "The Home-Market Effect and Bilateral Trade Patterns." *American Economic Review* 94(4): 1108–129.
- . 2011. "Trade Barriers and Trade Flows with Product Heterogeneity: An Application to U.S. Motion Picture Exports." *Journal of International Economics* 83(1): 14–26.
- Hardy, G. H., J. E. Littlewood, and G. Polya. 1934. *Inequalities*. London: Cambridge University Press.
- Harper, Michael J., Ernst R. Berndt, and David O. Wood. 1989. "Rates of Return and Capital Aggregation Using Alternative Rental Prices." Pages 331–372 in Dale W. Jorgenson and Ralph Landau, eds., *Technology and Capital Formation*. Cambridge: MIT Press.
- Harrigan, James. 1995. "Factor Endowments and the International Location of Production: Econometric Evidence for the OECD, 1970–1985." *Journal of International Economics* 39(1/2): 123–41.
- . 1996. "Openness to Trade in Manufactures in the OECD." *Journal of International Economics* 40(1/2): 23–39.
- . 1997. "Technology, Factor Supplies and International Specialization: Estimating the Neoclassical Model." *American Economic Review* 87(4): 475–94.
- . 2003. "Specialization and the Volume of Production: Do the Data Obey the Laws?" In Kwan Choi and James Harrigan, eds., *Handbook of International Trade, Volume 1*. Oxford: Blackwell.
- Harrigan, James, and Egon Zakrajsek. 2000. "Factor Supplies and Specialization in the World Economy." NBER Working Paper No. 7848, and *Federal Reserve Bank of New York Staff Report*, no. 107, August.
- Harris, Richard. 1984. "Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition." *American Economic Review* 74(5): 1016–32.
- . 1985. "Why Voluntary Export Restraints Are 'Voluntary'." *Canadian Journal of Economics* 18(4): 799–809.
- Harrison, Ann E. 1994. "Productivity, Imperfect Competition and Trade Reform: Theory and Evidence." 36(1/2): 53–73.
- . 1996. "Openness and Growth: A Time-series, Cross-section Analysis for Developing Countries." *Journal of Development Economics* 48: 419–47.
- Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98: 1119–58.
- . 1999. "Foundations of Incomplete Contracts." *Review of Economics Studies* 66: 115–38.
- Hartigan, James C. 1996. "Predatory Dumping." *Canadian Journal of Economics* 29(1): 228–39.
- Head, Keith C. 1994. "Infant Industry Protection in the Steel Rail Industry." *Journal of International Economics* 37(3/4): 141–66.
- Head, Keith C., and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." Pages 131–96 in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Trade, Volume 4*. Amsterdam: Elsevier.
- Head, Keith, Thierry Mayer, and Mathias Thoenig. 2014. "Welfare and Trade without Pareto," *American Economic Review* 104(5): 310–16.
- Head, Keith C., and John C. Ries. 1998. "Immigration and Trade Creation: Econometric Evidence from Canada." *Journal of International Economics* 31(1): 47–62.
- . 1999. "Rationalization Effects of Tariff Reductions." *Journal of International Economics* 47(2): 295–320.

- . 2000. "Offshore Production and Skill Upgrading by Japanese Manufacturing Firms." *Journal of International Economics* 58(1): 81–106.
- . 2001. "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade." *American Economic Review* 91(4): 858–76.
- Head, Keith C., John Ries, and Barbara Spencer. 2004. "Vertical Networks and U.S. Auto Parts Exports: Is Japan Different?" *Journal of Economics and Management Strategy* 13(1): 37–67.
- Heckscher, Eli, and Bertil Ohlin. 1991. Edited by Harry Flam and M. June Flanders, *Heckscher-Ohlin Trade Theory*. Cambridge: MIT Press.
- Helpman, Elhanan. 1981. "International Trade in the Presence of Product Differentiation, Economics of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach." *Journal of International Economics* 11: 305–40.
- . 1984a. "The Factor Content of Foreign Trade." *Economic Journal* 94: 84–94.
- . 1984b. "Increasing Returns. Imperfect Markets and Trade Theory." Pages 325–365 in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, vol. 1. Amsterdam: North Holland.
- . 1984c. "A Simple Theory of International Trade with Multinational Corporations." *Journal of Political Economy* 92: 451–71.
- . 1985. "Multinational Corporations and Trade Structure." *Review of Economic Studies* 102: 443–57.
- . 1987. "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries." *Journal of the Japanese and International Economies* 1: 62–81.
- . 1990. "Monopolistic Competition in Trade Theory." Special Papers in International Finance, no. 16, International Finance Section, Princeton University.
- . 2006. "Trade, FDI, and the Organization of Firms." *Journal of Economic Literature* 44: 589–630.
- Helpman, Elhanan, and Oleg Itskhoki, 2010. "Labour Market Rigidities, Trade, and Unemployment." *Review of Economic Studies* 77(3): 1100–1137.
- Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding. 2008. "Wages, Unemployment and Inequality with Heterogeneous Firms and Workers." NBER Working Paper No. 14122.
- . 2010. "Inequality and Unemployment in a Global Economy." *Econometrica* 78(4): 1239–83.
- Helpman, Elhanan, Oleg Itskhoki, Stephen Redding, and Marc-Andreas Muendler. 2012. "Trade and Inequality: From Theory to Estimation." NBER Working Paper No. 17991.
- Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge: MIT Press.
- . 1989. *Trade Policy and Market Structure*. Cambridge: MIT Press.
- Helpman, Elhanan, Dalia Marin, and Thierry Verdier, eds. 2008. *The Organization of Firms in a Global Economy*. Cambridge: Harvard University Press.
- Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics* 123(2): 441–87.
- Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review* 94(1): 300–316.
- Hicks, John R. 1940. "The Valuation of Social Income." *Economica* 7: 105–24.
- Hillman, Arye L. 1989. *The Political Economy of Protection*. London: Harwood Academic.
- Holmes, Thomas J., Wen-Tai Hsu, and Sanghoon Lee. 2014. "Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade." *Journal of International Economics* 94: 195–206.
- Holmstrom, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review* 84(4): 972–91.
- Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–50.
- Horn, Henrik, and Lars Persson. 2001. "Equilibrium Ownership of an International Oligopoly." *Journal of International Economics* 53(2): 307–33.
- Hornig, Ellen, Richard N. Boisvert, and David Blandford. 1990. "Quota Rents and Subsidies: The Case of U.S. Cheese Import Quotas." *European Review of Agricultural Economics* 17: 421–34.

- Horstmann, Ignatius J., and James R. Markusen. 1986. "Up the Average Cost Curve: Inefficient Entry and the New Protectionism." *Journal of International Economics* 20: 225–47.
- . 1987. "Licensing versus Direct Investment: A Model of Internalization by the Multinational Enterprise." *Canadian Journal of Economics* 20: 464–81.
- . 1992. "Endogenous Market Structures in International Trade." *Journal of International Economics* 32: 109–29.
- . 1996. "Exploring New Markets: Direct Investment, Contractual Relationships, and the Multinational Enterprise." *International Economic Review* 37: 1–20.
- Hottman, Colin, Stephen J. Redding, and David E. Weinstein. 2014. "What Is 'Firm Heterogeneity' in Trade Models? The Role of Quality, Scope, Markups, and Cost." NBER Working Paper No. 20436.
- Howenstein, Ned. G., and William J. Zeile. 1994. "Characteristics of Foreign-Owned U.S. Establishments." *Survey of Current Business* 74: 34–59.
- Hsieh, Chang-Tai, and Ralph Ossa. 2011. "A Global View of Productivity Growth in China." NBER Working Paper No. 16778.
- Hsieh, Chang-Tai, and Keong T. Woo. 2005. "The Impact of Outsourcing to China on Hong Kong's Labor Market." *American Economic Review* 95(5): 1673–87.
- Hummels, David. 1999. "Towards a Geography of Trade Costs." Purdue University, manuscript.
- Hummels, David, Jun Ishii, and Kei-Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics* 54: 75–96.
- Hummels, David, Rasmus Jorgensen, Jakob Munch, and Chong Xiang. 2014. "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data." *American Economic Review* 104(6): 1597–629.
- Hummels, David, and Peter J. Klenow. 2002. "The Variety and Quality of a Nation's Trade." NBER Working Paper No. 8712.
- Hummels, David, and James Levinsohn. 1995. "Monopolistic Competition and International Trade: Reinterpreting the Evidence." *Quarterly Journal of Economics* 110: 799–836. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- Hummels, David, Jakob R. Munch, and Chong Xiang. 2015. "Offshoring and Labor Markets." <http://goo.gl/NDq24i> (accessed 8/6/2015).
- Hummels, David, and Alexandre Skiba. 2004. "Shipping the Good Applies Out? An Empirical Confirmation of the Alchian-Allen Conjecture." *Journal of Political Economy* 112(6): 1384–402.
- Hunter, Linda C., and James R. Markusen. 1988. "Per-Capita Income as a Determinant of Trade." Pages 89–109 in Robert C. Feenstra, ed., *Empirical Methods for International Trade*. Cambridge: MIT Press.
- Irwin, Douglas A., 2002. *Free Trade under Fire*. Princeton: Princeton University Press.
- . 2003. "The Optimal Tax on Antebellum U.S. Cotton Exports." *Journal of International Economics* 60(2): 275–91.
- . 2005. "The Welfare Cost of Autarky: Evidence from the Jeffersonian Trade Embargo, 1807–1809." *Review of International Economics* 13(4): 631–45.
- Irwin, Douglas A., and Nina Pavcnik. 2004. "Airbus versus Boeing Revisited: International Competition in the Aircraft Market." *Journal of International Economics* 64(2): 223–45.
- Itoh, Motoshige, and Kazuharu Kiyono. 1987. "Welfare-Enhancing Export Subsidies." *Journal of Political Economy* 95(1): 115–37.
- Janeba, Eckhard. 1998. "Tax Competition in Imperfectly Competitive Markets." *Journal of International Economics* 44(1): 135–54.
- Johnson, George, and Frank Stafford. 1999. "The Labor Market Implications of International Trade." Pages 2215–88 in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, vol. 3B. Amsterdam: Elsevier.
- Johnson, Robert C. 2012. "Trade and Prices with Heterogeneous Firms," *Journal of International Economics* 86: 43–56.
- Johnson, Robert C., and Guillermo Noguera. 2012. "Accounting for Intermediates: Production Sharing and Trade in Value Added." *Journal of International Economics* 86(2): 224–36.

- Jones, Charles I. 1995a. "Time Series Tests of Endogenous Growth Models." *Quarterly Journal of Economics* 110(2): 495–526.
- . 1995b. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103(4): 759–84.
- Jones, Charles I., and Robert E. Hall. 1999. "Why Do Some Countries Produce So Much More Output Per Worker than Others?" *Quarterly Journal of Economics* 116(1): 83–116.
- Jones, Ronald W. 1956–57. "Factor Proportions and the Heckscher-Ohlin Theorem." *Review of Economic Studies* 24: 1–10.
- . 1965. "The Structure of Simple General Equilibrium Models." *Journal of Political Economy* 73: 557–72. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 1971. "A Three-Factor Model in Theory, Trade and History." Pages 3–21 in Jagdish Bhagwati et al., eds., *Trade, Balance of Payments and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*. Amsterdam: North Holland.
- . 1979. "Technical Progress and Real Incomes in a Ricardian Trade Model." Chapter 17 in Ronald W. Jones, *International Trade: Essays in Theory*. Amsterdam: North-Holland.
- . 2000. *Globalization and the Theory of Input Trade*. Ohlin Lectures, vol. 8. Cambridge and London: MIT Press.
- Jones, Ronald, and Henryk Kierzkowski. 2001. "Globalization and the Consequences of International Fragmentation." In Rudiger Dornbusch, ed., *Money, Capital Mobility and Trade: Essays in Honor of Robert A. Mundell*. Cambridge: MIT Press.
- Jones, Ronald W., J. Peter Neary, and Francis P. Ruane. 1987. "International Capital Mobility and the Dutch Disease." Pages 86–98 in Henryk Keirzkowski, ed., *Protection and Competition in International Trade: Essays in Honour of W. M. Corden*. Oxford: Basil Blackwell.
- Jones, Ronald W., and José A. Scheinkman. 1977. "The Relevance of the Two-Sector Production Model in Trade Theory." *Journal of Political Economy* 85(5): 909–35.
- Ju, Jiandong, and Kala Krishna. 2000a. "Evaluating Trade Reform with Many Consumers." *Canadian Journal of Economics* 33(3): 787–98.
- . 2000b. "Welfare and Market Access Effects of Piecemeal Tariff Reform." *Journal of International Economics* 51: 305–16.
- Judd, Kenneth. 1985. "On the Performance of Patents." *Econometrica*, 53: 567–86.
- Katz, Lawrence F., and David Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, vol. 3A, Amsterdam: Elsevier, 1463–555.
- Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107: 35–78.
- Katz, Lawrence F., and Lawrence H. Summers. 1989a. "Can Inter-Industry Wage Differentials Justify Strategic Trade Policy?" Pages 85–124 in Robert C. Feenstra, ed., *Trade Policies for International Competitiveness*. Chicago: University of Chicago Press, and NBER.
- . 1989b. "Industry Rents: Evidence and Implications." *Brookings Papers on Economic Activity, MicroEconomics* 2: 209–75.
- Kee, Hiau Looi. 2001. "Productivity versus Endowments: A Study of Singapore's Sectoral Growth 1974–92." Policy Research Working Paper No. WPS 2702. Washington, D.C.: World Bank.
- Keller, Wolfgang. 1998. "Are International R&D Spillovers Trade-Related? Analyzing Spillovers among Randomly Matched Trade Partners." *European Economic Review* 42: 1469–81.
- . 2002. "Geographic Localization of International Technology Diffusion." *American Economic Review* 92(1): 120–42.
- Kemp, Murray C. 1962. "The Gain from International Trade." *Economic Journal* 72: 803–19.
- Kemp, Murray C., and Takashi Negishi. 1970. "Variable Returns to Scale, Commodity Taxes, Factor Market Distortions, and Implications for Trade Gains." *Swedish Journal of Economics* 72: 1–11.
- Kemp, Murray C., and Henry Wan, Jr. 1976. "An Elementary Proposition Concerning the Formation of Customs Unions." In Murray Kemp, ed., *Three Topics in the Theory of International Trade: Distribution, Welfare and Uncertainty*. Amsterdam: North-Holland.

- Kenney, Martin, and Richard Florida. 1994. "Japanese Maquiladoras: Production Organization and Global Commodity Chains." *World Development*, 22(1): 27–44.
- Khandelwal, Amit. 2010. "The Long and Short (of) Quality Ladders," *Review of Economics Studies* 77: 1450–76.
- Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei. 2013. "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters." *American Economic Review* 103(6): 2169–195.
- Kichko, Sergey, Sergey Kokovin, and Evgeny Zhelobodko. 2013. "Trade Patterns and Export Pricing under Non-CES Preferences." <http://ideas.repec.org/p/wsr/wpaper/y2013i124.html> (accessed 3/25/2015).
- Kimball, Miles S. 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit and Banking* 27(4): 1241–77.
- Klenow, Peter J., and Jonathan L. Willis. 2006. "Real Rigidities and Nominal Price Changes." <http://ssrn.com/abstract=896325> or <http://dx.doi.org/10.2139/ssrn.896325> (accessed 10/24/2014).
- Knetter, Michael M. 1989. "Price Discrimination by U.S. and German Exports." *American Economic Review* 79(1): 198–210.
- . 1993. 1993. "International Comparisons of Pricing-to-Market Behavior." *American Economic Review* 83(3): 473–86.
- Knittel, Christopher R., and Konstantinos Metaxoglou. 2014. "Estimation of Random Coefficient Demand Models: Two Empiricists' Perspective." *Review of Economics and Statistics* 96(1): 34–59.
- Kohler, Wilhelm. 2001. "A Specific-factors View on Outsourcing." *North American Journal of Economics and Finance* 12: 31–53.
- . 2004. "International outsourcing and factor prices with multistage production." *Economic Journal* 114: 166–85.
- Kohli, Ulrich R. 1978. "A Gross National Product Function and the Derived Demand for Imports and Supply of Exports." *Canadian Journal of Economics* 11: 167–82.
- . 1981. "Nonjointness and Factor Intensity in U.S. Production." *International Economic Review* 22: 3–18.
- . 1983. "Nonjoint Technologies." *Review of Economics Studies* 50: 209–19.
- . 1990a. "Price and Quantity Elasticities in Foreign Trade." *Economic Letters* 33(3): 277–81.
- . 1990b. "Growth Accounting in an Open Economy: Parametric and Nonparametric Estimates." *Journal of Economic and Social Measurement* 16: 125–36.
- . 1991. *Technology, Duality and Foreign Trade: The GNP Function Approach to Modeling Imports and Exports*. Ann Arbor: University of Michigan Press, London: Harvester Wheatsheaf.
- . 1993a. "A Symmetric Normalized Quadratic GDP Function and the U.S. Demand for Imports and Supply of Exports." *International Economic Review* 34(1): 243–55.
- . 1993b. "U.S. Technology and the Specific Factors Model." *Journal of International Economics* 34: 115–36.
- . 2004. "Real GDP, Real Domestic Income and Terms-of-Trade Changes." *Journal of International Economics* 62(1): 83–106.
- Kokko, Ari. 1994. "Technology, Market Characteristics and Spillovers." *Journal of Development Economics* 43: 279–93.
- Koopman, Robert, Zhi Wang, and Shang-Jin Wei. 2014. "Tracing Value-Added and Double Counting in Gross Exports." *American Economic Review*, 104(2): 459–94.
- Kortum, Samuel. 1997. "Research, Patenting, and Technological Change." *Econometrica*, 65: 1389–419.
- Kranton, Rachel E., and Deborah F. Minehart. 2001. "A Theory of Buyer-Seller Networks." *American Economic Review* 91: 485–508.
- Kreickemeier, Udo, and Douglas Nelson. 2006. "Fair Wages, Unemployment and Technological Change in a Global Economy." *Journal of International Economics* 70: 451–69

- Kremer, Michael. 1993. "Population Growth and Technological Change: One Million B.C. to 1990." *Quarterly Journal of Economics* 108: 681–716.
- Kremer, Michael, and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill." NBER Working Paper No. 5718.
- . 2006. "Globalization and Inequality." http://dev.wcfia.harvard.edu/sites/default/files/Kremer_GlobalizationInequality.pdf (10/16/2014)
- Krishna, Kala. 1987. "Tariffs Versus Quotas with Endogenous Quality." *Journal of International Economics* 23(1/2): 97–113.
- . 1989. "Trade Restrictions as Facilitating Practices." *Journal of International Economics* 26: 251–270. Reprinted in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- . 1990. "Protection and the Product Line: Monopoly and Product Quality." *International Economic Review* 31(1): 87–102.
- . 2002. "Understanding Rules of Origin." Pennsylvania State University, manuscript.
- Krishna, Kala, and Anne O. Krueger. 1995. "Implementing Free Trade Areas: Rules of Origin and Hidden Protection." In James Levinsohn, Alan Deardorff, and Robert Stern, eds., *New Directions in Trade Theory*. Ann Arbor: University of Michigan Press.
- Krishna, Pravin. 1998. "Regionalism and Multilateralism: A Political Economy Approach." *Quarterly Journal of Economics* 108: 227–51.
- . 2003. "Are Regional Trading Partners 'Natural?'" *Journal of Political Economy* 111(1): 202–26.
- Krishna, Pravin, and Arvind Panagariya. 2002. "On Necessarily Welfare-Enhancing Free Trade Areas." *Journal of International Economics* 57(2): 353–67.
- Krueger, Alan B. 1997. "Labor Market Shifts and the Price Puzzle Revisited." NBER Working Paper No. 5924.
- Krueger, Alan B., and Lawrence Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica*, 56: 69–93.
- Krueger, Anne O. 1974. "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64(3): 291–303.
- Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9: 469–79. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70: 950–59. Reprinted as chapter 11 in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- . 1981. "Intra-Industry Specialization and the Gains from Trade." *Journal of Political Economy* 89: 959–73.
- . 1984. "Import Protection as Export Promotion: International Competition in the Presence of Oligopoly and Economics of Scale." In Henryk Kierzkowski, ed., *Monopolistic Competition and International Trade*. Oxford: Oxford University Press. Reprinted in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- . 1991a. "Is Bilateralism Bad?" In Elhanan Helpman and Assaf Razin, eds., *International Trade and Trade Policy*. Cambridge: MIT Press.
- . 1991b. "The Move Towards Free Trade Zones." In *Policy Implications of Trade and Currency Zones*. Federal Reserve Bank of Kansas City, 7–43.
- . 1994. "Does Third World Growth Hurt First World Prosperity?" *Harvard Business Review* 72(4): 113–21.
- . 1995. "Growing World Trade: Causes and Consequences." *Brookings Paper on Economic Activity*. 1: 327–62.
- . 2000. "Technology, Trade, and Factor Prices." *Journal of International Economics* 50(1): 51–72.
- Krugman, Paul R., and Anthony J. Venables, 1995 "Globalization and the Inequality of Nations." *Quarterly Journal of Economics* 110(4): 857–80.

- Lancaster, Kelvin. 1975. "Socially Optimal Product Differentiation." *American Economic Review* 65: 567–85.
- . 1979. *Variety, Equity and Efficiency*. New York: Columbia University Press.
- . 1980. "Intra-Industry Trade under Perfect Monopolistic Competition." *Journal of International Economics* 10: 151–75.
- Lawrence, Colins, and Pablo T. Spiller. 1983. "Product Diversity, Economics of Scale and International Trade." *Quarterly Journal of Economics* 98: 63–83.
- Leamer, Edward E. 1980. "The Leontief Paradox, Reconsidered." *Journal of Political Economy* 88(3): 495–503. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 1984. *Source of Comparative Advantage*. Cambridge: MIT Press.
- . 1987. "Paths of Development in the 3-Factor, N-Good General Equilibrium Model." *Journal of Political Economy* 95: 961–99.
- . 1994. "Trade, Wages and Revolving Door Trade Ideas." NBER Working Paper No. 4716.
- . 1995. "The Heckscher-Ohlin Model in Theory and Practice." *Princeton Studies in International Finance*, no. 77.
- . 1996. "The Effects of Trade in Services, Technology Transfer, and Delocalisation on Local and Global Income Inequality." *Asia-Pacific Economic Review* 2: 44–60.
- . 1998. "In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages." Pages 141–203 in Susan M. Collins, ed., *Imports, Exports, and the American Worker*. Washington, D.C.: Brookings Institution Press. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 2000. "What's the Use of Factor Contents?" *Journal of International Economics* 50: 17–50.
- Leamer, Edward E., and Harry P. Bowen. 1981. "Cross-Section Tests of the Heckscher-Ohlin Theorem: Comment." *American Economic Review* 71(5): 1040–43. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data or Rising Demand for Skill?" *American Economic Review*: 461–98.
- Leontief, Wassily W. 1953. Domestic Production and Foreign Trade: The American Capital Position Re-examined." Proceedings of the American Philosophical Society, 97: 332–49. Reprinted in Richard Caves and Harry G. Johnson, eds. *Readings in International Economics*. Homewood, Ill.: Irwin, 1968.
- Lerner, Abba P. 1952. "Factor Prices and International Trade." *Economica* (new series) 19: 1–15.
- Levinsohn, James. 1993. "Testing the Imports-as-Market-Discipline Hypothesis." *Journal of International Economics* 35(1/2): 1–22.
- Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70(2): 317–41.
- Levy, Philip I. 1997. "A Political-Economic Analysis of Free-Trade Agreements." *American Economic Review* 87(4): 506–19.
- Li, David D., and Qi Chen. 1998. "Why Aren't Foreign Direct Investments Always Welcome? A Political Economy Analysis." University of Michigan, manuscript.
- Lipsey, Robert E. 1999. "The Role of FDI in International Capital Flows." Pages 307–62 in Martin Feldstein, ed., *International Capital Flows*. Chicago: University of Chicago Press,
- . 2003. "Foreign Direct Investment and the Operations of Multinational Firms: Concepts, History and Data." In Kwan Choi and James Harrigan, eds., *Handbook of International Trade*, vol. 1. Oxford: Blackwell.
- List, F. 1856. *National System of Political Economy*. Translated by G. A. Matile. Philadelphia: Lippincott.
- Liu, Runjuan, and Daniel Trefler. 2011. "A Sorted Tale of Globalization: White-Collar Jobs and the Rise of Service Offshoring." NBER Working Paper No. 17559.
- Lopez de Silanes, Florencio, James R. Markusen, and Thomas F. Rutherford. 1994. "Complemenarity and Increasing Returns in Intermediate Inputs." *Journal of Development Economics* 45(1): 101–19

- Lorentowicz, Andzelika, Dalia Marin, and Alexander Raubold. 2008. "Is Human Capital Losing from Outsourcing? Evidence for Austria and Poland." Pages 225–58 in Steven Brakman and Harry Garretsen, eds., *Foreign Direct Investment and the Multinational Enterprise*. CESifo Seminar Series. Cambridge, Mass., and London: MIT Press.
- Ma, Hong. 2009. "Firm Heterogeneity and Multiproduct Firms in International Trade." University of California, Davis, Ph.D. diss.
- MacDougall, G.D.A. 1960. "The Benefits and Costs of Private Investment from Abroad: A Theoretical Approach." *Economic Record*, Special Issue, March.
- Magee, Christopher. 2003. "Endogenous Preferential Trade Agreements: An Empirical Analysis." *The B.E. Journal of Economic Analysis & Policy* 2(1): Article 15.
- Maggi, Giovanni. 1996. "Strategic Trade Policies with Endogenous Mode of Competition." *American Economic Review* 86(1): 237–58.
- . 1999. "The Role of Multilateral Institutions in International Trade Cooperation." *American Economic Review* 89(1): 190–214.
- Maggi, Giovanni, and Robert W. Staiger. 2011. "The Role of Dispute Settlement Procedures in International Trade Agreements." *Quarterly Journal of Economics* 126(1): 475–515.
- Manasse, Paolo, and Alessandro Turrini. 2001. "Trade, Wages, and 'Superstars.'" *Journal of International Economics* 54: 97–117.
- Mandel, Benjamin. 2010. "Heterogeneous Firms and Import Quality: Evidence from Transaction-Level Prices." Board of Governors of the Federal Reserve Bank System, International Finance Discussion Paper No. 991.
- Mandel, Michael. 2007. "The Real Cost of Offshoring." *Business Week*, June 18, pp. 28–34.
- Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107(2): 407–38.
- Marin, Dalia. 2004. "'A Nation of Poets and Thinkers'—Less So with Eastern Enlargement? Austria and Germany." Discussion Papers in Economics no. 329. University of Munich, Department of Economics.
- Marjit, Sugata, and Rajat Acharyya. 2003. *International Trade, Wage Inequality and the Developing Economy: A General Equilibrium Approach*. Heidelberg: Physica Verlag.
- Markusen, James R. 1981. "Trade and the Gains from Trade with Imperfect Competition." *Journal of International Economics* 11: 531–51. Reprinted as chapter 3 in Gene M. Grossman, ed., *Imperfect Competition and International Trade*. Cambridge: MIT Press, 1992.
- . 1984. "Multinationals, Multi-Plant Economies, and the Gains from Trade." *Journal of International Economics* 16: 205–26.
- . 1989. "Trade in Producer Services and in Other Specialized, Intermediate Inputs." *American Economic Review* 79: 85–95.
- . 1990. "De-rationalizing Tariffs with Specialized Intermediate Inputs and Differentiated Final Goods." *Journal of International Economics* 28(3–4): 375–83.
- . 1995. "The Boundaries of Multinational Firms and International Trade." *Journal of Economic Perspectives* 9: 169–89.
- . 2001. "Contracts, Intellectual Property Rights, and Multinational Investment in Developing Countries." *Journal of International Economics* 53(1): 189–204.
- . 2002. *Multinational Firms and the Theory of International Trade*. Cambridge: MIT Press.
- Markusen, James R., and Keith E. Maskus. 2001. "Multinational Firms: Reconciling Theory and Evidence." Pages 71–98 in Magnus Blomstrom and Linda S. Goldberg, eds., *Topics in Empirical Economics: A Festschrift in Honor of Robert E. Lipsey*. Chicago: University of Chicago Press, and NBER.
- . 2002. "Discriminating among Theories of the Multinational Enterprise." *Review of International Economics* 10(4): 694–707.
- Markusen, James R., and Anthony J. Venables. 1998. "Multinational Firms and the New Trade Theory." *Journal of International Economics* 46: 183–203.
- . 2000. "The Theory of Endowment, IntraIndustry, and Multinational Trade." *Journal of International Economics* 52: 209–35.

- Martin, Julien, and Isabelle Méjean. 2012. "Low-Wage Countries' Competition, Reallocation across Firms and the Quality Content of Exports." CEPII Working Paper No. 2012-04.
- Mataloni, Raymond. 1995. "A Guide to BEA Statistics on U.S. Multinational Companies." *Survey of Current Business* (March).
- Matsuyama, Kiminori. 2008. "Ricardian Trade Theory." In Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*. 2nd ed. New York and London: Palgrave Macmillan. http://www.dictionaryofeconomics.com/article?id=pde2008_R000276.
- . 2013. "Endogenous Ranking and Equilibrium Lorenz Curve Across (ex ante) Identical Countries." *Econometrica* 81(5): 2009–31.
- Mayer, Thierry, Marc Melitz, and Gianmarco I. P. Ottaviano. 2014. "Market size, Competition, and the Product Mix of Exporters." *American Economic Review* 104(2): 495–536.
- Mayer, Wolfgang. 1974. "Short and Long Run Equilibria for a Small Open Economy." *Journal of Political Economy* 82: 955–67.
- . 1984. "Endogenous Tariff Formation." *American Economic Review* 74(5): 970–85.
- McCallum, John. 1995. "National Borders Matter." *American Economic Review* 85(3): 615–23.
- McCalman, Phillip. 2002. "Multilateral Trade Negotiations and the Most Favored Nation Clause." *Journal of International Economics* 57(1): 151–76.
- . 2004. "Protection for Sale and Trade Liberalization: An Empirical Investigation." *Review of International Economics* 12(1): 81–94.
- McFadden, Daniel. 1978. "Modeling the Choice of Residential Location." Pages 75–96 in Anders Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*. Amsterdam: North-Holland.
- . 1981. "Econometric Models of Probabilistic Choice." In Charles F. Manski and Daniel McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press, 198–272.
- McFadden, Daniel, and Kenneth Train. 2000. "Mixed MNL Models for Discrete Response." *Applied Econometrics* 15: 447–70.
- McLaren, John. 2000. "'Globalization' and Vertical Structure." *American Economic Review* 90(5): 1239–54.
- . 2002. "A Theory of Insidious Regionalism." *Quarterly Journal of Economics* 98(2): 571–608.
- Melitz, Marc. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–725.
- . 2005. "When and How Should Infant Industries Be Protected?" *Journal of International Economics* 66(1): 177–96.
- Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2008. "Market Size, Trade, and Productivity." *Review of Economic Studies* 75(1): 295–316.
- Melitz, Marc J., and Stephen J. Redding. 2013. "Firm Heterogeneity and Aggregate Welfare." NBER Working Paper No. 18919.
- . 2014a. "Heterogeneous Firms and Trade." Pages 1–54 in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Trade*, vol. 4. Amsterdam: Elsevier.
- . 2014b. "Missing Gains from Trade?" *American Economic Review* 104(5): 317–21.
- Mill, John Stuart. 1909. *The Principles of Political Economy*. London: Longmans, Green.
- Minhas, Bagicha S. 1962. "The Homohypallagic Production Function, Factor-Intensity Reversals and the Hecksher-Ohlin Theorem." *Journal of Political Economy* 70(2): 138–56.
- Mitra, Devashish. 1999. "Endogenous Lobby Formation and Endogenous Protection: A Long-Run Model of Trade Policy Determination." *American Economic Review* 89(5): 1116–34.
- Mitra, Devashish, and Priya Ranjan. 2010. "Offshoring and Unemployment: The Role of Search Frictions and Labor Mobility." *Journal of International Economics* 81(2): 219–29.
- Mitra, Devashish, Dimitrios Thomakos, and Mehmet Ulubasoglu. 2002. "'Protection for Sale' in a Developing Country: Democracy versus Dictatorship." *Review of Economics and Statistics* 84(3): 497–508.
- Monte, Ferdinando. 2011. "Skill Bias, Trade, and Wage Dispersion." *Journal of International Economics* 83 (2): 202–18.

- Morales, Eduardo, Gloria Sheu, and Andrés Zahler. 2014. "Gravity and Extended Gravity: Using Moment Inequalities to Estimate a Model of Export Entry." NBER Working Paper No. 19916.
- Mundell, Robert A. 1957. "International Trade and Factor Mobility." *American Economic Review* 47: 321–35.
- Mussa, Michael. 1974. "Tariffs and the Distribution of Income: The Importance of Factor Specificity, Substitutability, and Intensity in the Short and Long Run." *Journal of Political Economy* 82: 1191–203.
- . 1979. "The Two-Sector Model in Terms of its Dual." *Journal of International Economics* 9(4): 513–26.
- Mrázová, Monika, and J. Peter Neary. 2013. "Not so Demanding: Preference Structure, Firm Behavior and Welfare." University of Oxford, Discussion Paper No. 691. <http://www.economics.ox.ac.uk/materials/papers/13172/paper691.pdf> (accessed 10/16/2014).
- . 2014. "Together at Last: Trade Costs, Demand Structure, and Welfare." *American Economic Review* 104(5): 298–303.
- Naughton, Barry. 1996. "China's Emergence and Prospects as a Trading Nation." *Brookings Papers on Economic Activity* 2: 273–343.
- . 2003. "How Much Can Regional Integration Do to Unify China's Markets?" Pages 204–32 in Nicholas C. Hope, Dennis Tao Yang, and Mu Yang Li, eds., *How Far Across the River? Chinese Policy Reform at the Millennium*, 2003. Stanford: Stanford University Press.
- Neary, J. Peter. 1978. "Short-Run Capital Specificity and the Pure Theory of International Trade." *Economic Journal* 88: 488–510. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 1988. "Tariffs, Quotas, and Voluntary Export Restraints with and without Internationally Mobile Capital." *Canadian Journal of Economics* 21(4): 714–35.
- . 2003. "Globalization and Market Structure." *Journal of the European Economic Association* 1(2–3): 245–71.
- Nikaido, Hukukane. 1972. "Relative Shares and Factor Price Equalization." *Journal of International Economics* 2(3): 257–63.
- Nunn, Nathan. 2007. "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade." *Quarterly Journal of Economics* 122(2): 569–600.
- Nunn, Nathan, and Daniel Trefler. 2013. "Incomplete Contracts and the Boundaries of the Multinational Firm." *Journal of Economic Behavior and Organization* 94(1): 330–44.
- Obstfeld, Maurice, and Kenneth Rogoff. 2000. "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" Pages 339–89 in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual*, 2000. Cambridge: MIT Press.
- Odenski, Lindsay. 2012. "The Task Composition of Offshoring by U.S. Multinationals." *International Economics* 131: 5–21.
- . 2014. "Offshoring and the Polarization of the U.S. Labor Market." *Industrial and Labor Relations Review* 67 (2.5): 734–61.
- Ohlin, Bertil. 1933. *Interregional and International Trade*. Cambridge: Harvard University Press.
- Ohyama, M. 1972. "Trade and Welfare in General Equilibrium." *Keio Economic Studies* 9: 37–73.
- Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–97.
- Ossa, Ralph. 2011. "A 'New Trade' Theory of GATT/WTO Negotiations." *Journal of Political Economy* 119(1): 122–52.
- . 2012. "Why Trade Matters After All." NBER Working Paper No. 18113.
- Ottaviano, Gianmarco I. P., Giovanni Peri, and Greg C. Wright. 2013. "Immigration, Offshoring, and American Jobs." *American Economic Review* 103(5): 1925–59.
- Ottaviano, Gianmarco I. P., and Jacques-François Thisse. 2011. "Monopolistic Competition, Multiproduct Firms, and Product Diversity." *The Manchester School* 79(5): 938–51.
- Overman, Henry, Stephen Redding, and Anthony J. Venables. 2003. "The Economic Geography of Trade, Production and Income: A Survey of Empirics." In Kwan Choi and James Harrigan, eds., *Handbook of International Trade, Volume 1*. Oxford: Blackwell, 353–87.

- Panagariya, Arvind. 2000. "Evaluating the Factor-Content Approach to Measuring the Effect of Trade on Wage Inequality?" *Journal of International Economics* 50: 91–116.
- Pavcnik, Nina. 2002a. "Trade Disputes in the Commercial Aircraft Industry." *World Economy* 25: 733–51.
- . 2002b. "Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants." *Review of Economic Studies*, 69(1): 245–76.
- Perla, Jesse, Christopher Tonetti, and Michael E. Waugh. 2014. "Equilibrium Technology Diffusion, Trade, and Growth." http://christophertonetti.com/files/papers/PerlaTonettiWaugh_DiffusionTradeAndGrowth.pdf (accessed 11/30/2014).
- Perry, Martin K. 1989. "Vertical Integration: Determinants and Effects." Pages 185–255 in Richard Schmalansee and Robert Willig, eds., *Handbook of Industrial Organization*, vol. 1. Amsterdam: North-Holland.
- Phelps, Edmund S. 1997. *Rewarding Work*. Cambridge: Harvard University Press.
- Pollak, Robert A. 1989. *The Theory of the Cost-of-Living Index*. Oxford: Oxford University Press.
- Prebisch, Raul. 1950. *The Economic Development of Latin America and Its Principal Problems*. New York: United Nations.
- Prusa, Thomas J. 1991. "The Selection of Anti-Dumping Cases for Withdrawal." In Robert E. Baldwin, ed., *Empirical Studies of Commercial Policy*. Chicago: University of Chicago Press, and NBER.
- . 1992. "Why Are So Many Antidumping Petitions Withdrawn?" *Journal of International Economics* 33: 1–20.
- . 2001. "On the Spread and Impact of Anti-Dumping." *Canadian Journal of Economics* 34(3): 591–611.
- Raff, Horst, and Nicolas Schmitt. 2016. "Manufacturers and Retailers in the Global Economy." *Canadian Journal of Economics*, forthcoming.
- Raff, Horst, and Joaquim Wagner. 2013. "Productivity and the Product Scope of Multi-Product Firms: A Test of Feenstra-Ma." *Economic Bulletin* 33(1): 415–19.
- Rauch, James E. 1999. "Networks versus Markets in International Trade." *Journal of International Economics* 48(1): 7–37.
- . 2001. "Business and Social Networks in International Trade." *Journal of Economic Literature* 39: 1177–203.
- Rauch, James E., and Alessandra Casella. 2001. *Networks and Markets*. New York: Russell Sage.
- . 2003. "Overcoming Informational Barriers to International Resource Allocation: Prices and Ties." *Economic Journal* 113: 21–42.
- Rauch, James E., and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics*, 84(1): 116–30.
- . 2003. "Information, International Substitutability, and Globalization." *American Economic Review* 93(3): 775–91.
- Rauch, James E., and Joel Watson. 2002. "Entrepreneurship and International Trade." NBER Working Paper No. 8708.
- Redding, Stephen J. 2002. "Specialization Dynamics." *Journal of International Economics* 58(2): 299–334.
- Redding, Stephen J., Daniel Sturm, and Nikolaus Wolf. 2011. "History and Industrial Location: Evidence from German Airports." *Review of Economics and Statistics* 93(3): 814–31.
- Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, vol. 5. Amsterdam: Elsevier.
- Redding, Stephen J., and Anthony J. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics* 62(1): 53–82.
- Reid, Peter C., 1990. *Made Well in America: Lessons from Harley-Davidson on Being the Best*. New York: McGraw-Hill.
- Revelt, David, and Kenneth Train. 2000. "Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier." Department of Economics, University of California, Berkeley, working paper no. E00-274.

- Ricardo, David. 1951. *The Works and Correspondence of David Ricardo*. Edited by P. Sfaffa. Cambridge: Cambridge University Press.
- Richardson, J. David. 1982. "Trade Adjustment Assistance under the Trade Act of 1974: An Analytical Examination and Worker Survey." Pages 321–57 in Jagdish N. Bhagwati, ed., *Import Competition and Response*. Chicago: University of Chicago Press, and NBER.
- . 1995. "Income Inequality and Trade: How to Think, What to Conclude." *Journal of Economic Perspectives*, 9(3): 33–56.
- Ries, John. 1993. "Windfall Profits and Vertical Relationships: Who Gained in the Japanese Auto Industry from VERs?" *Journal of Industrial Economics* 41(3): 259–76.
- Rivera-Batiz, Luis A., and Paul Romer. 1991a. "Economic Integration and Economic Growth." *Quarterly Journal of Economics* 106: 531–55.
- . 1991b. "International Trade with Endogenous Technological Change." *European Economic Review* 35: 971–1004.
- Robertson, Raymond. 2004. "Relative Prices and Wage Inequality: Evidence from Mexico." *Journal of International Economics* 64(2): 387–409.
- Robinson, Joan. 1933. *The Economics of Imperfect Competition*. London: Macmillan.
- Rodriguez, Carlos A. 1979. "The Quality of Imports and the Differential Effects of Tariffs, Quotas, and Quality Controls as Protective Devices." *Canadian Journal of Economics* 12(3): 439–49.
- Rodriguez, Francisco, and Dani Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." Pages 261–325 in Ben S. Gernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2000*. Cambridge: MIT Press.
- Rodríguez-López, José Antonio. 2011. "Prices and Exchange Rates: A Theory of Disconnect." *Review of Economic Studies* 78(3): 1135–77.
- . 2014. "What Drives Offshoring Decisions? Selection and Escape-Competition Mechanisms." University of California, Irvine, manuscript. http://www.socsci.uci.edu/~jantonio/Papers/offshoring_decisions.pdf (accessed 10/5/2014).
- Rodrik, Dani. 1995a. "The Political Economy of Trade Policy." Pages 1457–94 in Gene M. Grossman and Kenneth Rogoff, eds., *Handbook of International Economics*, vol. 3. Amsterdam: Elsevier.
- . 1995b. "Taking Trade Policy Seriously: Export Subsidization as a Case Study in Policy Effectiveness." Pages 347–84 in James Levinsohn, Alan V. Deardorff, and Robert M. Stern, eds., *New Directions in Trade Theory*. Michigan: University of Michigan Press.
- . 1997. *Has Globalization Gone Too Far?* Washington, D.C.: Institute for International Economics.
- Romalis, John. 2004. "Factor Proportions and the Structure of Commodity Trade." *American Economic Review* 94(1): 67–97.
- . 2007. "NAFTA's and CUSFTA's Impact on International Trade." *Review of Economics and Statistics* 89(3): 416–35.
- Romer, Paul. 1990. "Endogenous Technical Change." *Journal of Political Economy* 98(5, Part 2): S71–S102.
- . 1994. "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions." *Journal of Development Economics* 43: 5–38.
- Rose, Andrew K. 2000. "One Money, One Market: Estimating the Effect of Common Currencies on Trade." *Economic Policy* 30: 9–45.
- Rose, Andrew K., and Eric van Wincoop. 2001. "National Money as a Barrier to International Trade: The Real Case for Currency Union." *American Economic Review* 91(2): 386–90.
- Rosen, Daniel. 1998. *Behind the Open Door: Foreign Firms in China*. Washington, D.C.: Institute for International Economics.
- Ruhl, Kim. 2014. "The Aggregate Impact of Antidumping Policies." New York University, manuscript. http://www.kimjruhl.com/storage/antidumping_newest.pdf (accessed 10/10/2014).
- Rybczynski, T. N. 1955. "Factor Endowments and Relative Commodity Prices." *Economica* 22: 336–41.
- Sachs, Jeffrey D. 2001. "Tropical Underdevelopment." NBER Working Paper No. 8119.

- Sachs, Jeffrey D., and Howard J. Shatz. 1994. "Trade and Jobs in U.S. Manufacturing." *Brookings Papers on Economic Activity* 1: 1–84.
- . 1998. "International Trade and Wage Inequality: Some New Results." Pages 215–40 in Susan M. Collins, ed., *Imports, Exports, and the American Worker*. Washington, D.C.: Brookings Institution Press.
- Sachs, Jeffrey, and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–118.
- Samuelson, Paul A. 1949. "International Factor Price Equalization Once Again." *Economic Journal* 59: 181–197. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 1952. "The Transfer Problem and Transport Costs: The Terms of Trade when Impediments Are Absent." *Economic Journal* 62: 278–304.
- . 1953–54. "The Prices of Goods and Factor in General Equilibrium." *Review of Economic Studies* 21(1): 1–20.
- . 1956. "Social Indifference Curves." *Quarterly Journal of Economics* 70(1): 1–22.
- . 1962. "The Gains from International Trade Once Again." *Economic Journal* 72: 820–29.
- . 2001. "The Ricardo-Sraffa Paradigm Comparing Gains from Trade in Inputs and Finished Products." *Journal of Economic Literature* 39: 1204–14.
- . 2004. "Why Ricardo and Mill Rebut and Confirm Arguments of Mainstream Economists Supporting Globalization." *Journal of Economic Perspectives* 18(3): 135–46.
- Sampson, Thomas. 2013. "Assignment Reversals: Trade, Skill Allocation and Wage Inequality." London School of Economics. <http://personal.lse.ac.uk/sampsont/Assignment%20Reversals.pdf> (accessed 10/5/2014).
- . 2014. "Selection into Trade and Wage Inequality." *American Economic Journal: Microeconomics* 6(3): 157–202.
- . 2016. "Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth." *Quarterly Journal of Economics*, forthcoming.
- Santos Silva, J.M.C., and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88(4): 641–58.
- Sato, Kazuo. 1976. "The Ideal Log-Change Index Number." *Review of Economics and Statistics* 58: 223–28.
- Saure, Philip. 2009. "Bounded Love of Variety and Patterns of Trade." Swiss National Bank.
- Schiff, Maurice, and Won Chang. 2003. "Market Presence, Contestability, and the Terms of Trade Effects of Regional Integration." *Journal of International Economics* 60(1): 161–75.
- Schott, Peter K. 2003. "One Size Fits All? Heckscher-Ohlin Specialization in Global Production." *American Economic Review* 93(3): 686–708.
- . 2004. "Across-Product versus Within-Product Specialization in International Trade." *Quarterly Journal of Economics* 119(2): 647–78.
- . 2008. "The Relative Sophistication of Chinese Exports." *Economic Policy* 23: 5–49.
- Segerstrom, Paul, T.C.A. Anant, and Elias Dinopoulos. 1990. "A Schumpeterian Model of the Product Life Cycle." *American Economic Review* 80: 1077–99.
- Shiells, Clinton R., Robert M. Stern, and Alan V. Deardorff. 1986. "Estimates of the Elasticities of Substitution between Imports and Home Goods for the United States." *Weltwirtschaftliches Archiv* 122(3): 497–519.
- Simonovska, Ina. 2015. "Income Differences and Prices of Tradables: Insights from an Online Retailer." *Review of Economic Studies*, forthcoming.
- Simonovska, Ina, and Michael E. Waugh. 2014a. "The Elasticity of Trade: Estimates and Evidence." *Journal of International Economics* 92(1): 34–50.
- . 2014b. "Trade Models, Trade Elasticities, and the Gains from Trade." NBER Working Paper No. 20495.
- Sitchinava, Nino. 2008. "Trade, Technology, and Wage Inequality: Evidence from U.S. Manufacturing, 1989–2004." University of Oregon, Ph.D. diss.
- Slater, Joanna and Eriko Amaha. 1999. "Masters of the Trade." *Far Eastern Economic Review* (July 22): 10–14.

- Slaughter, Matthew J. 2000. "What Are the Results of Product-Price Studies and What Can We Learn from Their Differences?" Pages 129–65 in Robert C. Feenstra, ed., *The Effects of International Trade on Wages*. Chicago: University of Chicago Press.
- . 2001. "International Trade and Labor-Demand Elasticities." *Journal of International Economics* 54: 27–56.
- Smith, M. Alasdair. 1977. "Capital Accumulation in the Open Two-Sector Economy." *Economic Journal* 87(346): 273–82.
- Soderbury Anson. 2010. "Investigating the Asymptotic Properties of Elasticity of Substitution Estimates." *Economic Letters* 190(2): 57–62.
- . 2012. "Estimating Import Supply and Demand Elasticities: Analysis and Implications." Purdue University, manuscript.
- . 2014. "Trade Elasticities, Heterogeneity, and Optimal Tariffs." http://web.ics.purdue.edu/~asoderbe/Papers/soderbery_heterogeneity.pdf (accessed 11/10/2014).
- Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70: 65–94.
- . 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 39: 312–20.
- Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *The Bell Journal of Economics* 6(2): 417–29.
- . 1976. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* 43(2): 217–35.
- Spencer, Barbara J., and Larry D. Qiu. 2001. "Keiretsu and Relationship-Specific Investment: A Barrier to Trade?" *International Economic Review* 42(4): 871–901.
- Staiger, Robert. 1986. "Measurement of the Factor Content of Foreign Trade with Traded Intermediate Goods." *Journal of International Economics* 21: 361–68.
- Staiger, Robert W., and Frank A. Wolak. 1992. "The Effect of Antidumping Law in the Presence of Foreign Monopoly." *Journal of International Economics* 32: 265–87.
- . 1994. "Measuring Industry-Specific Protection: Antidumping in the United States." *Brookings Papers on Economic Activity: Microeconomics* 1: 51–118.
- Stigler, George J. 1951. "The Division of Labor Is Limited by the Extent of the Market." *Journal of Political Economy* 59(3): 185–93.
- Stolper, Wolfgang F., and Paul A. Samuelson. 1941. "Protection and Real Wages." *Review of Economic Studies* 9: 58–73.
- Suomela, John W., 1993. *Free Trade versus Fair Trade: The Making of American Trade Policy in a Political Environment*. Turku, Finland: Institute for European Studies.
- Swan, Peter L. 1970. "Durability of Consumption Goods." *American Economic Review* 60(5): 884–94.
- Swenson, Deborah. 2004. "Foreign Investment and the Mediation of Trade Flows." *Review of International Economics* 12(4): 609–29.
- Taylor, Christopher T. 2004. "The Economic Effects of Withdrawn Antidumping Investigations: Is There Evidence of Collusive Settlements?" *Journal of International Economics* 62(2): 295–312.
- Taylor, M. Scott. 1993. "'Quality Ladders' and Ricardian Trade." *Journal of International Economics* 34: 225–43.
- . 1994. "'Once-off' and Continuing Gains from Trade." *Review of Economic Studies* 61: 589–601.
- Timmer, Marcel P., Abdul Azeez Erumban, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries. 2014. "Slicing Up Global Value Chains." *Journal of Economic Perspectives* 28(2): 99–118.
- Tinbergen, Jan. 1962. *Shaping the World Economy*. New York: Twentieth Century Fund.
- Törnqvist, L. 1936. "The Bank of Finland's Consumption Price Index." *Bank of Finland Monthly Bulletin* 10: 1–8.
- Train, Kenneth. 1986. *Qualitative Choice Analysis*. Cambridge: MIT Press.
- Trefler, Daniel. 1993a. "International Factor Price Differences: Leontief was Right!" *Journal of Political Economy* 101(6): 961–87.

- . 1993b. "Trade Liberalization and the Theory of Endogenous Protection." *Journal of Political Economy* 101(1): 138–60.
- . 1995. "The Case of Missing Trade and Other Mysteries." *American Economic Review* 85(5): 1029–46. Reprinted in Edward E. Leamer, ed., *International Economics*. New York: Worth, 2001.
- . 2004. "The Long and Short of the Canada-U.S. Free Trade Agreement." *American Economic Review* 94: 870–95.
- Trefler, Daniel, and Susan Chun Zhu. 2000 "Beyond the Algebra of Explanation: HOV for the Technology Age." *American Economic Review* 145–49.
- . 2010. "The Structure of Factor Content Predictions." *Journal of International Economics* 82(3): 195–207.
- Turunen-Red, Arja H., and Alan D. Woodland. 1991. "Strict Pareto Improving Reforms of Tariffs." *Econometrica* 59: 1127–52.
- . 2000. "Multilateral Policy Reforms and Quantity Restrictions on Trade." *Journal of International Economics* 52(1): 153–68.
- Tybout, James R., Jamie de Melo, and Vittorio Corbo. 1991. "The Effects of Trade Reform on Scale and Technical Efficiency: New Evidence from Chile." *Journal of International Economics* 31(3/4): 231–50.
- Tybout, James R., and M. Daniel Westbrook. 1995. "Trade Liberalization and Dimensions of Efficiency Change in Mexican Manufacturing Industries." *Journal of International Economics* 39(1/2): 53–78.
- Tyson, Laura D'Andrea. 1992. *Who's Bashing Whom: Trade Conflict in High Technology Industries*. Washington, D.C.: Institute for International Economics.
- United Nations. 1996. *World Investment Report*. New York.
- United States Department of Commerce, Bureau of Economic Analysis. 1995. *Foreign Direct Investment in the United States, 1992 Benchmark Survey, Final Results*. Washington, D.C.
- United States International Trade Commission. 1983. *Heavyweight Motorcycles, and Engines and Power Train Subassemblies Therefor*. Publication 1342. Washington, D.C.: USITC.
- . 1983–84. *Heavyweight Motorcycles: Quarterly Report on Selected Economic Indicators*. Various publications. Washington, D.C.: USITC.
- Van Ark, Bart, Erik Monnikhof, and Marcel Timmer. 1999. "Prices, Quantities, and Productivity in Industry: A Study of Transition Economies in a Comparative Perspective?" Pages 327–64 in Alan Heston and Robert E. Lipsey, eds., *International and Inter-area Comparisons of Income, Output, and Prices*. NBER Studies in Income and Wealth, No. 61. Chicago: University of Chicago Press.
- Vaneck, Jaroslav. 1968. "The Factor Proportions Theory: The N-Factor Case." *Kyklos* 21: 749–54.
- Vartia, Y.O. 1976a. *Relative Changes and Index Numbers*. Helsinki: Research Institute of the Finnish Economy.
- . 1976b. "Ideal Log-Change Index Numbers." *Scandinavian Journal of Statistics* 3: 121–26.
- Ventura, Jaume. 1997. "Growth and Interdependence." *Quarterly Journal of Economics* 107(1): 57–84.
- Viner, Jacob. 1950. *The Customs Union Issue*. New York: Carnegie Endowment for International Peace.
- . 1966. *Dumping: A Problem in International Trade*. New York: Augustus M. Kelley.
- Voigtländer, Nico. 2014. "Skill Bias Magnified: Intersectoral Linkages and White-Collar Labor Demand in U.S. Manufacturing." *Review of Economics and Statistics* 96(3): 495–513.
- Wei, Shang-Jin. 2000. "Why Does China Attract So Little Foreign Investment?" Pages 239–61 in Takatoshi Ito and Anne O. Krueger, eds., *The Role of Foreign Direct Investment in East Asian Economic Development, 2000*. Chicago: University of Chicago Press.
- Weinstein, David E. 1992. "Competition and Unilateral Dumping." *Journal of International Economics* 32: 379–88.
- Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

- . 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Winkler, Deborah, and William Milberg. 2012. "Bias in the 'Proportionality Assumption' Used in the Measurement of Offshoring." *World Economics* 13(4): 39–60.
- Wong, Kar-jiu. 1989. "Optimal Threat of Trade Restriction and Quid Pro Quo Foreign Investment." *Economics and Politics*, 1: 277–300.
- Wood, Adrian. 1995. "How Trade Hurt Unskilled Workers." *Journal of Economic Perspectives*, 9(3): 57–80.
- Woodland, Alan D. 1977. "A Dual Approach to Equilibrium in the Production Sector in International Trade Theory." *Canadian Journal of Economics* 10: 50–68.
- . 1982. *International Trade and Resource Allocation*. Amsterdam: North-Holland.
- World Bank. 1993. *The East Asian Miracle: Economic Growth and Public Policy*. Oxford: Oxford University Press.
- Wright, Greg. C. 2014. "Revisiting the Employment Impact of Offshoring." *European Economic Review* 66: 63–83.
- Xu, Bin. 2001. "Factor Bias, Sector Bias, and the Effects of Technical Progress on Relative Factor Prices." *Journal of International Economics* 54(1): 5–26.
- Yang, Mu-Jeung. 2014. "Firm Heterogeneity and Trade in a Distorted-Gravity World." University of Washington, Seattle. http://faculty.washington.edu/mjyang/research/MJY_DGW_v1.pdf (accessed 10/16/2014).
- Yeaple, Stephen R. 2005. "A Simple Model of Firm Heterogeneity, International Trade, and Wages." *Journal of International Economics* 65: 1–20.
- Yeung, Henry Wai-chung. 2001. "Organizing Regional Production Networks in Southeast Asia: Implications for Production Fragmentation, Trade and Rules of Origin." *Journal of Economic Geography*, 1(3): 299–321.
- Yi, Kei-Mu. 2003. "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy* 111(1): 52–102.
- . 2005. "Vertical Specialization and the Border Effect." Federal Reserve Bank of Philadelphia no. 05-24.
- Yoffie, David B., and Benjamin Gomes-Casseres. 1994. *International Trade and Competition*. New York: McGraw-Hill.
- Young, Alwyn. 2000. "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China." *Quarterly Journal of Economics* 115(4): 1091–135.
- Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse. 2011. "Monopolistic Competition in General Equilibrium: Beyond the CES." CORE Discussion Papers 2011010. Université Catholique de Louvain.
- . 2012. "Monopolistic Competition: Beyond the Constant Elasticity of Substitution." *Econometrica* 80(6): 2765–84.
- Zhu, Susan Chun. 2004. "Trade, Product Cycles and Inequality within and between Countries." *Canadian Journal of Economics* 37(4): 1042–60.
- . 2005. "Can Product Cycles Explain Skill Upgrading?" *Journal of International Economics* 66: 131–35.
- Zhu, Susan Chun, and Daniel Trefler. 2005. "Trade and Inequality in Developing Countries: A General Equilibrium Analysis." *Journal of International Economics* 65: 21–48.

INDEX

- Abramowitz, Moses, 333
absolute advantage, 3, 73
ad valorem tariffs, 180, 226–27, 230, 232–33, 237, 240, 243, 256, 263, 269, 281, 311, 316–17
Airbus, 294–96
Akerlof, George, 195
Amiti, Mary, 106, 115, 243
Anderson-Van Wincoop theorem, 140–44, 174–75
Antràs-Helpman model, 378–79, 382–86, 391
Arkolakis, Costinot, Rodríguez-Clare model, 156, 166–68, 171, 173, 181–83
asset specificity, 378
asymmetric information, 392–93
Atkeson, Andrew, 241–42
autarky: comparative advantage and, 2; endogenous growth and, 339–52, 359; factor prices and, 190, 194, 208; gains from trade and, 190–200, 207–10; Heckscher-Ohlin model and, 26–28, 45–46; import tariffs and, 248–49; many goods and factors and, 71–72, 75; monopolistic competition and, 122, 130–32, 157–67, 183–84; partial reform of tariffs and, 196–98; two-sector models and, 2–3; utility maximization and, 190
automobiles: discrete choice models and, 424; gains from trade and, 195; import quotas and, 258–60, 270, 274, 276–77; import tariffs and, 234; measuring quality in, 274–84; voluntary export restraints (VERs) and, 258
Autor, David H., 117

Bagwell, Kyle, 293, 328
Bagwell-Staiger theorem, 313–17
Baldwin, Richard E., 34–35, 114, 245
Bandyopadhyay, Usree, 307, 328
Barro, Robert J., 353
Basco, Sergi, 107–8
Ben-David, Dan, 354
Berman, Eli, 86–87
Bernheim, B. Douglas, 304–5
Bernstein-Weinstein condition, 70–71
Berry, Steven T., 282, 284, 419, 422–25
Bertrand competition: duopolies and, 230–34, 252, 290–92; export subsidies and, 290–92; imperfect competition and, 290–92; import quotas and, 261–62, 287, 290–93, 295, 297; import tariffs and, 219, 227, 230–34, 243, 250, 252, 254–55; monopolistic competition and, 181

Bhagwati, Jagdish N., 70, 256, 260–61, 318, 337–39
bilateral trades: Heckscher-Ohlin model and, 33, 44, 46–47; import quotas and, 271; intermediate inputs and, 115; monopolistic competition and, 132, 138, 141, 178; multinationals and, 398
blue collar labor, 84, 86
Boeing, 294–96
Bound, John, 86–87
Bowen, Harry P., 35, 38–39
Brander, James A., 214, 248
Brander-Spencer theorem, 226, 292–93
Branstetter, Lee, 323, 355–56
Broda, Christian, 129–31, 152, 167, 221, 223, 244
Burgess, David F., 63–64
Burstein, Ariel, 241–43

Caliendo, Lorenzo, 206
Canada: abundance of natural resources in, 28; free-trade agreement with United States and, 4, 123–25, 133–36, 187, 195, 203, 205–7, 211; gains from trade and, 187, 203, 205–6; gravity equation and, 133–36; Heckscher-Ohlin model and, 47; monopolistic competition and, 119, 123–25, 133–36, 141–43, 148, 177t, 178–79; NAFTA and, 187, 195, 203, 205–7, 211; rules of origin (ROO) and, 203
Canada-U.S. Free Trade Agreement (CUSFTA), 187, 205–7
cannibalization, 266–68, 270
capital: adjusted net exports and, 34; capital-abundant countries and, 10–11, 20, 22, 29, 32, 34–35, 92, 301–2, 320, 382; fully mobile, 4; gains from trade and, 195–96; Heckscher-Ohlin model and, 25–38, 42, 47, 49–50; import tariffs and, 214, 244–45; induced wage changes and, 7; Leontief's paradox and, 25, 28–35, 37, 39, 42, 48; many goods and factors and, 51, 56–59, 63–70, 76–82; Rybczynski theorem and, 14–18; two-sector models and, 1, 4, 7–24
capital flows: intermediate inputs and, 92, 94, 98; multinationals and, 361–65, 398–401
capital gains, 341–42
Carmichael, Calum M., 292
Carr, David L., 373–74
Caves-Christensen-Diewert theorem, 408–9, 414
Chamberlin, Edward, 119
Chaney, Thomas, 155–56, 163–74, 180, 182–83

- Chicken War, 235
- China: foreign investment in, 322–28; gains from trade and, 186; import quotas and, 260, 287; import tariffs and, 216–17; model of processing trade and, 383, 385; multinationals and, 383, 385, 387–88, 390–92, 398; offshoring and, 107; political economy and, 300, 322–28, 330; processing trade model and, 383; property-rights model and, 387–91; “protection for sale” model and, 323, 328; United States and, 77–78; wages and, 20
- Choi, Yong-Seok, 45, 47–48
- c.i.f. (cost, insurance, freight) prices: import quotas and, 265, 270, 272; import tariffs and, 246
- Cobb-Douglas preferences, 71, 90–91, 103–4, 116, 130, 233, 380–82, 385, 394, 401, 428
- Coe, David T., 354–55
- commodity taxes, 193–97, 210, 213, 217
- comparative advantage: Deardorff’s test of, 45–46; gains from trade and, 210; Heckscher-Ohlin model and, 45–46; import quotas and, 286; intermediate inputs and, 107, 113–14; many goods and factors and, 77, 80; two-sector models and, 2–3
- compensation principle, 197
- constant elasticity of substitution (CES): discrete choice models and, 419, 427–29; endogenous growth and, 331, 340, 343, 345, 348–51; import quotas and, 264, 266–68, 270–72, 277; import tariffs and, 223, 241–44; monopolistic competition and, 124–26, 136, 144–45, 148, 155, 157–58, 161, 163–67, 170–74, 181–82; multinationals and, 366, 370–73, 375, 381; political economy and, 316, 323–24; two-sector models and, 23
- consumer welfare, 303–4, 307, 326, 328, 401, 418
- consumption: autarky and, 193; gains from trade and, 190–93, 197, 201, 203–4, 208, 210, 212–13; Heckscher-Ohlin model and, 26–27, 31–33, 37, 43, 45, 49–50; Heckscher-Ohlin-Vanek (HOV) model and, 31–39; home market effect and, 144–48; import quotas and, 261, 263–65; import tariffs and, 218, 224, 226–28, 235, 237–38, 251, 253; indexes and, 415, 418
- consumption gains, 192
- contaminated estimates, 35
- continuum of goods: Heckscher-Ohlin model and, 75–77; import quotas and, 286, 296; intermediate inputs and, 84, 92, 95, 109, 113, 115; many goods and factors and, 52, 70–77, 80–81; monopolistic competition and, 133, 148, 173. *See also* Ricardian model
- corner solutions, 68
- cost curves: intermediate inputs and, 101, 105; iso-cost curves and, 8–9, 13–14, 19, 101, 105; many goods and factors and, 53; monopolistic competition and, 119, 121–22; two-sector models and, 8–9, 13–14, 19; unit, 53
- Costinot, Arnaud, 109–14, 156, 166–68, 171, 173, 181–83
- cost of living: indexes and, 414–17; monopolistic competition and, 126, 128, 130, 161; terms of trade and, 414–17
- Cournot competition: duopolies and, 226–30, 249, 252, 287–90; export subsidies and, 287–90; imperfect competition and, 287–90; import quotas and, 257, 261, 287–90, 292–97; import tariffs and, 219, 226–27, 230, 234, 245, 248–49, 252, 254; multinationals and, 373
- Cournot-Nash competition, 257
- Crinò, Rosario, 106–7
- customs unions, 187, 189t, 199–205, 210, 317
- Das, Satya P., 269n15
- Dasgupta, P., 244
- Davis, Donald R., 43–44, 48, 115, 147–48
- deadweight loss: import quotas and, 256, 259–60, 284–85, 296; import tariffs and, 221, 239, 252; multinationals and, 361, 365; political economy and, 328
- Deardorff, Alan V., 45–46, 70, 306
- demand curves: aggregate, 117; import quotas and, 257, 260–61, 284–85; import tariffs and, 220–29, 232, 240–41, 246, 248, 254; intermediate inputs and, 86, 92, 117; linear, 120, 181; monopolistic competition and, 120, 126, 129, 149–51, 181; multinationals and, 263; two-sector models and, 2 demand functions, 27, 50, 230–31, 240–41, 264, 329, 336
- Diewert theorems, 406–9, 414–17
- Dinopoulos, Elias, 281
- discrete choice models: automobiles and, 424; China and, 429; competition and, 419; with continuous quantities, 425–29; elasticity and, 422, 427–29; imperfect competition and, 419; Japan and, 424; logit demand system and, 421–22; McFadden theorem and, 420–25; nested logit demand system and, 422–25; outsourcing and, 429; price indexes and, 428–29; relative demand and, 428; subsidies and, 424; utility and, 419–28; welfare and, 421, 427
- Dixit, Avinash, 18, 52, 69, 119–20, 155–56, 186–87, 193–97, 210, 213, 293, 367
- Dollar, David, 354
- Donnenfeld, Shabtai, 269n15
- dumping, xi; antidumping and, 188t, 214–17, 221, 249–53, 260, 292, 365; Feenstra-Markusen-Rose theorem and, 247–48; gains from trade and, 188; import quotas and, 260, 292; import tariffs and, 214–17, 221, 245–54; multinationals and, 365; predatory, 245; reciprocal, 214, 245–48; Weinstein theorem and, 247–48
- Dutch Disease, 16
- Dutt, Pushan, 301–2, 310, 328
- Easterly, William, 353
- Eaton-Grossman theorem, 292–93
- Eaton-Kortum model, xi, 81, 156, 172–74, 180–82, 184, 206–7

- efficiency: endogenous growth and, 346–47; gains from trade and, 192–94, 199–200, 204; Heckscher-Ohlin model and, 40–41, 49; import tariffs and, 219, 229, 244, 249, 251; intermediate inputs and, 113; many goods and factors and, 70; monopolistic competition and, 119, 123, 149; multinationals and, 378, 381–82; Pareto distribution and, 401 (*see also* Pareto distribution); political economy and, 313–16, 328; production costs and, 4
- endogenous growth, xii; autarky and, 339–52, 359; differentiated intermediate inputs and, 340; efficiency and, 346–47; equilibrium and, 331, 333, 337, 340, 342–45, 348, 350–51, 354; factor endowments and, 339; factor prices and, 338, 357; Feenstra theorem and, 346–47; GDP function and, 335, 338–39, 357; gravity equation and, 354; Heckscher-Ohlin model and, 357; heterogeneous firms and, 347–48; homogeneity and, 333; imperfect competition and, 332–33; inequality and, 338, 384; intermediate inputs and, 335, 340–47, 356, 358; Pareto distribution and, 348–49; research and development (R&D) and, 342, 344–47, 354–56; returns to scale and, 332–34, 340, 356–57; Rybczynski effects and, 338, 354; Sampson theorem and, 352–53, 357; scale effects and, 332, 344, 347–52, 357; selection effect and, 348; small countries and, 344, 346–47, 356; spillovers and, 332, 343–44, 346–47, 353–56; Stolper-Samuelson theorem and, 354; total factor productivity (TFP) and, 333, 339, 354–56
- endogenous lobbies, 307–10
- Ethier, Wilfred, 207
- European Economic Community (EEC), 187, 204–5, 211, 235
- exchange rates: endogenous growth and, 335; import quotas and, 282; import tariffs and, 234–38, 243–44, 252; monopolistic competition and, 151, 154; multinationals and, 360
- export promotion, 214, 245, 287
- export subsidies: Bertrand duopoly and, 290–92; commercial aircraft and, 293–96; Cournot duopoly and, 287–90; imperfect competition and, 287–92; many goods and factors and, 285–87; perfect competition and, 284–85
- factor bias, 87–91
- factor endowments: changes in, 13–14; endogenous growth and, 339; Heckscher-Ohlin model and, 25–26, 35, 37–38, 40–42, 45, 47–48; indexes and, 409–10, 412, 414–16; many goods and factors and, 51–52, 56–57, 61–62, 64–65, 68, 76, 78–80; multinationals and, 367–68, 374; political economy and, 319; Rybczynski theorem and, 14–18; two-sector models and, 1, 8, 10, 14–15, 18–20, 22
- factor intensity reversals (FIRs): many goods and factors and, 53; two-sector models and, 8–10, 14, 18, 20, 23; United States and, 9–10
- factor payments, 194n1
- factor price equalization (FPE): Heckscher-Ohlin model and, 40, 45–48; many goods and factors and, 51–52, 55, 68–69, 75–76; multinationals and, 365–70, 374; statement of, 10–11; two-sector models and, 10–11, 16, 18–19, 21
- factor prices: autarky and, 190, 194, 208; commodity taxes and, 193–96; determination of, 8; equalization theorem and, 10–11; gains from trade and, 190, 193–94, 197; gradient vectors and, 8–9, 14; Heckscher-Ohlin model and, 25–26, 28, 30, 32, 40–41, 46–47, 361; import quotas and, 265; import tariffs and, 238; insensitivity and, 8–10; intermediate inputs and, 87–88, 90, 92, 95–97, 99, 101, 103, 113, 116–17; Lerner diagram and, 21–23, 357; magnified effects and, 12; many goods and factors and, 51–56, 58, 61–62, 64, 68–70, 75–80; two-sector models and, 6–23; weighted average and, 12
- factor requirements matrix, 40–42
- Feenstra, Robert C., xi–xii; dumping and, 247–48; elasticity of substitution and, 151–52; endogenous growth and, 346–47; export subsidies and, 286; gains from trade and, 195; generalized method of moments (GMM) and, 129; import quotas and, 267–74, 278, 281, 286; import tariffs and, 221, 223, 244; intermediate input theorem of, 84, 91, 93–95, 97, 99, 103–8, 113–15; monopolistic competition and, 127–28; moving subsidies and, 195; multinationals and, 377–78, 383–84, 390–91; political economy and, 323; translog expenditure function and, 181; U.S. competitive gains and, 182–83; voluntary export restraints (VERs) and, 278, 281
- Fernandes-Tang theorem, 386–91
- Fernandez, Raquel, 302
- final tariff vector, 199
- Fisher's ideal price index, 403–7, 410, 413, 417
- f.o.b. (free on board) prices, 136, 139–40, 246, 250, 265, 270–72, 370
- Ford Motor Company, 234, 274
- foreign direct investment (FDI), 360–61, 365, 370, 375–76, 398
- Frankel, Jeffrey A., 211, 354
- Fréchet distribution, 156, 168, 172–74, 180, 182, 184
- free-trade agreements (FTAs): CUSFTA, 187, 205–7; European Economic Community (EEC), 187, 204–5, 211, 235; gains from trade and, 187, 202–6, 211; Grinols-Silva theorem and, 203; import tariffs and, 253; Krishna-Panagariya theorem and, 203; MERCOSUR and, 187, 205; monopolistic competition and, 119, 123–25, 180; NAFTA, 187, 195, 203, 205–7, 211; political economy and, 322, 329; rules of origin (ROO) and, 203; two-sector models and, 4; U.S.–Canada, 4, 123–25, 136, 187, 195, 203, 205–7, 211
- Fujita, Masahisa, 148
- functional forms, 239–44

- gains from trade: autarky and, 190–200, 207–10; automobiles and, 195; budget constraints and, 190–91, 196–98, 207–8; Canada and, 187, 203, 205–6; capital and, 195–96; change in tariff revenue and, 199; China and, 186; commodity taxes and, 193–97, 210, 213; comparative advantage and, 210; compensation principle and, 197; competition and, 187, 195, 204, 207–11; consumption and, 190–93, 197, 201, 203–4, 208, 210, 212–13; customs unions and, 187, 189t, 199–205, 210; dumping and, 188t; Eaton-Kortum model and, 206–7; efficiency and, 192–94, 199–200, 204; factor prices and, 190, 193–94, 197; final tariff vector and, 199; free-trade agreements (FTAs) and, 187, 202–6, 211; GATT and, 186–87, 188t–189t, 202, 214–17, 223, 252–53; Grinols theorems and, 198–99, 203, 209–10; harmonic mean and, 131–32; Heckscher-Ohlin model and, 186; imperfect competition and, 207–9; import variety and, 125–32; inequality and, 192, 198, 209; Ju-Krishna theorem and, 198–99; Kemp-Wan rule and, 200–202; Korea and, 205, 253; Krishna-Panagariya theorem and, 203; many goods and factors and, 72–75; Melitz-Chaney model and, 164–66; Mexico and, 187, 203, 206; partial reform of tariffs and, 196–98
- Gawande, Kishore, 307, 328
- GDP function: endogenous growth and, 335, 338–39, 357; estimating, 60–66; Heckscher-Ohlin model and, 46–48; indexes and, 410–16; invisible hand and, 5; many goods and factors and, 51–53, 55, 57, 60–66, 68, 80–81; relative price of good and, 5–6; two-sector models and, 5–7; United States and, 51n1, 63–64, 80
- GDP price index, 409; Caves-Christensen-Diewert theorem and, 414; constant technology and, 411; cross-country real GDP measurement and, 417–18; Diewert-Morrison theorem and, 415–17; Fisher's ideal price index and, 410, 413, 417; Gerschenkron effect and, 411–12; Laspeyres price index and, 410–12, 414–15; Paasche price index and, 410–12, 414–15; production possibility frontiers (PPFs) and, 411–12, 414, 417; production quantities and, 418; terms of trade and, 414–17
- General Agreement on Tariffs and Trade (GATT): gains from trade and, 186–87, 202, 214–17, 223, 252–53; home market effect and, 316–17; key provisions of, 188t–189t; most favored nation (MFN) provision and, 187, 223, 253, 299, 317; political economy and, 299, 311, 314–17, 328
- generalized method of moments (GMM), 129, 149, 151, 154
- Germany: Chicken War and, 235; coal exports of, 81; endogenous growth and, 355; Heckscher-Ohlin model and, 47; import quotas and, 294; import tariffs and, 235; MERCOSUR and, 205; offshoring and, 106n17; steel imports of, 81; unification of, 129n15, 149, 186, 195–96
- Gerschenkron effect, 411–12
- globalization, 83, 107–8
- Goldberg, Pinelopi, 306–7, 328
- government: gains from trade and, 186, 188t, 190–98, 201, 203, 208–9, 213; import quotas and, 257–59, 263, 276, 286–87, 290, 292–94, 297; intermediation activities and, 418; many goods and factors and, 62–63; multinationals and, 399; political economy and, 303–5, 307, 311–16, 323–28; services of, 62–63
- gradient vectors, 8–9, 14
- Graham, Frank, 119, 207
- gravity equation: Anderson-Van Wincoop theorem and, 140–44, 174–75; Arkolakis, Costinot, Rodríguez-Clare model and, 156, 171, 173, 181–83; asymmetric effects and, 134; border effects and, 136–40; endogenous growth and, 354; estimation of, 137–40, 177–79; fixed effects and, 143–44; Head-Mayer theorem and, 169–81; Heckscher-Ohlin model and, 44n20; home market effect and, 144–48; iceberg transport costs and, 139, 141, 161, 168, 171–72, 174, 180; import quotas and, 272; Krugman theorem and, 146–47; monopolistic competition and, 132–56, 168–81; multinationals and, 398; national product differentiation and, 147–48; Newton's law of gravity and, 133; Poisson estimator and, 175–79; political economy and, 329; price index data and, 137–39; Probit equation and, 174; pseudo maximum likelihood (PML) estimator and, 176–79; RESET test and, 177–79; trade cost estimation and, 180–81; U.S.-Canada trade and, 133–36
- Griliches, Zvi, 86–87
- Grinols, Earl L., 186
- Grinols-Silva theorem, 203
- Grinols theorem, 209–10
- Grinols-Wong theorem, 198–99
- Groizard, José L., 115
- Grossman, Gene M., xi; Eaton-Grossman theorem and, 292–93; endogenous growth and, 331–32, 340, 344, 347, 349–50, 354; intermediate inputs and, 99–104, 107, 114–15; political economy and, 303, 306–7, 323, 328; “protection for sale” model and, 223, 299, 303–7, 310–13, 315, 318, 323, 328
- Grossman-Helpman theorem, 313–17
- Hamilton, Alexander, 244
- Hamilton, Gary G., 377–78
- Hanson, Gordon: intermediate inputs and, 84, 91, 93–95, 97, 99, 104, 107–8, 113–15, 117; multinationals and, 383–84, 390–91
- Harley-Davidson, 235, 239
- harmonic mean, 131–32
- Harrigan, James, 56–57, 65–66, 80
- Head, Keith C., xi, 244
- Head-Mayer theorem, 156; Eaton-Kortum model and, 172–74; gravity equation and, 169–81;

- Krugman model and, 170; Melitz-Chaney model and, 170–71; Poisson estimator and, 175–79; Probit equation and, 174; pseudo maximum likelihood (PML) estimator and, 176–79; RESET test and, 177–79; trade cost estimation and, 180–81; zeros in trade and, 174–77
- Heckman estimator, 174
- Heckscher, Eli, 25, 83
- Heckscher-Ohlin model: abundant country factors and, 28; autarky and, 26–28, 45–46; capital and, 25–38, 42, 47, 49–50; capital flows and, 363–65; comparative advantage and, 45–46; competition and, 35; continuum of goods and, 75–77; differences in tastes and, 49; efficiency and, 40–41, 49; equilibrium and, 26–28, 46; factor endowments and, 25–26, 35, 37–38, 40–42, 45, 47–48; factor price equalization (FPE) and, 40, 45–48; factor price insensitivity and, 361; factor prices and, 25–26, 28, 30, 32, 40–41, 46–47; foreign technologies and, 29; gains from trade and, 186; GDP function and, 46–48; inequality and, 46–47; Leontief's paradox and, 25, 28–35, 37, 39, 42, 48, 63; sign patterns and, 35, 37–38
- Heckscher-Ohlin-Samuelson (HOS) model, 25–26
- Heckscher-Ohlin-Vanek (HOV) model: adjusted net exports and, 34; bilateral trades and, 33, 44, 46–47; complete test of, 38–39; consumption and, 31–39; contaminated estimates and, 35; different technologies across countries and, 40, 43–44; factor requirements matrix and, 40–42; Leamer theorem and, 31–39; many goods and factors and, 78–79; other tests of trade and, 45–48; partial tests of, 33–38; production and, 31–39; rank tests and, 37–39, 50; Rybczynski effects and, 35–37; sign tests and, 37–38, 43, 48, 50; Trefler theorem and, 40–43, 49–50; Trefler-Zhu theorem and, 44–45
- Helpman, Elhanan: Antràs-Helpman model and, 378–79, 382–86; endogenous growth and, 331–32, 340, 344, 347, 349–50, 354–55; gains from trade and, 207, 209; intermediate inputs and, 115; monopolistic competition and, 119–20, 174; multinationals and, 368, 370, 374, 376; political economy and, 303, 306–7, 323, 328; “protection for sale” model and, 223, 299, 303–7, 310–13, 315, 318, 323, 328
- heterogeneous firms, 115; endogenous growth and, 347–49, 353, 356, 359; gains from trade and, 211; imports and, 241, 268, 271; monopolistic competition and, 155–58, 167, 182–83; multinationals and, 375–76
- heteroskedasticity, 129, 151, 175, 178–79
- Hicks-neutral productivity, 65, 90–91, 103, 338, 357
- holdup problem, 378
- home market effect, 144–48, 316–17
- Honda, 237
- Hopenhayn, Hugo A., 156
- Horstmann, Ignatius J., 230, 391–92
- Hsieh, Chang-Tai, 117
- Huang, Deng-Shing, 377
- iceberg transport costs, 100, 136, 139, 141, 161, 168, 171–72, 174, 180, 210, 245–46, 316, 350, 370, 375
- immiserizing growth, 335–37
- imperfect competition: Bertrand duopoly and, 290–92; Cournot duopoly and, 287–90; discrete choice models and, 419; endogenous growth and, 332–33; export subsidies and, 287–92; gains from trade and, 207–9; import quotas and, 256, 260–63, 287, 296; import tariffs and, 214, 218–19, 223, 229, 234, 245, 252–53; monopolistic competition and, 181, 187, 204, 207–10; political economy and, 299; returns to scale and, 207–9; sheltered markets and, 260
- imperfect information, 377, 391–97
- import quotas: automobiles and, 258–60, 270, 274–84; Bertrand duopoly and, 290–92; bilateral trades and, 271; Brander-Spencer theorem and, 292–93; budget constraints and, 264, 297; cannibalization and, 266–68, 270; China and, 260, 287; c.i.f. prices and, 265, 270, 272; comparative advantage and, 286; continuum of goods and, 286, 296; Cournot duopoly and, 287–90; dead-weight loss and, 256, 259–60, 284–85, 296; difference from tariffs, 256; dumping and, 260, 292; equivalence and, 256–60; exchange rates and, 282; licenses and, 257–58; lobbies and, 258–59; Multi-Fiber Arrangement (MFA) and, 259–60; multiproduct firm pricing and, 266–68; non-equivalence and, 260–69; pass-through and, 281; quality and, 263–84; rent seeking and, 248, 260; textiles and, 258–60; trade policy and, 256–57, 296–97; voluntary export restraints (VERs) and, 258, 259n4, 262–63, 268–69, 274–84; welfare and, 256–61, 281, 284–93, 296–97
- import substitution, 213, 221, 361
- import tariffs: autarky and, 248–49; automobiles and, 234; Chicken War and, 235; China and, 216–17; Cournot duopoly and, 226–30; dumping and, 214–17, 221, 245–54; efficiency and, 219, 229, 244, 249, 251; estimating terms of trade effects and, 236–39; exchange rates and, 234–38, 243–44, 252; factor prices and, 238; foreign monopolies and, 224–26; free-trade agreements (FTAs) and, 253; functional forms and, 239–44; GATT and, 215–17; imperfect competition and, 214, 218–19, 223, 229, 234, 245, 252–53; import variety and, 244; Japan and, 234–39, 245, 253; Japanese trucks/motorcycles and, 234–36; pass-through and, xi, 224–25, 230, 235–44, 251–52; small countries and, 219–21, 223; terms of trade and, 221–39, 249–53; trade policy and, 214, 229, 234; truck/motorcycle tariffs and, 214–15, 234–39, 244, 252, 255; United States and, 215–17, 223, 234–39, 244–45, 249–53; utility and, 218, 220, 230, 233–34, 239–41, 253; welfare and, 214,

- import tariffs (*continued*)
 217–22, 225–34, 239, 244, 248–52; World Trade Organization (WTO) and, 215–17
- import variety: Broda-Weinstein model and, 129–30; gains from trade and, 125–32; harmonic mean and, 131–32; import tariffs and, 244; intermediate inputs and, 97; monopolistic competition and, 125–32, 149, 164, 166, 182
- income distribution, 302, 308–10
- inequality: endogenous growth and, 338, 384; gains from trade and, 192, 198, 209; Heckscher-Ohlin model and, 46–47; import tariffs and, 248, 251; intermediate inputs and, 90, 103–4, 108, 113, 115; many goods and factors and, 58; political economy and, 301–2, 310; two-sector models and, 3, 12
- infant industry, 214, 239, 244–45, 293
- input price indexes, 404–9
- intermediate inputs, xi; autarky and, 87, 115–17; automobiles and, 97, 115; bilateral trade and, 115; capital and, 87–98, 104, 106, 113, 117; China and, 105–7, 115–16; comparative advantage and, 107, 113–14; continuum of goods and, 84, 92, 95, 109, 113, 115; efficiency and, 113; factor prices and, 87–88, 90, 92, 95–97, 99, 101, 103, 113, 116–17; Feenstra theorem and, 103–8; gains from trade and, 206; Heckscher-Ohlin model and, 29–30, 45, 83–84, 105, 113; import tariffs and, 243; import variety and, 97; Leontief preferences and, 91, 99, 110–11; manufacturing and, 84–91, 95, 98–99, 105–7, 114; many goods and factors and, 63, 70, 81; Mexico and, 86–87, 92–93; monopolistic competition and, 144, 183; multinationals and, 361, 369, 378; nonproduction workers and, 84–87, 89–90, 96–99, 104–7, 114, 117–18; offshoring and, 83–84, 87, 92–95, 97–109, 113–16; Ricardian model and, 83, 95, 115; Rybczynski effects and, 103–4; sector bias and, 87–91; Stolper-Samuelson theorem and, 88, 90, 103; supermodular production and, 108–13; technological change and, 84, 86–88, 90–91, 95, 98–99, 101, 104; terms of trade and, 115–16; total factor productivity (TFP) and, 87–90, 95–97; United States and, 83–84, 86–87, 89–93, 97n12, 101, 104–8, 113; wages and, 83–118; zero-profit conditions and, 87–88, 95, 100–101, 105, 116
- International Trade Commission (ITC), 215–17, 234–35, 237, 239n18, 249–50, 274
- intra-industry trade, 132, 148, 248, 369
- investment: endogenous growth and, 339; foreign direct (FDI), 360–61, 365, 370, 375–76, 398; gains from trade and, 196; Heckscher-Ohlin model and, 43; intermediate inputs and, 97–98; many goods and factors and, 62–64, 66; multinationals and, 360–65, 369, 373–75, 398; political economy and, 322–24
- invisible hand, 5
- Irwin, Douglas A., 293–96
- Isuzu, 238
- Itoh, Motoshige, 286
- Itskhoki, Oleg, 115, 243
- Japan: endogenous growth and, 355–56; foreign direct investment and, 360; gains from trade and, 192, 205; Heckscher-Ohlin model and, 46; hedonic regression on exports of, 283t; import quotas and, 256, 258, 268, 270, 274, 276–84; import tariffs and, 234–39, 245, 253; many goods and factors and, 69–70, 80; Meiji restoration and, 46; Ministry of International Trade and Industry (MITI) and, 258, 268, 281; monopolistic competition and, 130t, 131, 147–48; multinationals and, 360, 365, 377; Ossa model and, 130t, 131; research and development (R&D) and, 356; truck/motorcycle tariffs and, 214–15, 234–39, 244, 252; voluntary export restraints (VERs) and, 365, 424; WTO protests and, 253
- joint ventures, 360–61, 363, 377, 393–97, 399
- Jones-Scheinkman theorem, 56–57, 59
- Ju-Krishna theorem, 198–99
- Keller, Wolfgang, 355
- Kemp, Murray C., 187, 200–202, 203n6
- Kiyono, Kazuharu, 286–87
- knowledge-capital model, 373–75
- Kohli, Ulrich R., 61–65, 80
- Konings, Jozef, 243
- Korea: compensation principle and, 197; gains from trade and, 205, 253; Heckscher-Ohlin model and, 47; many goods and factors and, 81; multinationals and, 377–78; political economy and, 302; two-sector models and, 20; WTO protests and, 253
- Kortum, Samuel, xi, 81, 156, 172–74, 180–82, 184, 206–7
- Kraay, Aart, 354
- Kreinin, Mordechai E., 281
- Kremer, Michael, 108–9
- Krishna, Kala, 186, 197–99, 256, 261, 263, 269n15, 281
- Krishna, Pravin, 45, 47–48, 202–3, 206, 211, 299, 328–29
- Krishna-Panagariya theorem, 203
- Krueger, Alan B., 90, 203
- Kreuger, Anne O., 258
- Krugman, Paul R.: gains from trade and, 207, 209–11; Head-Mayer theorem and, 170; import quotas and, 268, 293; import tariffs and, 214, 226–27, 240, 245–46, 248; intermediate inputs and, 90, 102–3; monopolistic competition and, 120–25, 129–30, 136–37, 144, 146–47, 155, 158, 160, 163–64, 167, 170–71, 174, 180–83; monopolistic competition model and, 322, 344, 347; reciprocal dumping model of, 214
- Krugman theorem, 123, 146–47
- Kyle, A. S., 293
- labor: adjusted net exports and, 34; blue collar, 84, 86; endowments and, 8, 15–17, 20, 26,

- 31–32, 50, 58, 60, 70, 78, 81, 117, 146, 301, 319–22, 343, 395, 397; factor bias and, 87–91; gains from trade and, 195–96; Heckscher-Ohlin model and, 25–36, 38, 40–42, 47, 49–50; import quotas and, 265, 286; import tariffs and, 218; induced wage changes and, 7 (*see also wages*); labor-abundant countries and, 10–11, 20–22, 26, 49, 77, 92, 301–2, 322, 364, 367, 374, 395; Leamer theorem and, 31–39; Leontief's paradox and, 25, 28–35, 37, 39, 42, 48; low-skilled, 84, 88, 90–92, 95, 99–105, 108, 113, 117, 366–70; nonproduction workers and, 84–87, 89–90, 96–99, 104–7, 114, 117–18; offshoring and, 75; Rybczynski theorem and, 14–18; sector bias and, 87–91; skilled, 29, 34, 49, 84, 86, 88, 90–95, 100–101, 103–5, 108, 113, 116–17, 366–70, 373–74; skilled-labor abundant countries and, 49, 92, 367, 374; supply of, 84, 86, 104, 218, 342; two-sector models and, 1–4, 7–22, 24; unemployment and, 115–16, 123, 195–96, 234, 274; unions and, 34, 144, 195, 215, 234–35, 274, 299; United States and, 83–93, 97n12, 101, 104–8, 113; unskilled, 30, 56–57, 369–70, 373–74; white collar, 84, 106
- labor-abundant countries, 10–11, 20–22, 26, 49, 77, 92, 301–2, 322, 364, 367, 374, 395
- Lagrangians, 24, 120, 264, 268–69, 277, 358
- Lancaster, Kelvin, 119–20
- Laspeyres price index, 403–7, 410–12, 414–15
- Leamer, Edward E., 87, 89–90, 101
- Leontief preferences: foreign technologies and, 29; input-output accounts and, 28–29; intermediate inputs and, 91, 99, 110–11; technology and, 91; trade in tasks and, 99
- Leontief production function, 76
- Leontief's paradox, 25, 28–35, 37, 39, 42, 48, 63
- Lerner diagram, 21–23, 357
- Lerner index, 267
- Levine, Ross, 353, 356
- Levinsohn, James, 282, 284, 335
- Levy, Philip I., 318, 328
- Levy theorem, 321–22
- Lewis, Tracy R., 193, 195
- licenses, 189t, 257–58
- lobbies: cost of forming, 309–10; endogenous, 307–10; group entry and, 308–9; import quotas and, 258–59; income distribution and, 308–10; political economy and, 299–300, 303–10, 328–29; welfare and, 308–9
- lump sum transfers: autarky and, 190–98, 210; consumption and, 190–93, 197, 201, 203, 208, 210; gains from trade and, 186–87, 190–98, 201, 203, 208, 210, 212–13; import tariffs and, 217; partial reform of tariffs and, 196–98
- Ma, Hong, 268
- McFadden theorem, 263–64, 420–25
- Machin, Stephen, 87
- Maggi, Giovanni, 293, 306–7, 323, 328
- many goods and factors: autarky and, 71–72, 75; Bernstein-Weinstein condition and, 70–71; comparative advantage and, 77, 80; continuum of goods and, 52, 70–77, 80–81; corner solutions and, 68; export subsidies and, 285–87; factor endowments and, 51–52, 56–57, 61–62, 64–65, 68, 76, 78–80; factor intensity reversals (FIRs) and, 53; factor price equalization (FPE) and, 51–52, 55, 68–69, 75–76; factor prices and, 51–56, 58, 61–62, 64, 68–70, 75–80; gains from trade and, 72–75, 186; GDP function and, 51–53, 55, 57, 60–66, 68, 80–81; Heckscher-Ohlin-Vanek (HOV) model and, 78–79; Jones-Scheinkman theorem and, 56–57, 59; more factors than goods and, 57–60; more goods than factors and, 67–70; product prices and, 51, 54, 59, 61, 68–69; Ricardian model and, 52, 70–75, 77, 80–81; Ricardo-Viner model and, 51, 57; Rybczynski effects and, 51, 55–57, 60, 62, 64, 66; same number of goods and factors and, 53–54; Samuelson-Nikaido lemma and, 54–56; Stolper-Samuelson theorem and, 51–57, 61–65; subsidies and, 70, 285–87; tariffs and, 64; technological change and, 63, 80; terms of trade and, 73–75; trade in goods and, 70, 81; welfare and, 72–75, 80; zero-profit conditions and, 52–53, 55, 57, 59, 67, 82
- Markusen, James R., 207, 230, 247–48, 370, 372–75, 391–92
- Maskin, Eric, 108–9
- Maskus, Keith E., 373–74
- matching, 393–97
- Mayer, Thierry, 156, 168. *See also Head-Mayer theorem*
- median voter model: capital-abundant countries and, 301–2; capital/labor ratio and, 301; Heckscher-Ohlin model and, 318–19, 321–22; majority vote and, 300–303; nontariff barriers and, 306–7; political economy and, 299–303, 306, 310–12, 315, 318–22, 328; Roy's Identity and, 301, 303; single peaked preferences and, 300; truthful contribution schedule and, 304–5; truthful Nash equilibrium and, 305; two-country model and, 311–13
- Meiji restoration, 46
- Melitz, Marc J., xi, 149, 156, 174, 182, 244, 376
- Melitz-Chaney model: endogenous growth and, 348; gains from trade and, 164–66; Head-Mayer theorem and, 170–71; monopolistic competition and, 155–56, 163–74, 180, 182–83
- MERCOSUR, 187, 205
- Mestieri, Marti, 107–8
- Mexico: gains from trade and, 187, 203, 206; import tariffs and, 253; intermediate inputs and, 86–87, 92–93; liberalization in, 123n3; multinationals and, 363; NAFTA and, 187, 195, 203, 205–7, 211; political economy and, 307; rules of origin (ROO) and, 203
- Ministry of International Trade and Industry (MITI), 258, 268, 281

- Mitra, Devashish, 115, 299, 301–4, 307–10, 328–29
- Mitsubishi, 238
- monopolistic competition: Anderson-Van Wincoop theorem and, 140–44, 174–75; Arkolakis, Costinot, Rodríguez-Clare model and, 156, 166–68, 171, 173, 181–83; bilateral trades and, 132, 138, 141, 178; Canada and, 119, 123–25, 133–36, 141–43, 148, 177t, 178–79; capital and, 153; c.i.f. prices and, 136, 139–40; continuum of goods and, 133, 148, 173; Eaton-Kortum model and, xi, 156, 172–74, 180–82; efficiency and, 119, 123, 149; exchange rates and, 151, 154; factor prices and, 133, 148, 153; Feenstra theorem and, 127–28; fixed effects and, 143–44; f.o.b. prices and, 136, 139–40; Fréchet distribution and, 156, 168, 172–74, 180, 182, 184; free-trade agreements (FTAs) and, 119, 123–25, 180; gains from import variety and, 125–32; gravity equation and, 132–56, 168–81; harmonic mean and, 131–32; Head-Mayer model and, 156, 169–81; Heckscher-Ohlin model and, 122, 133, 148; heterogeneous firms and, 157–63; ideal variety approach and, 119–20; import variety and, 125–32, 149, 164, 166, 182; intra-industry trade and, 132, 148; Japan and, 130t, 131, 147–48; Krugman model and, 120–25, 129–30, 136–37, 144, 146–47, 155, 158, 160, 163–64, 167, 170–71, 174, 180–83, 322, 344, 347; love of variety approach and, 119–20; Melitz-Chaney model and, 155–56, 163–74, 180, 182–83; Mexico and, 123n3; national product differentiation and, 147–48; Ossa theorem and, 132; Pareto distribution and, 156, 163–68, 171, 173–74, 180, 182–84; returns to scale and, 119, 148–49, 155, 172; Sato-Vartia theorem and, 126–27, 165; scale effect and, 123–25, 155; selection effect and, 123–25, 155, 164, 166–67; transportation costs and, 135–48, 168, 171, 180, 184; United States and, 119, 123–24, 129–35, 141–43, 147–48, 178–79, 182; utility and, 120, 124–31, 136–37, 144–45, 149, 152–53, 157, 164, 181; variable markups and, 181; wages and, 121–22, 126, 144–45, 148–53, 157–60, 163–64, 166, 170–74, 184–85; welfare and, 126, 129–31, 163–67, 182; zero-profit conditions and, 137, 145, 149, 155–62, 170–71, 348, 350–53; zeros in trade and, 174–77
- most favored nation (MFN) provision, 187, 223, 253, 299, 317
- motorcycles, 214–15, 234–39, 244, 252
- Multi-Fiber Arrangement (MFA), 259–60
- multinationals, xii; Antràs-Helpman model and, 378–79, 382–86, 391; asset specificity and, 378; asymmetric information and, 392–93; autarky and, 364–65, 393–97, 400–401; capital flows in one-sector economy and, 362–63; capital flows in two-sector economy and, 363–65; China and, 383, 385, 387–88, 390–92, 398; c.i.f. prices and, 370; deadweight loss and, 361, 365; dumping and, 365; efficiency and, 378, 381–82; exchange rates and, 360; factor endowments and, 367–68, 374; Fernandes-Tang theorem and, 386–91; f.o.b. prices and, 370; foreign direct investment (FDI) and, 360–61, 365, 370, 375–76, 398; gravity equation and, 398; Heckscher-Ohlin model and, 361, 397–98; heterogeneous firms and, 375–76; holdup problem and, 378; horizontal, 361, 365, 370–77, 398; imperfect information and, 377, 391–97; Japan and, 360, 365, 377; joint ventures and, 360–61, 363, 377, 393–97, 399; knowledge-capital model and, 373–75; Korea and, 377–78; labor and, 362–74, 377–78, 391, 393–402; location choice and, 379–82; market power and, 377–78; Markusen theorem and, 370, 372–73; matching firms across countries and, 393–97; Mexico and, 363; offshoring and, 369, 373, 379, 399; OLI framework and, 361; outsourcing and, 373, 379–83, 385, 387, 390, 399; ownership choice and, 376–77, 379–82; Pareto distribution and, 401; property-rights model and, 387–91; tariffs and, 361, 364–65, 399–400; transaction costs and, 361, 377–79, 391; United States and, 360, 363, 365, 374, 382, 398; utility and, 366, 370, 373, 400; vertical, 361, 364–70, 373–81, 383–85, 387, 390–91, 398–99; wages and, 361–63, 366–73, 375, 381–82, 385, 393–98, 401; welfare and, 361–65, 399–401
- Mundell, Robert A., 361, 363, 365, 398
- Nash, John, 379
- Nash bargaining, 379–83, 386–87, 395
- Nash equilibrium, 297, 304–5, 312–15
- national product differentiation, 147–48
- Newton's law of gravity, 133
- Norman, Victor, 18, 52, 69, 186–87, 193–97, 210, 367
- North American Free Trade Agreement (NAFTA), 187, 195, 203, 205–7, 211
- ONET database, 107
- offshoring: China and, 107; globalization and, 83, 107–8; intermediate inputs and, 83–84, 87, 92–95, 97–109, 113–16; many goods and factors and, 75; multinationals and, 369, 373, 379, 399; supermodularity and, 108–13
- Offshoring in the Global Economy: Theory and Evidence* (Feenstra), xii
- Ohlin, Bertil, 25, 83, 119
- Ohlin Lectures, xi
- OLI framework, 361
- Organization for Economic Cooperation and Development (OECD): endogenous growth and, 339, 354; Heckscher-Ohlin model and, 33; import tariffs and, 236; indexes and, 411; many goods and factors and, 51, 56–57, 65, 78–80; monopolistic competition and, 138–39, 147
- organization of the firm, 376–82
- Ossa, Ralph, xii, 116–17, 130–32, 167, 183, 316–17
- outsourcing: discrete choice models and, 429; domestic, 379, 382; foreign, 379, 382; intermediate

- inputs and, 83; multinationals and, 373, 379–83, 385, 387, 390, 399
- Paasche price index, 403–7, 410–12, 414–15
- Pakes, Ariel, 267, 277, 282, 284, 335
- Pareto distribution: endogenous growth and, 348–49; gains from trade and, 186, 193, 195–96, 199, 203, 208–10, 213; import tariffs and, 218; monopolistic competition and, 156, 163–68, 171, 173–74, 180, 182–84; multinationals and, 401
- Parro, Fernando, 206
- partial reform of tariffs, 196–98
- pass-through: complete, 243, 251; constant, 240–41; domestic prices and, 230; elasticity and, 237, 240, 244; import quotas and, 281; import tariffs and, xi, 224–25, 230, 235–44, 251–52; incomplete, 241; level form and, 240; lower, 243–44; nonconstant, 241; partial, 225, 252; symmetric, 236–38
- Pavcnik, Nina, 124, 293–96
- Penn World Table, 270
- perfect competition: endogenous growth and, 332, 335, 341, 356; export subsidies and, 284–85; gains from trade and, 207–8; import quotas and, 256–57, 265, 284–85, 296; import tariffs and, 218–24; indexes and, 409; large countries and, 221–24; many goods and factors and, 70; monopolistic competition and, 133, 156, 172, 181; small countries and, 219–21; two-sector models and, 4–5, 7, 21
- Perry, Martin K., 377
- Poisson estimator, 175–79
- political economy, xii; Bagwell-Staiger theorem and, 313–17; China and, 300, 322–28, 330; customs unions and, 317; efficiency and, 313–16, 328; free-trade agreements (FTAs) and, 322, 329; government and, 303–5, 307, 311–16, 323–28; Grossman-Helpman theorem and, 313–17; Heckscher-Ohlin model and, 300, 302, 318; home market effect and, 316–17; income distribution and, 302, 308–10; Korea and, 302; Levy theorem and, 321–22; lobbies and, 299–300, 303–10, 328–29; median voter model and, 299–303, 306, 310–12, 315, 318–22, 328; Mexico and, 307; most favored nation (MFN) provision and, 187, 223, 253, 299, 317; nontariff barriers and, 306–7; Ossa theorem and, 316–17; “protection for sale” model and, 223, 299, 303–7, 310–13, 315, 318, 323, 328; regional trade agreements and, 317–21; subsidies and, 300–309, 329; tariffs and, 299–318, 323–29; terms of trade and, 311–16, 328; trade policy and, 299–303, 328; United States and, 300, 306–7, 328; utility and, 300–303, 319–20, 322, 324–25, 335–36, 340, 343, 348, 358; wages and, 303–4, 316, 324–26, 327; welfare and, 302–11, 316–20, 322, 326, 328–31; zero-profit conditions and, 316
- poll subsidies, 193, 209, 213
- Prebisch, Raul, 331
- price indexes: Caves-Christensen-Diewert theorem and, 414; changing technology and, 407; constant technology and, 405–6; current period quantities and, 403–4; Diewert theorems and, 406–9; discrete choice models and, 428–29; endogenous growth and, 343, 346, 351, 359; Fisher's ideal, 403–7, 410, 413, 417; GDP, 409–18; Gerschenkron effect and, 411–12; gravity equation and, 137–39; import quotas and, 272, 274, 279; import tariffs and, 241–42; input, 404–9; Laspeyres, 403–7, 410–12, 414–15; monopolistic competition and, 126–27, 136–37, 143, 145, 148, 157–58, 161, 164–66, 170, 173; multinationals and, 371; Paasche, 403–7, 410–12, 414–15; political economy and, 316; quantities in the base period and, 403
- Probit equation, 174
- production possibility frontiers (PPFs): corner solution and, 68; endogenous growth and, 335; gains from trade and, 212; Heckscher-Ohlin model and, 26; indexes and, 411–12, 414, 417; many goods and factors and, 67–68, 80; political economy and, 319; two-sector models and, 2–6, 16–17
- productivity: Diewert-Morrison theorem and, 415–17; endogenous growth and, 332–35, 338–40, 346–53, 355–57; of factors, 40; gains from trade and, 206; Heckscher-Ohlin model and, 37, 40–44, 48, 50; Hicks-neutral, 65, 90–91, 103, 338–39, 357; import quotas and, 265–66, 268, 270–72; import tariffs and, 235, 239; indexes and, 403–5, 407, 409, 411, 413–17; intermediate inputs and, 85–87, 89, 91, 96, 99, 101, 103–4, 106, 113, 116–17; many goods and factors and, 51, 65, 73–75, 79–81; measurement of, 332–35; monopolistic competition and, 123–24, 129, 151, 155–67, 170–72, 174, 181–85; multinationals and, 363, 371, 373, 381; terms of trade and, 414–17; total factor productivity (TFP) and, 87–90, 95–97, 333, 339, 354–56, 369, 403, 409, 416; zero-cutoff-profit (ZCP) condition and, 158–62, 170–71, 348, 350–53
- productivity indexes, 403–4, 409, 414–17
- product prices: changes in, 11–12; changes in endowments and, 13–14; endogenous growth and, 354; Heckscher-Ohlin model and, 26, 31; indexes and, 403–5, 407, 409, 411–13, 415, 417; intermediate inputs and, 88, 90, 96, 117; many goods and factors and, 51, 54, 59, 61, 68–69; multinationals and, 362, 365, 398; political economy and, 303; two-sector models and, 4, 7–8, 10–13, 16–17, 21; wages exceeding, 13
- property-rights model, 387–91
- “protection for sale” model, 223, 299, 303–7, 310–13, 315, 318, 323, 328
- pseudo maximum likelihood (PML) estimator, 176–79
- quality: automobiles and, 274–84; cannibalization and, 266–68, 270; free trade equilibrium and,

- quality (*continued*)
 264–66; import quotas and, 263–84; Lerner index and, 267; McFadden aggregation theorem and, 263–64; measurement of traded goods and, 270–84; multiproduct firm pricing and, 266–68; nonequivalence and, 263–69; voluntary export restraints (VERs) and, 268–69, 274–84; without characteristics, 270–74
- Ranjan, Priya, 115
- rank tests, 37–39, 50
- Rauch, James E., 147n32, 362, 393–95, 397–99
- real wages, 12, 58, 93, 101, 104, 117, 122, 126, 153, 163–64, 191, 333, 335, 359
- reciprocal dumping, 214, 245–48
- reciprocity, 53, 56, 253, 299, 314–16, 328
- Redding, Stephen J., xi, 143–44, 148–49, 156, 167, 181–83, 268
- regional agreements, 187, 199–202, 204, 210–11, 299–300, 317–22, 328–29, 354
- rent seeking, 248, 260
- research and development (R&D): endogenous growth and, 342, 344–47, 354–56; import quotas and, 293; intermediate inputs and, 91; Japan and, 356
- RESET test, 177–79
- returns to scale: constant, 4, 7, 21, 26, 46–47, 55, 172, 190–91, 194, 197, 207–8, 265, 332, 356, 406, 408; endogenous growth and, 332–34, 340, 356–57; gains from trade and, 187, 190–91, 194, 197, 207–10; Heckscher-Ohlin model and, 26, 46–48; imperfect competition and, 207–9; import quotas and, 265; increasing, 4, 48, 80, 119, 148–49, 155, 187, 207, 209–10, 332–33, 357, 377; indexes and, 406, 408; many goods and factors and, 55, 59, 64, 80; monopolistic competition and, 119, 148–49, 155, 172; two-sector models and, 4, 7, 21
- Ricardian model, xi; absolute advantage and, 3; Heckscher-Ohlin model and, 25; import quotas and, 286, 296; intermediate inputs and, 83, 95, 115; many goods and factors and, 52, 70–75, 77, 80–81; monopolistic competition and, 148; relative demand and, 2–3; as two-goods, one-factor model, 1; two-sector models and, 1–4, 21; world relative demand and, 2–3
- Ricardo, David, 331, 399
- Ricardo-Viner model, 21, 51, 57
- Ries, John C., 281–82, 398
- Robinson, Joan, 119
- Rodríguez-Clare, Andrés, 156, 166–68, 171, 173, 181–83
- Rodríguez-López, Antonio, 115
- Rodrik, Dani, 287, 301–3, 328, 353
- Romalis, John, xi, 267, 270–74
- Romer, Paul, 221, 354
- Rose, Andrew K., 247–48
- Rossi-Hansberg, Esteban, xi, 99–104, 107, 114
- Roy's Identity, 301, 303, 421, 425–28
- Rubinstein, Yona, 174
- Ruhl, Kim, 251–52
- rules of origin (ROO), 203
- Rybczynski effects: elasticity and, 62–63, 65–66; endogenous growth and, 338, 354; Heckscher-Ohlin-Vanek (HOV) model and, 35–37; intermediate inputs and, 103–4; many goods and factors and, 56–57, 60, 62, 66; multinationals and, 400
- Rybczynski elasticity, 62
- Rybczynski theorem: endogenous growth and, 338; intermediate inputs and, 102; many goods and factors and, 51, 55–57, 64; monopolistic competition and, 153; two-sector models and, 14–18, 21
- Sachs, Jeffrey D., 90, 353–54
- Salai-i-Martin, Xavier, 353
- Sampson theorem, 352–53, 357
- Samuelson, Paul, 18, 25, 99, 115–16, 286. *See also* Stolper-Samuelson theorem
- Samuelson factor price equalization theorem, 10–11. *See also* factor price equalization (FPE)
- Samuelson-Nikaido lemma, 54–56
- Santo Silva, J. M. C., 175–79
- Sato-Vartia theorem, 126–27, 165
- scale effect: endogenous growth and, 332, 344, 347–52, 357; monopolistic competition and, 123–25, 155
- Scheinkman, José A., 56–57, 59
- sector bias, 87–91
- selection effect, 123–25, 155, 164, 166–67, 348
- self-financing tariffs, 199
- Shatz, Howard J., 90
- sign patterns, 35, 37–38, 63, 375, 390
- sign tests, 37–38, 43, 48, 50, 79
- skill-biased technological change, 86–87, 90, 95, 98, 101, 104
- skilled-labor abundant countries, 49, 92, 367, 374
- small countries: endogenous growth and, 344, 346–47, 356; gains from trade and, 199–200, 207, 209; import quotas and, 256, 261, 284–85; import tariffs and, 219–20, 223; intermediate inputs and, 90, 101, 104, 114; monopolistic competition and, 135, 146
- Smith, Adam, 5, 377
- social welfare, 212–13, 217–20, 228–29, 232, 284, 289, 302, 304, 336
- Solow, Robert M., 339, 356
- Special Economic Zones (SEZs), 388–89, 392
- Spence, A. Michael, 119, 155, 264, 269n15
- Spencer, Barbara, 226, 292–93
- spillovers, 332, 343–44, 346–47, 353–56, 363
- Staiger, Robert W., 299, 311–17, 328
- Standard Industrial Trade Classification (SITC), 35, 147n32, 271, 273, 391
- STATA program, 50, 118, 152, 154, 176–77, 185, 255
- Stigler, George J., 377
- Stiglitz, Joseph, 119–20, 155–56, 244

- Stolper-Samuelson theorem: endogenous growth and, 354; Heckscher-Ohlin model and, 28; intermediate inputs and, 88, 90, 103; many goods and factors and, 51–57, 61–65; monopolistic competition and, 153; multinationals and, 364; two-sector models and, 12–13, 21, 28, 51–57, 61–65, 88, 90, 103, 153, 186, 354, 364; weighted average and, 12
- subsidies: Bertrand duopoly and, 290–92; commercial aircraft and, 293–96; commodity taxes and, 193–96; compensation principle and, 197; Cournot duopoly and, 287–90; discrete choice models and, 424; endogenous growth and, 344; factor prices and, 193–96; gains from trade and, 186, 189t, 190–97, 199, 209–10, 213; imperfect competition and, 287–92; import quotas and, 256–57, 284–97; import tariffs and, 214, 217, 225, 234; many goods and factors and, 70, 285–87; moving, 195; multinationals and, 369, 371, 398; partial reform of tariffs and, 196–98; perfect competition and, 284–85; political economy and, 300–309, 329; poll, 193, 209, 213
- supermodularity, 108–13
- Suzuki, 237
- Sveikauskas, Leo, 35, 37–39
- symmetry: endogenous growth and, 342, 350, 356; gains from trade and, 210; import quotas and, 266; import tariffs and, 236–38, 243, 246, 255; indexes and, 406, 413; many goods and factors and, 61–63, 66; monopolistic competition and, 120–21, 123, 126, 134, 140, 142, 144, 149, 153, 155, 170, 180, 183; multinationals and, 375, 392; political economy and, 308–9, 314–15, 329
- Tang, Heiwai, 386–91
- tariffs: ad volorem, 180, 226–27, 230, 232–33, 237, 240, 243, 256, 263, 269, 281, 311, 316–17; deadweight loss and, 221, 239, 252, 256, 259–60, 284–85, 296, 328, 361, 365; endogenous growth and, 331, 334–35; equivalence and, 256–60; gains from trade and, 186–87, 188t–189t, 196–207, 211–13; GATT and, 214–17 (*see also* General Agreement on Tariffs and Trade (GATT)); Heckscher-Ohlin model and, 46; imperfect competition and, 256, 260–63; import quotas and, 256–63, 269, 278–79, 281–82, 284–85, 292–93, 296; import, 214–53 (*see also* import tariffs); Japanese trucks/motorcycles and, 234–36; many goods and factors and, 64; monopolistic competition and, 123–24, 129, 135–39, 180; most favored nation (MFN) provision and, 187, 223, 253, 299, 317; multinationals and, 361, 364–65, 399–400; Nash equilibrium, 312; non-equivalence and, 260–69; political economy and, 299–318, 323–29; politically optimal, 312; symmetric pass-through and, 236–38; tariff-jumping FDI and, 361; two-country model and, 311–13; two-sector models and, 12; World Trade Organization (WTO) and, 215–17
- Tariff Schedule of the United States (TSUSA), 129
- technological change: factor bias and, 87–91; indexes and, 405–14, 417; intermediate inputs and, 84, 86–88, 90–91, 95, 98–99, 101, 104; many goods and factors and, 63, 80; sector bias and, 87–91; skill-biased, 86–87, 90, 95, 98, 101, 104
- Tenreyro, Silvana, 175–79
- terms of trade: Diewert-Morrison theorem and, 415–17; endogenous growth and, 331, 335, 346; estimating effect of, 236–39; gains from trade and, 187, 198–99, 202, 204–6, 211; import quotas and, 256, 274, 276, 285–87, 296; import tariffs and, 221–39, 249–53; indexes and, 403–5, 407, 409, 411, 413–17; intermediate inputs and, 115–16; many goods and factors and, 73–75; MERCOSUR and, 187, 205; multinationals and, 398; political economy and, 311–16, 328; productivity and, 414–17
- textiles, 20, 81, 258–60
- Tinbergen, Jan, 132–33
- total factor productivity (TFP): endogenous growth and, 333, 339, 354–56; indexes and, 403, 409, 416; intermediate inputs and, 87–90, 95–97; multinationals and, 369
- Trade Act of 1974, 195, 215
- Trade Adjustment Assistance (TAA), 195
- trade in goods: endogenous growth and, 344, 346–47, 356; Heckscher-Ohlin model and, 25–26; indexes and, 418; intermediate inputs and, 107; many goods and factors and, 70, 81; monopolistic competition and, 131, 181; multinationals and, 363, 365–66; two-sector models and, 11, 19
- trade in tasks, 84, 99–103, 107
- trade policy: endogenous lobbies and, 307–10; home market effect and, 316–17; import quotas and, 256, 296–97; import tariffs and, 214, 229, 234; median voter model and, 299–303, 306, 310–12, 315, 318–22, 328; most favored nation (MFN) provision and, 187, 223, 253, 299, 317; multinationals and, 365; nontariff barriers and, 306–7; Ossa theorem and, 316–17; political economy and, 299–303, 328; “protection for sale” model and, 223, 299, 303–7, 310–13, 315, 318, 323, 328; Special Economic Zones (SEZs) and, 388–89, 392; two-country model and, 311–13
- transaction costs, 361, 377–79, 391
- translog expenditure function, 181, 243–44
- transportation costs: endogenous growth and, 350, 370–75; gains from trade and, 210–11; import quotas and, 265–66, 269, 277, 297; import tariffs and, 240, 245–50; intermediate inputs and, 100, 114; many goods and factors and, 70; monopolistic competition and, 135–48, 168, 171, 180, 184
- Trefler, Daniel, 38–45, 48–50, 78–81, 124, 206
- Trindade, Vitor, 47n32, 362, 393–99
- trucks: import quotas and, 278–82, 283t, 298; import tariffs and, 214–15, 234–39, 252, 255
- TSLS estimation, 326–27
- two-good, two-factor model, 1, 4, 11–12, 91, 357

- two-sector models: autarky and, 2–3; capital and, 1, 4, 7–24; changes in endowments and, 13–14; changes in product prices and, 11–12; comparative advantage and, 2–3; determination of factor prices and, 8; equilibrium and, 2–3, 6–8, 10, 13, 16, 18–19, 22; factor endowments and, 1, 8, 10, 14–15, 18–20, 22; factor intensity reversals (FIRs) and, 8–10, 14, 18–21, 23; factor price equalization (FPE) and, 10–11, 16, 18–19, 21; factor price insensitivity and, 8–10; GDP function and, 5–7; Heckscher-Ohlin model and, 1, 4, 10–11; Lerner diagram and, 21–23; returns to scale and, 4, 7, 21, 265; Ricardian model and, 1–4, 21; Rybczynski theorem and, 14–18; Stolper-Samuelson theorem and, 12–13, 21, 28, 51–57, 61–65, 88, 90, 103, 153, 186, 354, 364; tariffs and, 12; trade in goods and, 11, 19; unit-cost functions and, 6–8, 13, 21; wages and, 1–3, 7–8, 10–13, 20–22; zero-profit conditions and, 7–12, 19
- unemployment, 115–16, 123, 195–96, 234, 274
- unit-cost functions: import quotas and, 265; indexes and, 406–7, 409; intermediate inputs and, 91–92, 117; many goods and factors and, 52–53, 55, 64, 76–77; two-sector models and, 6–8, 13, 21
- unit cylinders, 394
- United Automobile Workers, 234, 274
- United Kingdom, 47, 86–87, 187, 204–5, 294
- United States: as capital-abundant country, 20; Chicken War and, 235; China and, 77–78; CUSFTA and, 187, 205–7; discrete choice models and, 424; elasticity and, 80; endogenous growth and, 347, 355–56; factor intensity reversals (FIRs) and, 9–10; free-trade agreement with Canada and, 4, 123–25, 133–36, 187, 195, 203, 205–7, 211; gains from trade and, 187, 192, 195–96, 203, 205–6; GDP function and, 51n1, 63–64, 80; GDP function for, 51n1; gravity equation and, 133–36; Heckscher-Ohlin model and, 28–29, 32–35, 40–42, 47–48; import quotas and, 256–58, 260, 270, 274–82, 293–94; import tariffs and, 215–17, 223, 234–39, 244–45, 249–53; intermediate inputs and, 83–84, 86–87, 89–93, 97n12, 101, 104–8, 113; manufacturing wages over time in, 84–91; many goods and factors and, 51n1, 63–64, 73–74; monopolistic competition and, 119, 123–24, 129–35, 141–43, 147–48, 178–79, 182; multinationals and, 360, 363, 365, 374, 382, 398; NAFTA and, 187, 195, 203, 205–7, 211; nontariff barriers and, 306–7; offshoring and, 105–8 (*see also* offshoring); political economy and, 300, 306–7, 328; rules of origin (ROO) and, 203; skilled labor and, 83–93, 97n12, 101, 104–8, 113; Tariff Schedule of (TSUSA) and, 129; Trade Act of 1974 and, 195, 215; Trade Adjustment Assistance (TAA) and, 195; truck/motorcycle tariffs and, 214–15, 234–39, 244, 252
- U.S. Department of Commerce, 217, 237, 246
- U.S. Export-Import Bank, 292
- U.S. Federal Reserve Bank, 234
- utility: autarky and, 190; discrete choice models and, 419–28; gains from trade and, 190, 193–94, 197–98, 210, 212–13; Heckscher-Ohlin model and, 27, 49; import quotas and, 263–64, 266, 269, 276–79, 282, 297; import tariffs and, 218, 220, 230, 233–34, 239–41, 253; indexes and, 412, 414, 417–18; monopolistic competition and, 120, 124–31, 136–37, 144–45, 149, 152–53, 157, 164, 181; multinationals and, 366, 370, 373, 400; nested CES function and, 241–43; political economy and, 300–303, 319–20, 322, 324–25, 335–36, 340, 343, 348, 358
- value added to gross exports (VAX ratio), 115
- Vanek, Jaroslav, 25, 30
- Van Wincoop, Eric, 140–44, 174–75
- Vartia, Y. O., 126–27, 165
- Venables, Anthony J., 114, 136, 143–44, 148, 372–73
- Ventura, Jaume, 354
- vertical specialization, 83, 95
- Viner, Jacob, 21, 51, 57, 187, 200, 204, 245
- Vogel, Jonathan, 109–14
- voluntary export restraints (VERs): automobiles and, 258, 274–84; correcting for trade policies and, 279, 281; discrete choice models and, 424; import quotas and, 258, 259n4, 262–63, 268–70, 274–84, 296; Japan and, 365, 424; multinationals and, 365; optimal prices and, 277–78; price effects of, 282, 283t; quality and, 268–69, 274–84
- wage fund, 331
- wages: China and, 20; endogenous growth and, 331, 333, 335, 341, 345–46, 348, 350, 359; equilibrium, 13, 58, 71, 121, 144, 363, 394–95; exceeding product prices, 13; gains from trade and, 190–96, 206, 208; Heckscher-Ohlin model and, 28, 41; import quotas and, 266, 270–72, 277, 282, 286–87; induced changes and, 7; inequality and, 90, 108, 115; intermediate inputs and, 83–118; mandated wage equation and, 88, 95–96; many goods and factors and, 53, 58–60, 64, 71–76, 82; minimum, 367; monopolistic competition and, 121–22, 126, 144–45, 148–53, 157–60, 163–64, 166, 170–74, 184–85; multinationals and, 361–63, 366–73, 375, 381–82, 385, 393–98, 401; over time, 84–91; political economy and, 303–4, 316, 324–26, 327t; real, 12, 58, 93, 101, 104, 117, 122, 126, 153, 163–64, 191, 333, 335, 359; two-sector models and, 1–3, 7–8, 10–13, 20–22; United States and, 84–91
- Wan, Henry, Jr., 187, 200–202, 203n6
- Warner, Andrew, 353–54
- Wei, Shang-Jin, 106, 260, 324n19
- weighted averages: endogenous growth and, 332–33, 355–56; factor prices and, 12; gains from trade and, 209; import tariffs and, 237; indexes and, 407, 409, 411, 416; intermediate inputs and,

- 96; many goods and factors and, 55, 59; monopolistic competition and, 131–32, 141
- Weinstein, David E.: Bernstein-Weinstein theorem and, 70–71; dumping and, 247–48; import quotas and, 268; import tariffs and, 221, 223, 244, 247–48; many goods and factors and, 70–71, 77–81; monopolistic competition and, 129–31, 147–48, 152, 167, 181–82; U.S. competitive gains and, 182–83
- welfare: consumer, 303–4, 307, 326, 328, 401, 418; discrete choice models and, 421, 427; endogenous growth and, 331, 336–37, 343, 346–47, 352; gains from trade and, 187, 192, 198–99, 203–4, 206, 210–13; import quotas and, 256–61, 281, 284–93, 296–97; import tariffs and, 214, 217–22, 225–34, 239, 244, 248–52; indexes and, 416–18; many goods and factors and, 72–75, 80; monopolistic competition and, 126, 129–31, 163–67, 182; multinationals and, 361–65, 399–401; political economy and, 302–11, 316–20, 322, 326, 328–31; social, 212–13, 217–20, 228–29, 232, 284, 289, 302, 304, 336
- Whinston, Michael D., 304–5
- white collar labor, 84, 106
- Wong, Kar-yiu, 186
- World Trade Organization (WTO): gains from trade and, 186, 201, 204; input quotas and, 260; input tariffs and, 214–17, 223, 252–53; intermediate inputs and, 107, 116; political economy and, 299–300, 311, 317, 323, 328
- Wright, Greg C., 107, 115–16
- Yamaha, 237
- Yeaple, Stephen R., 149, 155n1, 375–76
- zero-cutoff-profit (ZCP) condition: endogenous growth and, 348, 350–53; monopolistic competition and, 158–62, 170–71, 348, 350–53
- zero-profit conditions: endogenous growth and, 342, 357; factor prices and, 7–12, 19; import quotas and, 271–72; intermediate inputs and, 87–88, 95, 100–101, 105, 116; many goods and factors and, 52–53, 55, 57, 59, 67, 82; monopolistic competition and, 137, 145, 149, 155–56, 158, 161; political economy and, 316; two-sector models and, 7–12, 19
- zeros in trade, 174–77
- Zhu, Susan Chun, 44–45, 78–81