# Poverty Traps [⋆]

**Prepared for the Handbook of Economic Growth**
**(Philippe Aghion and Steven Durlauf, eds.)**

Costas Azariadis [a] and John Stachurski [b]

[a] *Department of Economics, University of California Los Angeles*
*405 Hilgard Avenue, Los Angeles, CA 90095-1477*

[b] *Department of Economics, The University of Melbourne, VIC 3010, Australia*

---

> In the problem of economic development, a phrase that crops up frequently is 'the vicious circle of poverty.' It is generally treated as something obvious, too obvious to be worth examining. I hope I may be forgiven if I begin by taking a look at this obvious concept. (R. Nurkse, 1953)

## 1 Introduction

Despite the considerable amount of research devoted to economic growth and development, economists have not yet discovered how to make poor countries rich. As a result, poverty remains the common experience of billions. One half of the world's people live on less than \$2 per day. One fifth live on less than \$1. [1] If modern production technologies are essentially free for the taking, then why is it that so many people are still poor?

[⋆] This chapter draws on material contained in two earlier surveys by the first author (Azariadis 1996, 2004). Support from the Program of Dynamic Economics at UCLA is acknowledged with thanks, as is research assistance from Athanasios Bolmatis, and discussions with David de la Croix, Cleo Fleming, Oded Galor, Karla Hoff, Kirdan Lees and Yasusada Murata. The second author thanks the Center for Operations Research and Econometrics at Université Catholique de Louvain for their hospitality during a period when part of this survey was written. All simulations and estimations use the open source programming language R.

*Email addresses:* `azariadi@ucla.edu` (Costas Azariadis),
`j.stachurski@econ.unimelb.edu.au` (John Stachurski).

[1] Figures are based on Chen and Ravallion (2001). Using national surveys they calculate a total head-count for the \$1 and \$2 poverty lines of 1.175 and 2.811

The literature that we survey here contains the beginnings of an answer to this question. First, it is true that technology is the primary determinant of a country's income. However, the most productive techniques will not always be adopted: There are self-reinforcing mechanisms, or "traps," that act as barriers to adoption. Traps arise both from market failure and also from "institution failure;" that is, from traps within the set of institutions that govern economic interaction. Institutions—in which we include the state, legal systems, social norms, conventions and so on—are determined endogenously within the system, and may be the direct cause of poverty traps; or they may interact with market failure, leading to the perpetuation of an inefficient status quo.

There is no consensus on the view that we put forward. Some economists argue that the primary suspect for the unfortunate growth record of the least developed countries should be bad domestic policy. Sound governance and free market forces are held to be not only necessary but also *sufficient* to revive the poor economies, and to catalyze their convergence. Because good policy is available to all, there are no poverty traps.

The idea that good policy and the invisible hand are sufficient for growth is at least vacuously true, in the sense that an all-seeing and benevolent social planner who completes the set of markets can succeed where developing country governments have failed. But this is not a theory of development, and of course benevolent social planners are not what the proponents of good governance and liberalization have in mind. Rather, their argument is that development can be achieved by the poor countries if only governments allow the market mechanism to function effectively—to get the prices right—and permit economic agents to fully exploit the available gains from trade. This requires not just openness and non-distortionary public finance, but also the enforcement of property rights and the restraint of predation. [2]

In essence, this is the same story that the competitive neoclassical benchmark economy tells us: Markets are complete, entry and exit is free, transaction costs are negligible, and technology is convex at an efficient scale relative to the size of the market. As a result, the private and social returns to production and investment are equal. A complete set of "virtual prices" ensures

---

billion respectively in 1998. Their units are 1993 purchasing power adjusted US dollars.

[2] Development theory then reduces to Adam Smith's famous and compelling dictum, that "Little else is requisite to carry a state to the highest degree of opulence from the lowest barbarism but peace, easy taxes, and a tolerable administration of justice."

that all projects with positive net social benefit are undertaken. Diminishing returns to the set of reproducible factor inputs implies that when capital is scarce the returns to investment will be high. The dynamic implications of this benchmark were summarized by Solow (1956), Cass (1965), and Koopmans (1965). Even for countries with different endowments, the main conclusion is convergence.

There are good reasons to expect this benchmark will have relevance in practice. The profit motive is a powerful force. Inefficient practices and incorrect beliefs will be punished by lost income. Further, at least one impetus shaping the institutional environment in which the market functions is the desire to mitigate or correct perceived social problems; and one of the most fundamental of all social problems is scarcity. Over time institutions have often adapted so as to relieve scarcity by addressing sources of market inefficiency. [3]

In any case, the intuition gained from studying the neoclassical model has been highly influential in the formulation of development policy. A good example is the structural adjustment programs implemented by the International Monetary Fund. The key components of the Enhanced Structural Adjustment Facility—the centerpiece of the IMF's strategy to aid poor countries and promote long run growth from 1987 to 1999—were prudent macroeconomic policies and the liberalization of markets. Growth, it was hoped, would follow automatically.

Yet the evidence on whether or not non-distortionary policies and diminishing returns to capital will soon carry the poor to opulence is mixed. Even relatively well governed countries have experienced little or no growth. For example, Mali rates as "free" in recent rankings by Freedom House. Although not untroubled by corruption, it scores well in measures of governance relative to real resources (Radlet 2004; Sachs et al. 2004). Yet Mali is still desperately poor. According to a 2001 UNDP report, 70% of the population lives on less than \$1 per day. The infant mortality rate is 230 per 1000 births, and household final consumption expenditure is down 5% from 1980.

Mali is not an isolated case. In fact for all of Africa Sachs et al. (2004) argue that

> With highly visible examples of profoundly poor governance, for example in Zimbabwe, and widespread war and violence, as in Angola, Democratic Republic of Congo, Liberia, Sierra Leone and Sudan, the impression of a continent-wide gov-

[3] See Greif, Milgrom and Weingast (1994) for one of many possible examples.

> ernance crisis is understandable. Yet it is wrong. Many parts of Africa are well governed, and yet remain mired in poverty. Governance is a problem, but Africa's development challenges are much deeper.

There is a further, more subtle, problem with the "no poverty traps" argument. While the sufficiency of good policy and good governance for growth is still being debated, what can be said with certainty is that they are both elusive. The institutions that determine governance and other aspects of market interaction are difficult to reform. Almost everyone agrees that corruption is bad for growth, and yet corruption remains pervasive. Some institutions important to traditional societies have lingered, inhibiting the transition to new techniques of production. The resistance of norms and institutions to change is one reason why the outcome of liberalization and governance focused adjustment lending by the IMF has often been disappointing.

To put the problem more succinctly, the institutional framework in which market interaction takes place is not implemented "from above" (Hoff 2000). Rather it is determined within the system. This includes the formal, legalistic aspects of the framework, but is particularly true for the informal aspects, such as social norms and conventions.

The above considerations lead us back to poverty traps. First, numerous deviations from the neoclassical benchmark generate market failure. Because of these failures, good technologies are not always adopted, and productive investments are not always undertaken. Inefficient equilibria exist. Second, institutions are not always simple choice variables for benevolent national planners. Bounded rationality, imperfect information, and costly transactions make institutions and other "rules of the game" critical to economic performance; and the equilibria for institutions may be inefficient.

Moreover, these inefficient equilibria have a bad habit of *reinforcing* themselves. Corrupt institutions can generate incentives which reward more corruption. Workers with imperfectly observed skills in an unskilled population may be treated as low skilled by firms, and hence have little incentive to invest large sums in education. Low demand discourages investment in increasing returns technology, which reduces productivity and reinforces low demand. That these inefficient outcomes are self-reinforcing is important—were they not then presumably agents would soon make their way to a better equilibrium.

Potential departures from the competitive neoclassical benchmark which cause market failure are easy to imagine. One is increasing returns to scale, both internal and external. Increasing returns matter because development is almost synonymous with industrialization, and with the adoption of modern production techniques in agriculture, manufacturing and services. These modern techniques involve both fixed costs—internal economies—and greater specialization of the production process, the latter to facilitate application of machines.

The presence of fixed costs for a *given* technology is more troubling for the neoclassical benchmark in poor countries because there market scale is relatively small. If markets are small, then the neoclassical assumption that technologies are convex at an efficient scale may be violated. The same point is true for market scale and specialization, in the sense that for poor countries a given increase in market scale may lead to considerably more opportunity to employ indirect production. [4]

Another source of increasing returns follows from the fact that modern production techniques are knowledge-intensive. As Romer (1990) has emphasized, the creation of knowledge is associated with increasing returns for several reasons. First, knowledge is non-rival and only partially excludable. Romer's key insight is that in the presence of productive non-rival inputs, *the entire replication-based logical argument for constant returns to scale immediately breaks down.* Thus, knowledge creation leads to positive technical externalities and increasing returns. Second, new knowledge tends in the aggregate to complement existing knowledge.

If scale economies, positive spillovers and other forms of increasing returns are important, then long run outcomes may not coincide with the predictions of the neoclassical benchmark. The essence of the problem is that when returns are increasing a rise in output *lowers* unit cost, either for the firm itself or for other firms in the industry. This sets in motion a chain of positive self-reinforcement. Lower unit cost encourages production, which further lowers unit cost, and so on. Such positive feedbacks can strongly reinforce either

[4] Domestic markets *are* small in many developing countries, despite the possibility of international trade. In tropical countries, for example, roads are difficult to build and expensive to maintain. In Sub-Saharan Africa, overland trade with European and other markets is cut off by the Sahara. At the same time, most Sub-Saharan Africans live in the continent's interior highlands, rather than near the coast. To compound matters, very few rivers from the interior of this part of the continent are ocean-navigable, in contrast to the geography of North America, say, or Europe (Limao and Venables 2001; Sachs et al. 2004). The potential for international trade to mitigate small market size is thus far lower than for a country with easy ocean access, such as Singapore or the UK.

poverty *or* development.

Another deviation from the competitive neoclassical benchmark that we discuss at length is failure in credit and insurance markets. Markets for loans and insurance suffer more acutely than most from imperfections associated with a lack of complete and symmetric information, and with all the problems inherent in anonymous trading over time. Borrowers may default or try not to pay back loans. The insured may become lax in protecting their own possessions.

One result of these difficulties is that lenders usually require collateral from their borrowers. Collateral is one thing that the poor always lack. As a result, the poor are credit constrained. This can lead to an inefficient outcome which is self-reinforcing: Collateral is needed to borrow funds. Funds are needed to take advantage of economic opportunities—particularly those involving fixed costs. The ability to take advantage of opportunities determines income; and through income is determined the individual's wealth, and hence their ability to provide collateral. Thus the poor lack access to credit markets, which is in turn the cause of their own poverty.

An important aspect of this story for us is that many modern sector occupations and production techniques have indivisibilities which are not present in subsistence farming, handicraft production or other traditional sector activities. Examples include projects requiring fixed costs, or those needing large investments in human capital such as education and training. The common thread is that through credit constraints the uptake of new technologies is inhibited.

With regards to insurance, it has been noted that—combined with limited access to credit—a lack of insurance is more problematic for the poor than the rich, because the poor cannot self-insure by using their own wealth. As a result, a poor person wishing to have a smooth consumption path may be forced to choose activities with low variance in returns, possibly at the cost of lower mean. Over time, lower mean income leads to more poverty.

Credit and insurance markets are not the only area of the economy where limited information matters. Nor is lack of information the only constraint on economic interaction: The world we seek to explain is populated with economic actors who are boundedly rational, not rational. The fact that people are neither all-knowing nor have unlimited mental capability is important to us for several other reasons.

One is that transactions become costly; and this problem is exacerbated as societies become larger and transactions more impersonal. Interaction with large societies requires more information about more people, which in turn requires more calculation and processing (North 1993, 1995). Second, if we concede that agents are boundedly rational then we must distinguish between the objective world and each agent's subjective interpretation of the world. These interpretations are formed on the basis of individual and local experience, of individual inference and deduction, and of the intergenerational transmission of knowledge, values and customs. The product of these inputs is a mental model or belief system which drives, shapes and governs individual action (Simon 1986; North 1993).

These two implications of bounded rationality are important. The first (costly transactions) because when transactions are costly institutions matter. The second (local mental models and subjective beliefs) because these features of different countries and economies shape their institutions.

In this survey we emphasize two related aspects of institutions and their connection to poverty traps. The first is that institutions determine how well inefficiencies arising within the market are resolved. A typical example would be the efforts of economic and political institutions to solve coordination failure in a given activity resulting from some form of complementary externalities. The second is that institutions themselves can have inefficient equilibria. Moreover, institutions are *path dependent*. In the words of Paul A. David, they are the "carriers of history" (David 1994).

Why are institutions characterized by multiple equilibria and path dependence? Although human history often shows a pattern of negotiation towards efficient institutions which mitigate the cost of transactions and overcome market failure, it is also true that institutions are created and perpetuated by those with political power. As North (1993, p. 3) has emphasized, "institutions are not necessarily or even usually created to be socially efficient; rather they, or at least the formal rules, are created to serve the interests of those with the bargaining power to create new rules."

Moreover, the institutional framework is path dependent because those who currently hold power almost always have a stake in its perpetuation. Consider for example the current situation in Burundi, which has been mired in civil war since its first democratically elected president was assassinated in 1993. The economic consequences have not been efficient. Market-based economic activity has collapsed along with income. Life expectancy has fallen from 54 years in 1992 to 41 in 2000. Household final consumption expenditure is down 35%

from 1980. Nevertheless, the military elite have much to gain from continuation of the war. The law of the gun benefits those with most guns. Curfew and identity checks provide opportunities for extortion. Military leaders continue to subvert a peace process that would lead to reform of the army.

Path dependence is strengthened by positive feedback mechanisms which *reinforce* existing institutions. For example, the importance of strong property rights for growth has been extensively documented. Yet Acemoglu, Johnson and Robinson (this volume) document how in Europe during the Middle Ages monarchs consistently failed to ensure property rights for the general population. Instead they used arbitrary expropriation to increase their wealth and the wealth of their allies. Increased wealth closed the circle of causation by reinforcing their own power. Engerman and Sokoloff (2004) discuss how initial inequality in some of Europe's colonial possessions led to policies which hindered broad participation in market opportunities and strengthened the position of a small elite. Such policies tended to reinforce existing inequality (while acting as a break on economic growth).

Path dependence is also inherent in the way that informal norms form the foundations of community adherence to legal stipulations. While the legal framework can be changed almost instantaneously, social norms, conventions and other informal institutions are invariably persistent (otherwise they could hardly be conventions). Often legislation is just the first step a ruling body must take when seeking to alter the de facto rules of the game. [5]

Finally, bounded rationality can be a source of self-reinforcing inefficient outcomes independent of institutions. For example, even in an otherwise perfect market a lack of global knowledge can cause agents to choose an inefficient technology, which is then reinforced by herd effects. [6] When there are market frictions or nonconvexities such outcomes may be exacerbated. For example, if technology is nonconvex then initial poor choices by boundedly rational agents can be locked in (Arthur 1994).

In summary, the set of all self-reinforcing mechanisms which can potentially cause poverty is large. Even worse, the different mechanisms can interact, and reinforce one another. Increasing returns may cause investment complementar-

[5] For example, Transparency International's 2004 Global Corruption Report notes that in Zambia courts have been reluctant to hand down custodial sentences to those convicted of corruption, "principally because it was felt that white-collar criminals *did not deserve* to go to jail." (Emphasis added.)

[6] This example is due to Karla Hoff.

ities and hence coordination failure, which is then perpetuated by pessimistic beliefs and conservative institutions. Rent-seeking and corruption may discourage investment in new technology, which lowers expected wages for skilled workers, decreasing education effort and hence the pool of skilled workers needed by firms investing in technology. The disaffected workers may turn to rent-seeking. Positive feedbacks reinforce other feedbacks. In these kinds of environments the relevance of the neoclassical benchmark seems tenuous at best.

Our survey of poverty traps proceeds as follows. Section 2 reviews key development facts. Section 3 considers several basic models associated with persistent poverty, and their implications for dynamics and the data. Section 4 looks at the empirics of poverty traps. Our survey of microfoundations is in Sections 5–8. Section 9 concludes.

There are already a number of surveys on poverty traps, including two by the first author (Azariadis 1996, 2004). The surveys by Hoff (2000) and Matsuyama (1995, 1997) are excellent, as is Easterly (2001). See in addition the edited volumes by Bowles, Durlauf and Hoff (2004) and Mookherjee and Ray (2001). Parente and Prescott (this volume) also focus on barriers to technology adoption as an explanation of cross-country variation in income levels. In their analysis institutions are treated as exogenous.

## 2 Development Facts

In Section 2.1 we briefly review key development facts, focusing on the vast and rising differences in per capita income across nations. Section 2.2 reminds the reader how these disparities came about by quickly surveying the economic history behind income divergence.

### *2.1 Poverty and riches*

What does it mean to live on one or two dollars per day? Poverty translates into hunger, lack of shelter, illness without medical attention. Calorie intake in the poorest countries is far lower than in the rich. The malnourished are less productive and more susceptible to disease than those who are well fed. Infant mortality rates in the poorest countries are up to 40 or 50 times higher than the OECD average. Many of the common causes, such as pneumonia or dehydration from diarrhea, cost very little to treat.

The poor are more vulnerable to events they cannot control. They are less able to diversify their income sources. They are more likely to suffer from famine, violence and natural disasters. They have lower access to credit markets and insurance, with which to smooth out their consumption. Their children risk exploitation, and are less likely to become educated.

The plight of the poor is even more striking when compared to the remarkable wealth of the rich. Measured in 1996 US dollars and adjusted for purchasing power parity, average yearly income per capita in Luxembourg for 2000 was over $46,000. [7] In Tanzania, by contrast, average income for 2000 was about $500. In other words, people in Luxembourg are nearly 100 times richer on average than those living in the very poorest countries. [8] Luxembourg is rather exceptional in terms of per capita income, but even in the US average income is now about 70 times higher than it is in Tanzania.

How has the gap between the richest and the poorest evolved over time? The answer is simple: It has increased dramatically, even in the postwar era. In 1960, per capita income in Tanzania was $478. After rising somewhat during the 1960s and 1970s it collapsed again in the 1980s. By 2000 it was $457. Many other poor countries have had similar experiences, with income hovering around the $500–1,000 mark. Meanwhile, the rich countries continued exponential growth. Income in the US grew from $12,598 in 1960 (26 times that of Tanzania) to $33,523 in 2000 (73 times). Other rich industrialized countries had similar experiences. In Australia over the same period per capita GDP rose from $10,594 to $25,641. In France it rose from $7,998 to $22,253, and in Canada from $10,168 to $26,983.

Figure 1 shows how the rich have gotten richer relative to the poor. The left hand panel compares an average of real GDP per capita for the 5 richest countries in the Penn World Tables with an average of the same for the 5 poorest. The comparison is at each point in time from 1960 to the year 2000.

[7] Unless otherwise stated, all income data in the remainder of this section is from the Penn World Tables Version 6.1 (Heston, Summers and Aten 2002). Units are PPP and terms of trade adjusted 1996 US dollars.

[8] Some countries record per capita income even lower than the figure given above for Tanzania. 1997 average income in Zaire is measured at $276. Sachs et al. (2004) use the World Bank's 2003 World Development Indicators to calculate a population-weighted average income for Sub-Saharan Africa at 267 PPP-adjusted US dollars, or 73 cents a day.

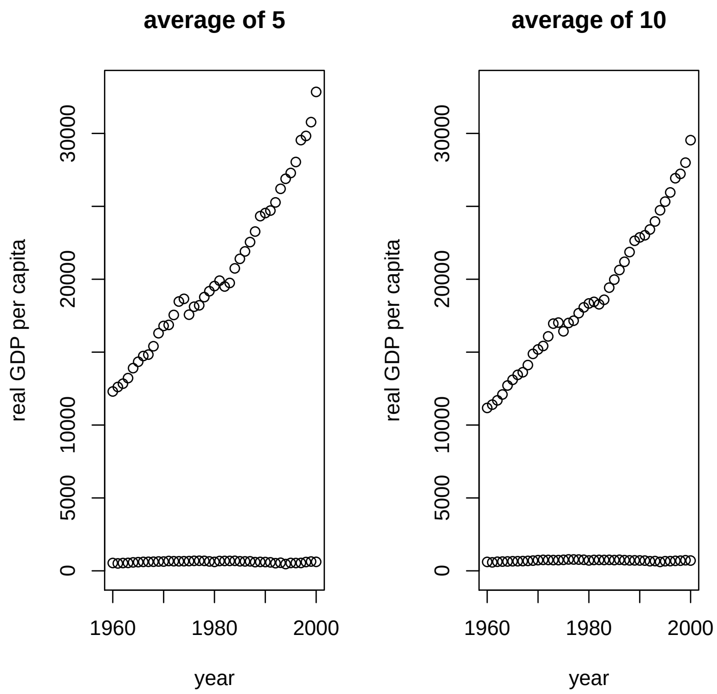


Fig. 1.

The right panel does the same comparison with groups of 10 countries (10 richest vs 10 poorest) instead of 5. Both panels show that by these measures income disparity has widened dramatically in the postwar era, and the rate of divergence is, if anything, increasing. The vast and growing disparity in output per person shown in Figure 1 is what growth and development theorists are obliged to explain. [9]

### *2.2 A brief history of economic development*

How did the massive disparities in income shown in Figure 1 arise? It is worth reviewing the broad history of economic development in order to remind ourselves of key facts. [10]

Although the beginnings of agriculture some ten thousand years ago marked the start of rapid human progress, for most of the subsequent millennia all but

[9] Of course the figure says nothing about mobility. The poor this year could be the rich next year. See Section 4.1 for some discussion of mobility.

[10] The literature on origins of modern growth is too extensive to list here. See for example the monographs of Rostow (1975) and Mokyr (2002).

a tiny fraction of humanity was poor as we now define it, suffering regularly from hunger and highly vulnerable to adverse shocks. Early improvements in economic welfare came with the rise of premodern city-states. Collective organization of irrigation, trade, communications and security proved more conducive to production than did autarky. Handicraft manufacture became more specialized over time, and agriculture more commercial. (Already the role of increasing returns and the importance of institutions are visible here.)

While such city-states and eventually large empires rose and fell over time, and the wealth of their citizens with them, until the last few hundred years no state successfully managed the transition to what we now call modern, self-sustaining growth. Increased wealth was followed by a rise in population. Malthusian pressure led to famine and disease.

The overriding reason for lack of sustained growth was that in the premodern world production technology improved only slowly. While the scientific achievements of the ancient Mediterranean civilizations and China were remarkable, in general there was little attempt to apply science to the economic problems of the peasants. Scientists and practical people had only limited interaction. Men and women of ability usual found that service to the state—or predation against other states—was more rewarding than entrepreneurship and invention.

Early signs of modern growth appeared in Western Europe around the middle of the last millennium. Science from the ancient world had been preserved, and now began to be extended. The revolutionary ideas of Copernicus led to intensive study of the natural world and its regularities. The printing press and movable type dramatically changed the way ideas were communicated. Innovations in navigation opened trade routes and new lands. Gunpowder and the cannon swept away local fiefdoms based on feudal castles.

These technological innovations led to changes in institutions. The weakening of local fiefdoms was followed in many countries by a consolidation of central authority, which increased the scale of markets and the scope for specialization.[11] Growing trade with the East and across the Atlantic produced a rich and powerful merchant class, who subsequently leveraged their political muscle to gain strengthened property and commercial rights.

---

[11] For example, in 1664 Louis XIV of France drastically reduced local tolls and unified import customs. In 1707 England incorporated Scotland into its national market. Russia abolished internal duties in 1753, and the German states instituted similar reforms in 1808.

Increases in market size, institutional reforms and progress in technology at first lead to steady but unspectacular growth in incomes. In 1820 the richest countries in Europe had average per capita incomes of around \$1,000 to \$1,500—some two or three times that of the poorest countries today. However, in the early 19th Century the vast majority of people were still poor.

In this survey we compare productivity in the poor countries with the economic triumphs of the rich. Richness in our sense begins with the Industrial Revolution in Britain (although the rise in incomes was not immediate) and, subsequently, the rest of Western Europe. Industrialization—the systematic application of modern science to industrial technology and the rise of the factory system—led to productivity gains *entirely different in scale from those in the premodern world.*

In terms of proximate causes, the Industrial Revolution in Britain was driven by a remarkable revolution in science that occurred during the period from Copernicus through to Newton, and by what Mokyr (2002) has called the "Industrial Enlightenment," in which traditional artisanal practices were systematically surveyed, cataloged, analyzed and generalized by application of modern science. Critical to this process was the interactions of scientists with each other and with the inventors and practical men who sought to profit from innovation.

Science and invention led to breakthroughs in almost all areas of production; particularly transportation, communication and manufacturing. The structure of the British economy was massively transformed in a way that had never occurred before. Employment in agriculture fell from nearly 40% in 1820 to about 12% in 1913 (and to 2.2% in 1992). The stock of machinery, equipment and non-residential structures per worker increased by a factor of five between 1820 and 1890, and then doubled again by 1913. The literacy rate also climbed rapidly. Average years of education increased from 2 in 1820 to 4.4 in 1870 and 8.8 in 1913 (Maddison 1995).

As a result of these changes, per capita income in the UK jumped from about \$1,700 in 1820 to \$3,300 in 1870 and \$5,000 in 1913. Other Western European countries followed suit. In the Netherlands, income per capita grew from \$1,600 in 1820 to \$4,000 in 1913, while for Germany the corresponding figures are \$1,100 and \$3,900. [12]

Looking forward from the start of the last century, it might have seemed likely

[12] The figures are from Maddison (1995). His units are 1990 international dollars.

that these riches would soon spread around the world. The innovations and inventions behind Britain's productivity miracle were to a large extent public knowledge. Clearly they were profitable. Adaptation to new environments is not costless, but nevertheless one suspects it was easy to feel that already the hard part had been done.

Such a forecast would have been far too optimistic. Relatively few countries besides Western Europe and its off-shoots have made the transition to modern growth. Much of the world remains mired in poverty. Among the worst performers are Sub-Saharan Africa and South Asia, which together account for some 70% of the 1.2 billion people living on less than $1 per day. But poverty rates are also high in East Asia, Latin America and the Carribean. Why is it that so many countries are still poorer than 19th Century Britain? Surely the different outcomes in Britain and a country such as Mali can—at least from a modeler's perspective—be Pareto ranked. What deviation from the neoclassical benchmark is it that causes technology growth in these countries to be retarded, and poverty to persist?

## 3 Models and Definitions

We begin our attempt to answer the question posed at the end of the last section with a review of the convex neoclassical growth model. It is appropriate to start with this model because it is the benchmark from which various deviations will be considered. Section 3.2 explains why the neoclassical model *cannot* explain the vast differences in income per capita between the rich and poor countries. Section 3.3 introduces the first of two "canonical" poverty trap models. These models allow us to address issues common to all such models, including dynamics and implications for the data. Section 3.4 introduces the second.

### *3.1 Neoclassical growth with diminishing returns*

The convex neoclassical model (Solow 1956) begins with an aggregate production function of the form

$$Y_t = K_t^\alpha (A_t L_t)^{1-\alpha} \xi_{t+1}, \quad \alpha \in (0,1), \tag{1}$$

where $Y$ is output of a single composite good, $A$ is a productivity parameter, $K$ is the aggregate stock of tangible and intangible capital, $L$ is a measure of

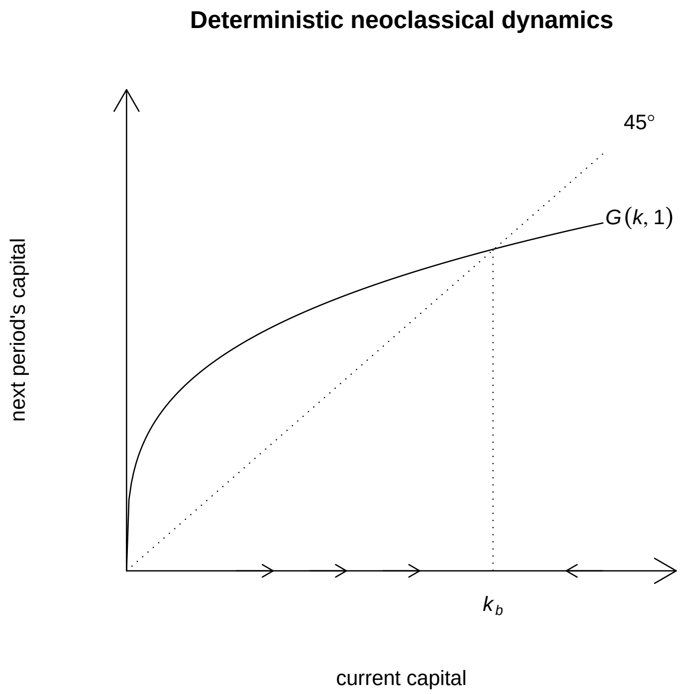


Fig. 2.

labor input, and $\xi$ is a shock. In this formulation the sequence $(A_t)_{t \geq 0}$ captures the persistent component of productivity, and $(\xi_t)_{t \geq 0}$ is a serially uncorrelated innovation.

The production function on the right hand side of (1) represents maximum output for a given set of inputs. That output is maximal follows from competitive markets, profit seeking and free entry. (Implicit is the assumption of no significant indivisibilities or nonconvexities.) The Cobb-Douglas formulation is suggested by relative constancy of factor shares with respect to the level of worker output.

Savings of tangible and intangible capital from current output occurs at constant rate $s$; in which case $K$ evolves according to the rule

$$K_{t+1} = sY_t + (1 - \delta)K_t. \tag{2}$$

Here $\delta \in (0, 1]$ is a constant depreciation rate. The savings rate can be made endogenous by specifying intertemporal preferences. However the discussion in this section is purely qualitative; endogenizing savings changes little.[13]

[13] See, for example, Brock and Mirman (1972) or Nishimura and Stachurski (2004) for discussion of dynamics when savings is chosen optimally.

If, for example, labor $L$ is undifferentiated and grows at exogenous rate $n$, and if productivity $A$ is also exogenous and grows at rate $\gamma$, then the law of motion for capital per effective worker $k_t := K_t/(A_tL_t)$ is given by

$$k_{t+1} = \frac{sk_t^\alpha \xi_{t+1} + (1-\delta)k_t}{\theta} =: G(k_t, \xi_{t+1}), \tag{3}$$

where $\theta := 1 + n + \gamma$. The evolution of output per effective worker $Y_t/(A_tL_t)$ and output per capita $Y_t/L_t$ are easily recovered from (1) and (3).

Because of diminishing returns, capital poor countries will extract greater marginal returns from each unit of capital stock invested than will countries with plenty of capital. The result is convergence to a long-run outcome which depends only on fundamental primitives (as opposed to beliefs, say, or historical conditions).

Figure 2 shows the usual deterministic global convergence result for this model when the shock $\xi$ is suppressed. The steady state level of capital per effective worker is $k_b$. Figure 3 illustrates stochastic convergence with three simulated series from the law of motion (3), one with low initial income, one with medium initial income and one with high initial income. Part (a) of the figure gives the logarithm of output per effective worker, while (b) is the logarithm of output per worker. All three economies converge to the balanced growth path.[14]

Average convergence of the sample paths for $(k_t)_{t\geq 0}$ and income is mirrored by convergence in probabilistic laws. Consider for example the sequence of marginal distributions $(\psi_t)_{t\geq 0}$ corresponding to the sequence of random variables $(k_t)_{t\geq 0}$. Suppose for simplicity that the sequence of shocks is independent, identically distributed and lognormal; and that $k_0 > 0$. It can then be shown that (a) the distribution $\psi_t$ is a density for all $t \geq 1$, and (b) the sequence $(\psi_t)_{t\geq 0}$ obeys the recursion

$$\psi_{t+1}(k') = \int_0^\infty \Gamma(k, k')\psi_t(k)dk, \quad \text{for all } t \geq 1, \tag{4}$$

where the *stochastic kernel* $\Gamma$ in (4) has the interpretation that $\Gamma(k, \cdot)$ is the probability density for $k_{t+1} = G(k_t, \xi_{t+1})$ when $k_t$ is taken as given and equal to $k$.[15] The interpretation of (4) is straightforward. It says (heuristically) that

[14] In the simulation the sequence of shocks $(\xi_t)_{t\geq 0}$ is lognormal, independent and identically distributed. The parameters are $\alpha = 0.3$, $A_0 = 100$, $\gamma = .025$, $n = 0$, $s = 0.2$, $\delta = 0.1$, and $\ln \xi \sim N(0, 0.1)$. Here and in all of what follows $X \sim N(\mu, \sigma)$ means that $X$ is normally distributed with mean $\mu$ and standard deviation $\sigma$.

[15] See the technical appendix for details.

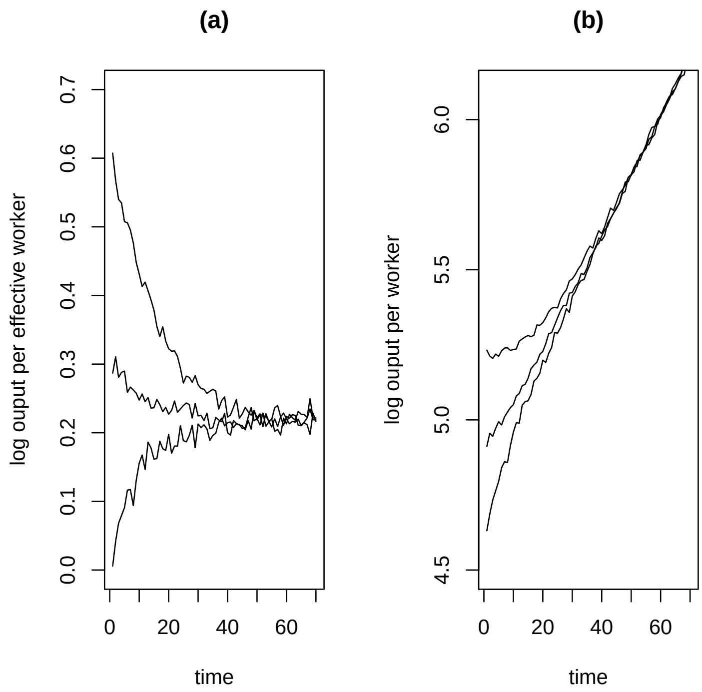


Fig. 3.

$\psi_{t+1}(k')$, the probability that the state takes value $k'$ next period, is equal to the probability of taking value $k'$ next period given that the current state is $k$ (that is, $\Gamma(k, k')$), summed across all $k$, and weighted by the probability that the current state actually takes the value $k$ (i.e., $\psi_t(k)dk$).

Here the conditional distribution $\Gamma(k, \cdot)$ of $k_{t+1}$ given $k_t = k$ is easily calculated from (3) and the familiar change-of-variable rule that if $\xi$ is a random variable with density $\varphi$ and $Y = h(\xi)$, where $h$ is smooth and strictly monotone, then $Y$ has density $\varphi(h^{-1}(y)) \cdot [dh^{-1}(y)/dy]$. Applying this rule to (3) we get

$$\Gamma(k, k') := \varphi\left[\frac{\theta k' - (1-\delta)k}{sk^{\alpha}}\right]\frac{\theta}{sk^{\alpha}}, \tag{5}$$

where $\varphi$ is the lognormal density of the productivity shock $\xi$.[16]

All Markov processes have the property that the sequences of marginal distributions they generate satisfies a recursion in the form of (4) for some stochastic kernel $\Gamma$.[17] Although the state variables usually do not themselves become

[16] Precisely, $z \mapsto \varphi(z)$ is this density when $z > 0$ and is equal to zero when $z \leq 0$.
[17] See the technical appendix for definitions. Note that we are working here with processes that generate sequences of *densities*. If the marginal distributions are not densities, and the conditional distribution contained in $\Gamma$ is not a density, then the

stationary (due to the ongoing presence of noise), the sequence of probabilities $(\psi_t)_{t\geq 0}$ may. In particular, the following behavior is sometimes observed:

**Definition 3.1 (Ergodicity)** *Let a growth model be defined by some stochastic kernel* $\Gamma$*, and let* $(\psi_t)_{t\geq 0}$ *be the corresponding sequence of marginal distributions generated by* (4)*. The model is called* ergodic *if there is a unique probability distribution* $\psi^*$ *supported on* $(0, \infty)$ *with the property that* (i)

$$\psi^*(k') = \int_0^\infty \Gamma(k, k')\psi^*(k)dk \quad \textit{for all } k';$$

*and* (ii) *the sequence* $(\psi_t)_{t\geq 0}$ *of marginal distributions for the state variable satisfies* $\psi_t \to \psi^*$ *as* $t \to \infty$ *for all non-zero initial states.*[18]

It is easy to see that (i) and (4) together imply that if $\psi_t = \psi^*$ (that is, $k_t \sim \psi^*$), then $\psi_{t+1} = \psi^*$ (that is, $k_{t+1} \sim \psi^*$) also holds (and if this is the case then $k_{t+2} \sim \psi^*$ follows, and so on). A distribution with this property is called a stationary distribution, or ergodic distribution, for the Markov chain. Property (ii) says that, conditional on a strictly positive initial stock of capital, the marginal distribution of the stock converges in the long run to the ergodic distribution.

Under the current assumptions it is relatively straightforward to prove that the Solow process (3) is ergodic. (See the technical appendix for more details.) Figures 4 and 5 show convergence in the neoclassical model (3) to the ergodic distribution $\psi^*$. In each of the two figures an initial distribution $\psi_0$ has been chosen arbitrarily. Since the process is ergodic, in both figures the sequence of marginal distributions $(\psi_t)_{t\geq 0}$ converges to the same ergodic distribution $\psi^*$. This distribution $\psi^*$ is determined purely by fundamentals, such as the propensity to save, the rate of capital depreciation and fertility.[19]

---

formula (4) needs to be modified accordingly. See the technical appendix. Other references include Stokey, Lucas and Prescott (1989), Futia (1982) and Stachurski (2004).

[18] Convergence refers here to that of measures in the total variation norm, which in this case is just the $L_1$ norm. Convergence in the norm topology implies convergence in distribution in the usual sense.

[19] The algorithms and code for computing marginal and ergodic distributions are available from the authors. All ergodic distributions are calculated using Glynn and Henderson's (2001) look-ahead estimator. Marginals are calculated using a variation of this estimator constructed by the authors. The parameters in (3) are chosen—rather arbitrarily—as $\alpha = 0.3$, $\gamma = .02$, $n = 0$, $s = 0.2$, $\delta = 1$, and $\ln \xi \sim N(3.6, 0.11)$.

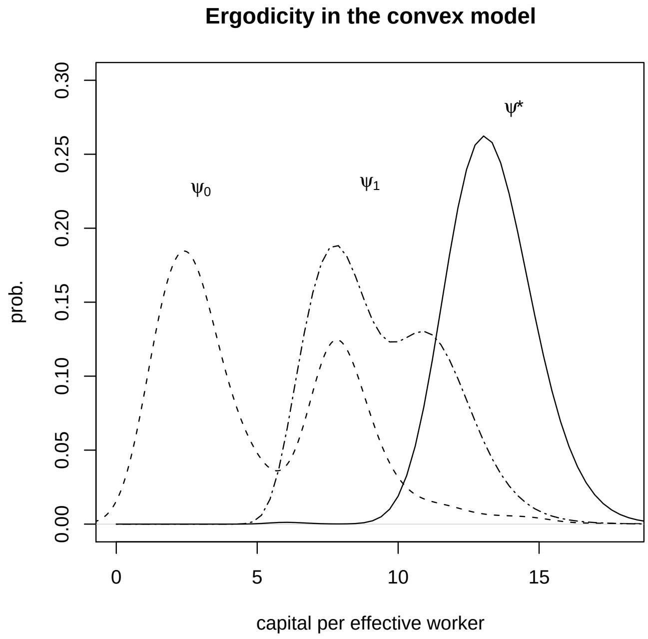


Fig. 4.

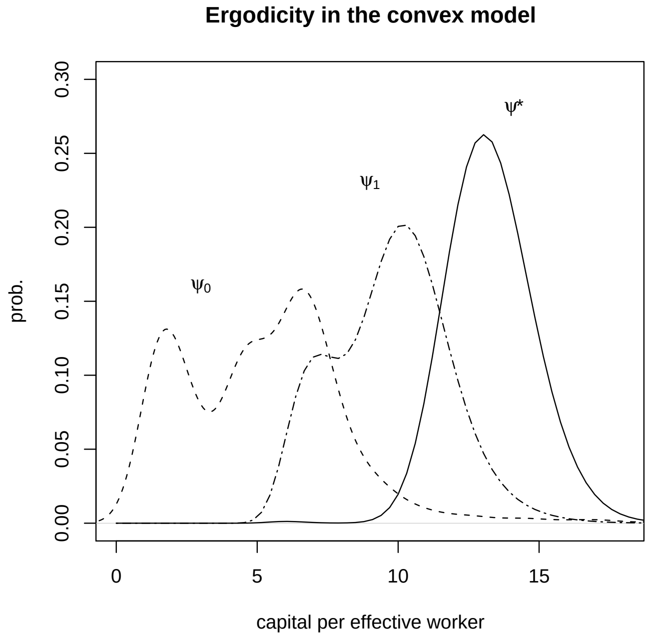


Fig. 5.

Notice in Figures 4 and 5 how initial differences are moderated under the convex neoclassical transition rule. We will see that, without convexity, initial differences often persist, and may well be amplified as the system evolves through time.

The convex neoclassical growth model described in the previous section predicts that per capita incomes will differ across countries with different rates of physical and human capital formation or fertility. Can the model provide a reasonable explanation then for the fact that per capita income in the US is more than 70 times that in Tanzania or Malawi?

The short answer to this question is no. First, rates at which people accumulate reproducible factors of production or have children (fertility rates) are endogenous—in fact they are choice variables. To the extent that factor accumulation and fertility are important, we need to know *why* some individuals and societies make choices that lead them into poverty. For poverty is suffering, and, all things being equal, few people will choose it.

This same observation leads us to suspect that the choices facing individuals in rich countries and those facing individuals in poor countries are very different. In poor countries, the choices that collectively would drive modern growth—innovation, investment in human and physical capital, etc.—must be perceived by individuals as *worse* than those which collectively lead to the status quo.[20]

A second problem for the convex neoclassical growth model as an explanation of level differences is that even when we regard accumulation and fertility rates as exogenous, they must still account for all variation in income per capita across countries. However, as many economists have pointed out, the differences in savings and fertility rates are not large enough to explain real income per capita ratios in the neighborhood of 70 or 100. A model ascribing output variation to these few attributes alone is insufficient. A cotton farmer in the US does not produce more cotton than a cotton farmer in Mali simply because he has saved more cotton seed. The production techniques used in these two countries are utterly different, from land clearing to furrowing to planting to irrigation and to harvest. A model which does not address the vast differences in production technology across countries cannot explain the observed differences in output.

Let us very briefly review the quantitative version of this argument.[21] To

[20] For this reason, endogenizing savings by specifying preferences is not very helpful, because to get poverty in optimal growth models we must assume that the poor are poor because they prefer poverty.

[21] The review is brief because there are many good sources. See, for example, Lucas (1990), King and Rebelo (1993), Prescott (1998), Hall and Jones (1999) or Easterly

begin, recall the aggregate production function (1), which is repeated here for convenience:

$$Y_t = K_t^{\alpha}(A_t L_t)^{1-\alpha}\xi_{t+1}. \tag{6}$$

All of the components are more or less observable besides $A_t$ and the shock.[22] Hall and Jones (1999) conducted a simple growth accounting study by collecting data on the observable components for the year 1988. They calculate that the geometric average of output per worker for the 5 richest countries in their sample was 31.7 times that of the 5 poorest countries. Taking $L$ to be a measure of human capital, variation in the two inputs $L$ and $K$ contributed only factors of 2.2 and 1.8 respectively. This leaves all the remaining variation in the productivity term $A$.[23]

This is not a promising start for the neoclassical model as a theory of level differences. Essentially, it says that there is no single map from total inputs to aggregate output that holds for every country. Why might this be the case? We know that the aggregate production function is based on a great deal of theory. Output is maximal for a given set of inputs because of perfect competition among firms. Free entry, convex technology relative to market size, price taking and profit maximization mean that the best technologies are used—and used efficiently. Clearly some aspect of this theory must deviate significantly from reality.

Now consider how this translates into predictions about level differences in income per capita. When the shock is suppressed ($\xi_t = 1$ for all $t$), output per capita converges to the balanced path

$$y_t := \frac{Y_t}{L_t} = A_t(s/\kappa)^{\alpha/(1-\alpha)}, \tag{7}$$

where $\kappa := n + \gamma + \delta$.[24] Suppose at first that the path for the productivity residual is the same in all countries. That is, $A_t^i = A_t^j$ for all $i, j$ and $t$. In this

and Levine (2000).

[22] The parameter $\alpha$ is the share of capital in the national accounts. Human capital can be estimated by collecting data on total labor input, schooling, and returns to each year of schooling as a measure of its productivity.

[23] The domestic production shocks $(\xi_t)_{t\geq 0}$ are not the source of the variation. This is because they are very small relative to the differences in incomes across countries, and, by definition, not persistent. (Recall that in our model they are innovations to the permanent component $(A_t)_{t\geq 0}$.)

[24] When considering income *levels* it is necessary to assume that countries are in the neighborhood of the balanced path, for this is where the model predicts they will be. Permitting them to be "somewhere else" is not a theory of income level variation.

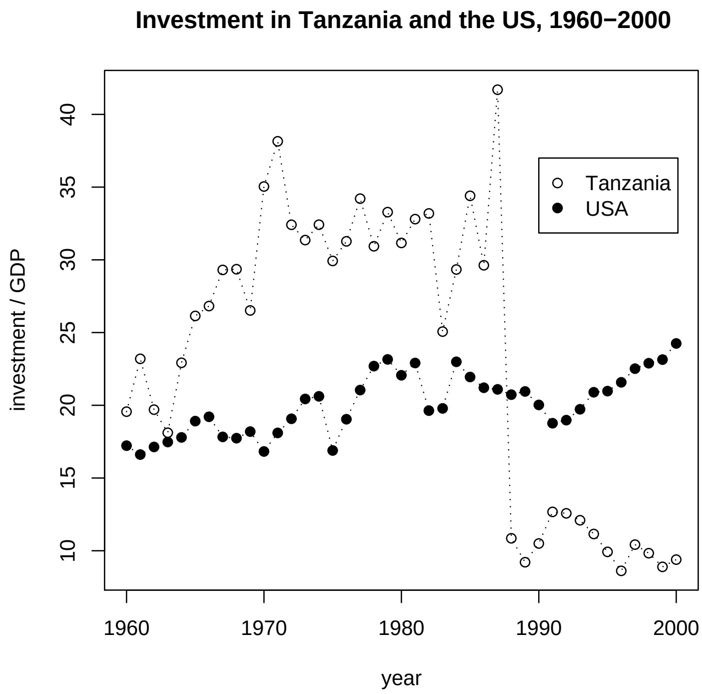


Fig. 6.

case, the ratio of output per capita in country $i$ relative to that in country $j$ is constant and equal to

$$\frac{y^i}{y^j} = \left(\frac{s^i \kappa^j}{s^j \kappa^i}\right)^{\alpha/(1-\alpha)} . \tag{8}$$

The problem for the neoclassical model is that the term inside the brackets is usually not very large. For example, if we compare the US and Tanzania, say, and if we identify capital with physical capital, then average investment as a fraction of GDP between 1960 and 2000 was about 0.2 in the US and 0.24 in Tanzania. (Although the rate in Tanzania varied a great deal around this average. See Figure 6.) The average population growth rates over this period were about 0.01 and 0.03 respectively. Since $A_t^i = A_t^j$ for all $t$ we have $\gamma^i = \gamma^j$. Suppose that this rate is 0.02, say, and that $\delta^i = \delta^j = 0.05$. This gives $s^i\kappa^j/(s^j\kappa^i) \approx 1$. Since payments to factors of production suggest that $\alpha/(1-\alpha)$ is neither very large nor very small, output per worker in the two countries is predicted to be roughly equal.

This is only an elementary calculation. The computation of investment rates in Tanzania is not very reliable. There are issues in terms of the relative ratios of consumption and investment good prices in the two countries which may distort the data. Further, we have not included intangible capital—most notably human capital. The rate of investment in human capital and training in the US is larger than it is in Tanzania. Nevertheless, it is difficult to get the term in (8) to contribute a factor of much more than 4 or 5—certainly not 70.[25]

However the calculations are performed, it turns out that to explain the ratio of incomes in countries such as Tanzania and the US, productivity residuals must absorb most of the variation. In other words, the convex neoclassical growth model cannot be reconciled with the cross-country income data *unless we leave most of the variation in income to an unexplained residual term about which we have no quantitative theory.* And surely *any* scientific theory can explain *any* given phenomenon by adopting such a strategy.

Different authors have made this same point in different ways. Lucas (1990) notes that if factor input differences *are* large enough to explain cross-country variations in income, the returns to investment in physical and human capital in poor countries implied by the model will be huge compared to those found in the rich. They are not. Also, productivity residuals are growing quickly in countries like the US.[26] On the other hand, in countries like Tanzania, growth in the productivity residual has been very small.[27] Yet the convex neoclassical model provides no theory on why these different rates of growth in productivity should hold.

On balance, the importance of productivity residuals suggests that the poor countries are not rich because for one reason or another they have failed or not been able to adopt modern techniques of production. In fact production technology in the poorest countries is barely changing. In West Africa, for example, almost 100% of the increase in per capita food output since 1960 has

[25] See in particular Prescott (1998) for detailed calculations. He concludes that convex neoclassical growth theory "fails as a theory of international income differences, even after the concept of capital is broadened to include human and other forms of intangible capital. It fails because differences in savings rates cannot account for the great disparity in per capita incomes unless investment in intangible capital is implausibly large."

[26] One can compute this directly, or infer it from the fact that interest rates in the US have shown no secular trend over the last century, in which case transitional dynamics can explain little, and therefore growth in output per worker and growth in the residual can be closely identified (King and Rebelo 1993).

[27] Again, this can be computed directly, or inferred from the fact that if it had been growing at a rate similar to the US, then income in Tanzania would have been at impossibly low levels in the recent past (Pritchett 1997).

come from expansion of harvest area (Baker 2004). On the other hand, the rich countries are becoming ever richer because of continued innovation.

Of course this only pushes the question one step back. Technological change is only a proximate cause of diverging incomes. What economists need to explain is why production technology has improved so quickly in the US or Japan, say, and comparatively little in countries such as Tanzania, Mali and Senegal.

We end this section with some caveats. First, the failure of the simple convex neoclassical model does *not* imply the existence of poverty traps. For example, we may discover successful theories that predict very low levels of the residual based on exogenous features which tend to characterize poor countries. (Although it may turn out that, depending on what one is prepared to call exogenous, the map from fundamentals to outcomes is not uniquely defined. In other words, there are multiple equilibria. In Section 4.2 some evidence is presented on this point.)

Further, none of the discussion in this section seeks to deny that factor accumulation matters. Low rates of factor accumulation are certainly correlated with poor performance, and we do not wish to enter the "factor accumulation versus technology" debate—partly because this is viewed as a contest between neoclassical and "endogenous" growth models, which is tangential to our interests, and partly because technology and factor accumulation are clearly interrelated: technology drives capital formation and investment boosts productivity.[28]

Finally, it should be emphasized that our ability to reject the elementary convex neoclassical growth model as a theory of level differences between rich and poor countries is precisely because of its firm foundations in theory and excellent quantitative properties. All of the poverty trap models we present in this survey provide far less in terms of quantitative, testable restrictions that can be confronted with the data. The power of a model depends on its falsifiability, not its potential to account for every data set.

[28] However, as we stressed at the beginning of this section, to the extent that factor accumulation is important it may in fact turn out that low accumulation rates are mere symptoms of poverty, not causes.

### 3.3 Poverty traps: historical self-reinforcement

How then are we to explain the great variation in cross-country incomes such as shown in Figure 1? In the introduction we discussed some deviations from the neoclassical benchmark which can potentially account for this variation by endogenously reinforcing small initial differences. Before going into the specifics of different feedback mechanisms, this section formulates the first of two abstract poverty trap models. For both models a detailed investigation of microfoundations is omitted. Instead, our purpose is to establish a framework for the questions poverty traps raise about dynamics, and for their observable implications in terms of the cross-country income data.

The first model—a variation on the convex neoclassical growth model discussed in Section 3.1—is loosely based on Romer (1986) and Azariadis and Drazen (1990). It exemplifies what Mookherjee and Ray (2001) have called *historical* self-reinforcement, a process whereby initial conditions of the endogenous variables can shape long run outcomes. Leaving aside all serious complications for the moment, let us fix at $s > 0$ the savings rate, and at zero the rates of exogenous technological progress $\gamma$ and population growth $n$. Let all labor be undifferentiated and normalize its total mass to 1, so that $k$ represents both aggregate capital and capital per worker. Suppose that the productivity parameter $A$ can vary with the stock of capital. In other words, $A$ is a function of $k$, and aggregate returns $k_t \mapsto A(k_t)k_t^\alpha$ are potentially increasing.[29]

The law of motion for the economy is then

$$k_{t+1} = sA(k_t)k_t^\alpha \xi_{t+1} + (1 - \delta)k_t. \tag{9}$$

Depending on the specification of the relationship between $k$ and productivity, many dynamic paths are possible. Some of them will lead to poverty traps. Figure 7 gives examples of potential dynamic structures. For now the shock $\xi$ is suppressed. The $x$-axis is current capital $k_t$ and the $y$-axis is $k_{t+1}$. In each case the plotted curve is just the right hand side of (9), all with different maps $k \mapsto A(k)$.

[29] In Romer (1986), for example, private investment generates new knowledge, some of which enters the public domain and can be used by other firms. In Azariadis and Drazen (1990) there are spillovers from human capital formation. See also Durlauf (1993) and Zilibotti (1995). See Matsuyama (1997) and references for discussion of how investment may feed back via pecuniary externalities into specialization and hence productivity. Our discussion of microfoundations begins in Section 5.

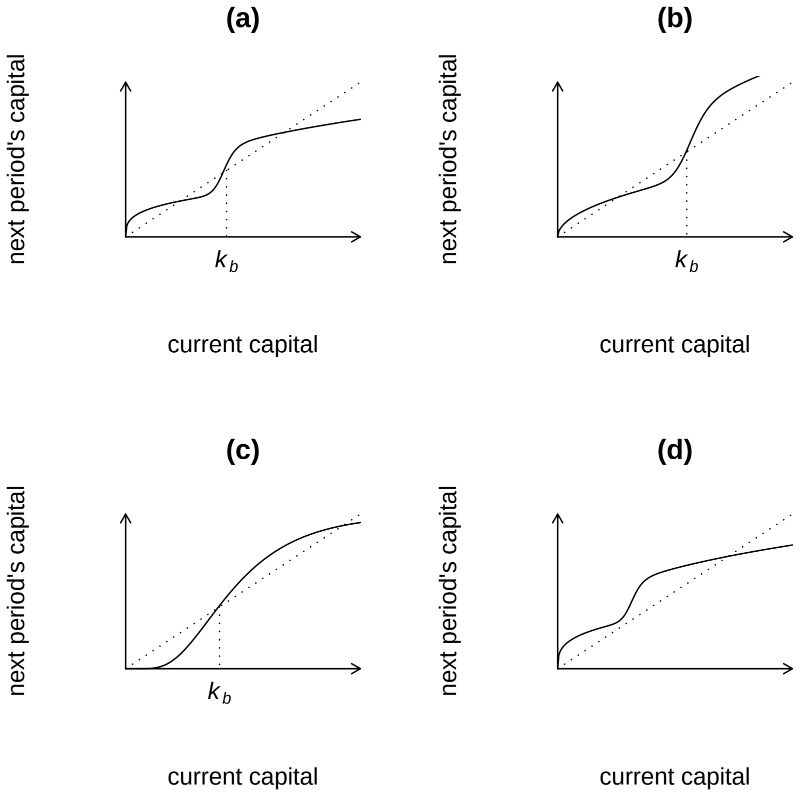


Fig. 7.

In part (a) of the figure the main feature is non-ergodic dynamics: long run outcomes depend on the initial condition. Specifically, there are two local attractors, the basins of attraction for which are delineated by the unstable fixed point $k_b$. Part (b) is also non-ergodic. It shows the same low level attractor, but now no high level attractor exists. Beginning at a state above $k_b$ leads to unbounded growth. In part (c) the low level attractor is at zero.

The figure in part (d) looks like an anomaly. Since the dynamics are formally ergodic, many researchers will not view this structure as a "poverty trap" model. Below we argue that this reading is too hasty: the model in (d) can certainly generate the kind of persistent-poverty aggregate income data we are hoping to explain.

In order to gain a more sophisticated understanding, let us now look at the stochastic dynamics of the capital stock. Deterministic dynamics are of course a special case of stochastic dynamics (with zero-variance shocks) but as in the case of the neoclassical model above, let us suppose that $(\xi_t)_{t \geq 0}$ is independently and identically lognormally distributed, with $\ln \xi \sim N(\mu, \sigma)$ and $\sigma > 0$. It then follows that the sequence of marginal distributions $(\psi_t)_{t \geq 0}$ for the capital stock sequence $(k_t)_{t\geq 0}$ again obeys the recursion (4) where the stochastic kernel $\Gamma$ is now

$$\Gamma(k,k') := \varphi\left[\frac{k'-(1-\delta)k}{sA(k)k^{\alpha}}\right]\frac{1}{sA(k)k^{\alpha}}, \tag{10}$$

with $\varphi$ the lognormal density on $(0,\infty)$ and zero elsewhere. All of the intuition for the recursion (4) and the construction of the stochastic kernel (10) is exactly the same as the neoclassical case.

How do the marginal distributions of the nonconvex growth model evolve? The following result gives the answer for most cases we are interested in.

**Proposition 3.1** *Let* $(\xi_t)_{t\geq 0}$ *be an independent sequence with* $\ln \xi_t \sim N(\mu,\sigma)$ *for all* $t$*, and let* $\sigma > 0$*. If* $k \mapsto A(k)$ *satisfies the regularity condition*

$$0 < \inf_k A(k) \leq \sup_k A(k) < \infty,$$

*then the stochastic nonconvex growth model defined by* (9) *is ergodic.*[30]

Ergodicity here refers to Definition 3.1 on page 18, which, incidentally, is the standard definition used in growth theory and macroeconomics (see, for example, Brock and Mirman 1972; or Stokey, Lucas and Prescott 1989). In other words, there is a unique ergodic distribution $\psi^*$, and the sequence of marginal distributions $(\psi_t)_{t\geq 0}$ converges to $\psi^*$ asymptotically, independent of the initial condition (assuming of course that $k_0 > 0$). A proof of this result is given in the technical appendix.

So why has a non-ergodic model become ergodic with the introduction of noise? The intuition is completely straightforward: Under our assumption of unbounded shocks there is always the potential—however small—to escape any basin of attraction. So in the long run initial conditions do not matter. (What *does* matter is how long this long run is, a point we return to below.)

Figure 8 gives the ergodic distributions corresponding to two poverty trap models.[31] Both have the same structural dynamics as the model in part (a) of Figure 7. The left hand panels show this structure with the shock suppressed. The right hand panels show corresponding ergodic distributions under the

[30] In fact we require also that $k \mapsto A(k)$ is a Borel measurable function. But this condition is very weak indeed. For example, $k \mapsto A(k)$ need be neither monotone nor continuous.

[31] Regarding numerical computation see the discussion for the neoclassical case above.

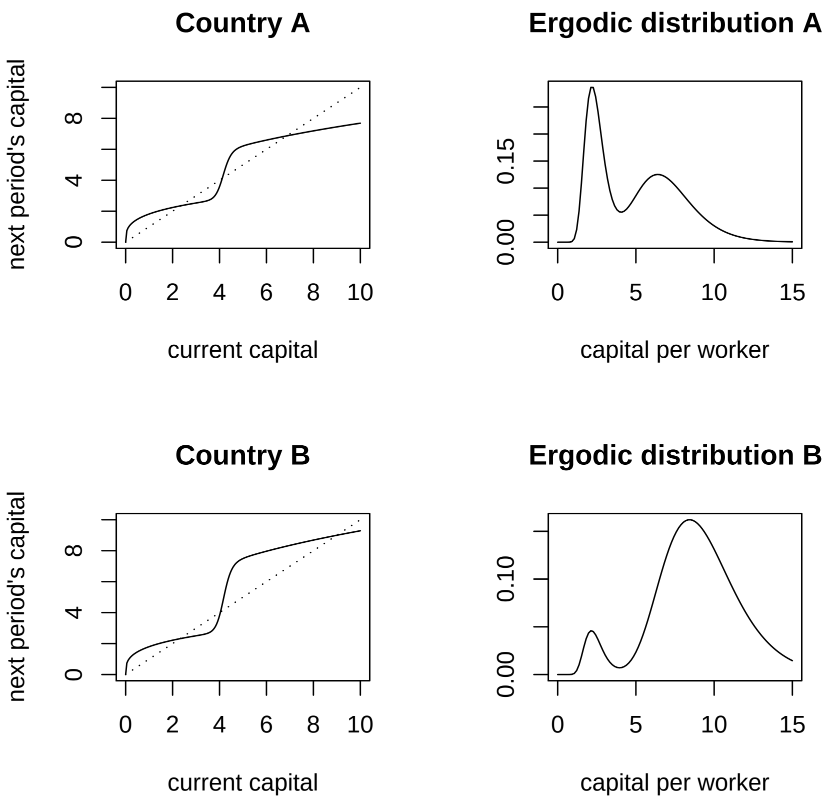


Fig. 8.

independent lognormal shock process. Both ergodic distributions are bimodal, with modes concentrated around the deterministic local attractors.

Comparing the two left hand panels, notice that although qualitatively similar, the laws of motion for Country A and Country B have different degrees of increasing returns. For Country B, the jump occurring around $k = 4$ is larger. As a result, the state is less likely to return to the neighborhood of the lower attractor once it makes the transition out of the poverty trap. Therefore the mode of the ergodic distribution corresponding to the higher attractor is large relative to that of Country A. Economies driven by law of motion B spend more time being rich.

Convergence to the ergodic distribution in a nonconvex growth model is illustrated in Figure 9. The underlying model is (a) of Figure 7.[32] As before, the

[32] The specification of $A(k)$ used in the simulation is $A(k) = a\exp(h\Psi(k))$, where $a = 15$, $h = 0.52$ and the transition function $\Psi$ is given by $\Psi(k) := (1+\exp(-\ln(k/k_T)/\theta))^{-1}$. The parameter $k_T$ is a "threshold" value of $k$, and is set at 6.9. The parameter $\theta$ is the smoothness of the transition, and is set at 0.09. The

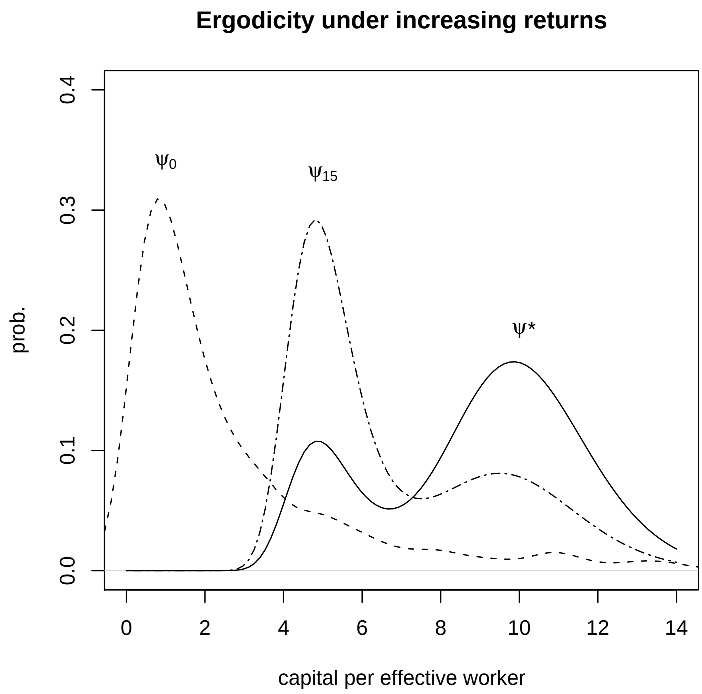


Fig. 9.

ergodic distribution is bimodal. In this simulation, the initial distribution was chosen arbitrarily. Note how initial differences tend to be magnified over the medium term despite ergodicity. The initially rich diverge to the higher mode, creating the kind of "convergence club" effect already seen in $\psi_{15}$, the period 15 marginal distribution.[33]

It is clear, therefore, that ergodicity is not the whole story. If the support of the shock $\xi$ is bounded then ergodicity may not hold. Moreover, even with ergodicity, historical conditions may be arbitrarily persistent. Just how long they persist depends mainly on (i) the size of the basins of attraction and (ii) the statistical properties of the shock. On the other hand, the non-zero degree of mixing across the state space that drives ergodicity is usually more realistic than deterministic models where poverty traps are absolute and can never be overcome. Indeed, we will see that ergodicity is very useful for framing empirical questions in Section 4.2.

Figures 10 and 11 illustrate how historical conditions persist for individual

---

other parameters are $\alpha = 0.3$, $s = 0.2$, $\delta = 1$, and $\ln \xi \sim N(0, 0.1)$.

[33] Incidentally, the change in the distributions from $\psi_0$ to $\psi_{15}$ is qualitatively quite similar to the change in the cross-country income distribution that has been observed in the post war period.

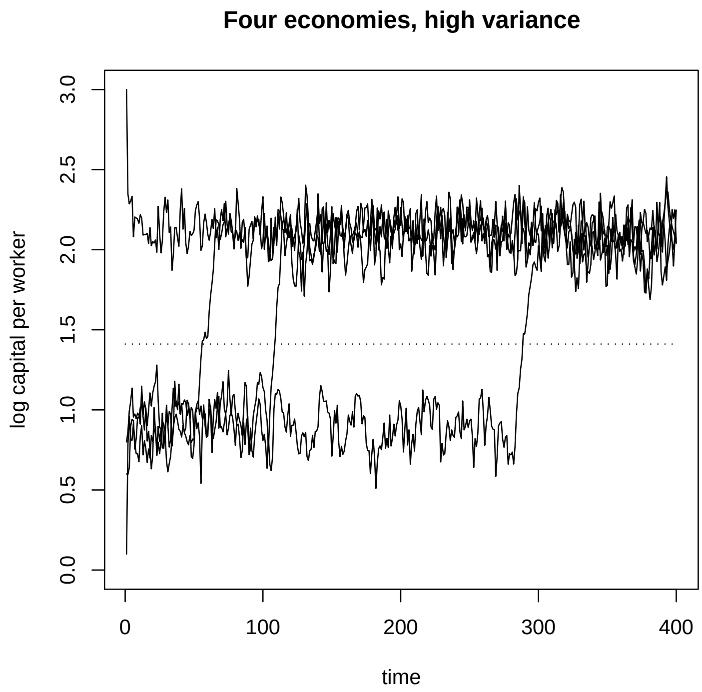


Fig. 10.

time series generated by a model in the form of (a) of Figure 7, regardless of ergodicity. In both figures, the $x$-axis is time and the $y$-axis is (the log of) capital stock per worker. The dashed line through the middle of the figure corresponds to (the log of) $k_b$, the point dividing the two basins of attraction in (a) of Figure 7. Both figures show the simulated time series of four economies. In each figure, all four economies are identical, apart from their initial conditions. One economy is started in the basin of attraction for the higher attractor, and three are started in that of the lower attractor. [34]

In the figures, the economies spend most of the time clustered in the neighborhoods of the two deterministic attractors. Economies starting in the portion of the state space (the $y$-axis) above the threshold are attracted on average to the high level attractor, while those starting below are attracted on average to the low level attractor. For these parameters, historical conditions are important in determining outcomes over the kinds of time scales economists are interested in, even though there are no multiple equilibria, and in the limit outcomes depend only on fundamentals.

[34] The specification of $A(k)$ is as in Figure 9, where now $k_T = 4.1$, $\theta = 0.2$, $h = 0.95$, $\alpha = 0.3$, $s = 0.2$, $\delta = 1$. For Figure 10 we used $\ln \xi \sim N(0, 0.1)$, while for Figure 11 we used $\ln \xi \sim N(0, 0.05)$ .

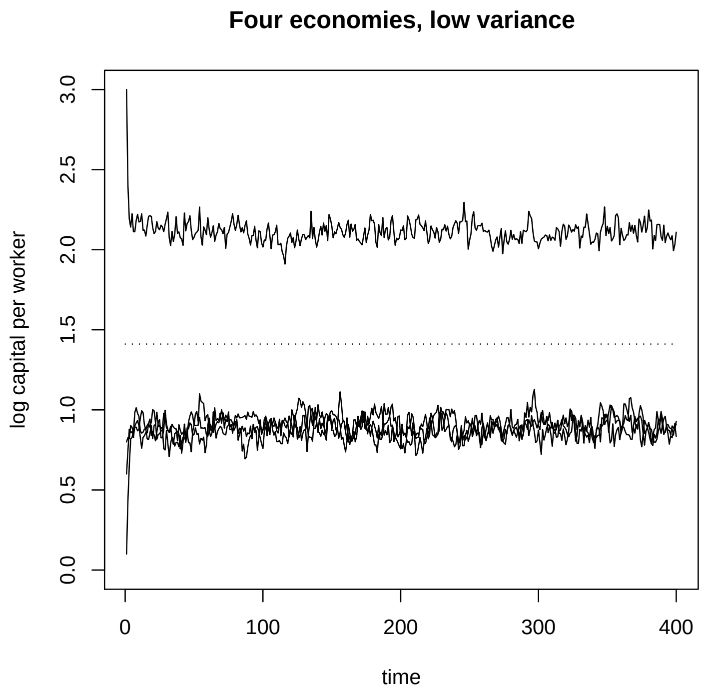


Fig. 11.

In Figure 10, all three initially poor economies eventually make the transition out of the poverty trap, and converge to the neighborhood of the high attractor. Such transitions might be referred to as “growth miracles.” In these series there are no “growth disasters” (transitions from high to low). The relative likelihood of growth miracles and growth disasters obviously depends on the structure of the model—in particular, on the relative size of the basins of attraction.

In Figure 10 the shock is distributed according to $\ln \xi \sim N(0, 0.1)$, while in Figure 11 the variance is smaller: $\ln \xi \sim N(0, 0.05)$. Notice that in Figure 11 no growth miracles occur over this time period. The intuition is clear: With less noise, the probability of a large positive shock—large enough to move into the basin of attraction for the high attractor—is reduced, and with it the probability of escaping from the poverty trap.

We now return to the model in part (d) of Figure 7, which is nonconvex, but at the same time is ergodic even in the deterministic case. This kind of structure is usually *not* regarded as a poverty trap model. In fact, since (d) is just a small perturbation of model (a), the existence of poverty traps is often thought to be very sensitive to parameters—a small change can cause a bifurcation of the dynamics whereby the poverty trap disappears. But, in

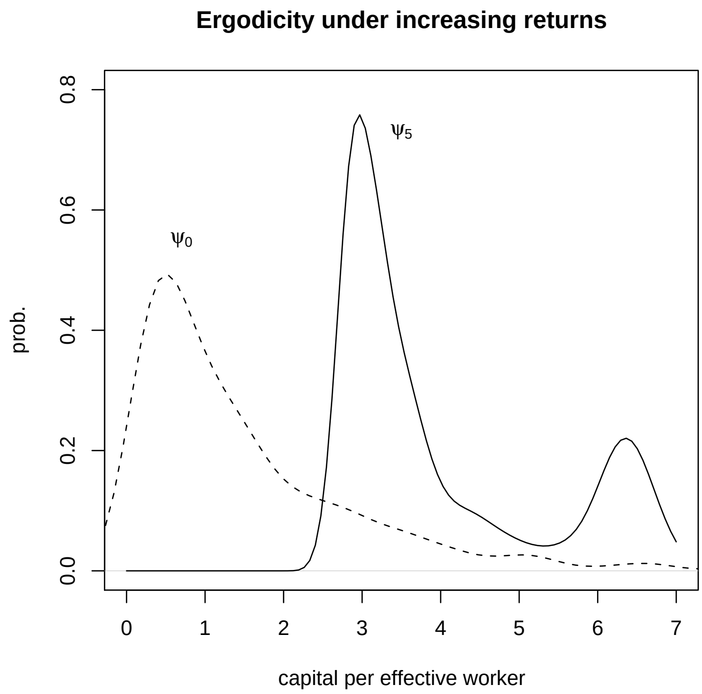


Fig. 12.

fact, the phenomenon of persistence is more subtle. In terms of their medium run implications for cross-country income patterns, the two models (a) and (d) are very similar.

To illustrate this, Figure 12 shows an arbitrary initial distribution and the resulting time 5 distribution for $k$ under the law of motion given in (d) of Figure 7.[35] As in all cases we have considered, the stochastic model is ergodic. Now the ergodic distribution (not shown) is unimodal, clustered around the single high level attractor of the deterministic model. Thus the long run dynamics are different to those in Figure 9. *However,* during the transition, statistical behavior is qualitatively the same as that for models that *do* have low level attractors (such as (a) of Figure 7). In $\psi_5$ we observe amplification of initial differences, and the formation of a bimodal distribution with two "convergence clubs."

How long is the medium run, when the transition is in progress and the distribution is bimodal? In fact one can make this transition arbitrarily long without changing the basic qualitative features of (d), such as the non-existence of a low level attractor. Its length depends on the degree of nonconvexity and the

[35] The specification of $A(k)$ is as before, where now $k_T = 3.1$, $\theta = 0.15$, $h = 0.7$, $\alpha = 0.3$, $s = 0.2$, $\delta = 1$, and $\ln \xi \sim N(0, 0.2)$.

variance of the productivity shocks $(\xi_t)_{t\geq 0}$. Higher variance in the shocks will tend to speed up the transition.

Incidentally, the last two examples have illustrated an important general principle: In economies with nonconvexities, the dynamics of key variables such as income can be highly sensitive to the statistical properties of the exogenous shocks which perturb activity in each period.[36] This phenomenon is consistent with the cross-country income panel. Indeed, several studies have emphasized the major role that shocks play in determining the time path of economic development (c.f., e.g., Easterly, Kremer, Pritchett and Summers 1993; den Haan 1995; Acemoglu and Zilibotti 1997; Easterly and Levine 2000).[37]

At the risk of some redundancy, let us end our discussion of the increasing returns model (9) by reiterating that persistence of historical conditions and formal ergodicity may easily coincide. (Recall that the time series in Figure 11 are generated by an ergodic model, and that (d) of Figure 7 is ergodic even in the deterministic case.) As a result, identifying history dependence with a lack of ergodicity can be problematic. In this survey we use a more general definition:

**Definition 3.2 (Poverty trap)** *A poverty trap is any self-reinforcing mechanism which causes poverty to persist.*

When considering a given quantitative model and its dynamic implications, the important question to address is, how persistent are the self-reinforcing mechanisms which serve to lock in poverty over the time scales that matter when welfare is computed?[38]

A final point regarding this definition is that the mechanisms which reinforce poverty may occur at any scale of social and spatial aggregation, from individuals to families, communities, regions, and countries. Traps can arise not just across geographical location such as national boundaries, but also within dispersed collections of individuals affiliated by ethnicity, religious beliefs or clan. Group outcomes are then summed up progressively from the level of the

[36] Such sensitivity is common to all dynamic systems where feedbacks can be positive. The classic example is evolutionary selection.

[37] This point also illustrates a problem with standard empirical growth studies. In general no information on the shock distribution is incorporated into calculation of dynamics.

[38] Mookherjee and Ray (2001) have emphasized the same point. See their discussion of “self-reinforcement as slow convergence.”

individual.[39]

## 3.4 *Poverty traps: inertial self-reinforcement*

Next we turn to our second "canonical" poverty trap model, which again is presented in a very simplistic form. (For microfoundations see Sections 5–8.) The model is static rather than dynamic, and exhibits what Mookherjee and Ray (2001) have described as *inertial* self-reinforcement.[40] Multiple equilibria exist, and selection of a particular equilibrium can be determined purely by beliefs or subjective expectations.

In the economy a unit mass of agents choose to work either in a traditional, rural sector or a modern sector. Labor is the only input to production, and each agent supplies one unit in every period. All markets are competitive. In the traditional sector returns to scale are constant, and output per worker is normalized to zero. The modern sector, however, is knowledge-intensive, and aggregate output exhibits increasing returns due perhaps to spillovers from agglomeration, or from matching and network effects.

Let the fraction of agents working in the modern sector be denoted by $\alpha$. The map $\alpha \mapsto f(\alpha)$ gives output per worker in the modern sector as a function of the fraction employed there. Payoffs are just wages, which equal output per worker (marginal product). Agents maximize individual payoffs taking the share $\alpha$ as exogenously given.

We are particularly interested in the case of strategic complementarities. Here, entry into the modern sector exhibits complementarities if the payoff to entering the modern sector increases with the number of other agents already there; in other words, if $f$ is increasing. We assume that $f' > 0$, and also that returns in the modern sector dominate those in the traditional sector only when the number of agents in the modern sector rises above some threshold. That is, $f(0) < 0 < f(1)$. This situation is shown in Figure 13. At the point $\alpha_b$ returns in the two sectors are equal.

Equilibrium distributions of agents are values of $\alpha$ such that $f(\alpha) = 0$, as well as "all workers are in the traditional sector," or "all workers are in the modern

[39] This point has been emphasized by Barrett and Swallow (2003) in their discussion of "fractal" poverty traps.
[40] By "static" we mean that there are no explicitly specified interactions between separate periods.


Fig. 13.

sector" (ignoring adjustments on null sets). The last two of these are clearly Pareto-ranked: The equilibrium $\alpha = 0$ has the interpretation of a poverty trap.

Immediately the following objection arises. Although the lower equilibrium is to be called a poverty trap, is there really a self-reinforcing mechanism here which causes poverty to persist? After all, it seems that as soon as agents coordinate on the good equilibrium "poverty" will disappear. And there are plenty of occasions where societies acting collectively have put in place the institutions and preconditions for successful coordination when it is profitable to do so.

Although the last statement is true, it seems that history still has a role to play in equilibrium selection. This argument has been discussed at some length in the literature, usually beginning with myopic Marshallian dynamics, under which factors of production move over time towards activities where returns are higher. In the case of our model, these dynamics are given by the arrows in Figure 13. If $(\alpha_0)_{t\geq 0}$ is the sequence of modern sector shares, and if initially $\alpha_0 < \alpha_b$, then $\alpha_t \to 0$. Conversely, if $\alpha_0 > \alpha_b$, then $\alpha_t \to 1$.

But, as many authors have noted, this analysis only pushes the question one step back. Why should the sectoral shares only evolve slowly? And if they can adjust instantaneously, then why should they depend on the initial condition at all? What are the sources of inertia here that prevent agents from immediately coordinating on good equilibria?[41]

Adsera and Ray (1997) have proposed one answer. Historical conditions may be decisive if—as seems quite plausible—spillovers in the modern sector arise only with a lag. A simplified version of the argument is as follows. Suppose that the private return to working in the modern sector is $r_t$, where now $r_0 = f(\alpha_0)$ and $r_t$ takes the lagged value $f(\alpha_{t-1})$ when $t \geq 1$. Suppose also that at the end of each period agents can move costlessly between sectors. Agent $j$ chooses location in order to maximize a discounted sum of payoffs given subjective beliefs $(\alpha_t^j)_{t\geq 0}$ for the time path of shares, where to be consistent we require that $\alpha_0^j = \alpha_0$ for all $j$.

Clearly, if $\alpha_0 < \alpha_b$, then switching to or remaining in the traditional sector at the end of time zero is a dominant strategy regardless of beliefs, because $r_1 = f(\alpha_0) < f(\alpha_b) = 0$. The collective result of these individual decisions is that $\alpha_1 = 0$. But then $\alpha_1 < \alpha_b$, and the whole process repeats. Thus $\alpha_t = 0$ for all $t \geq 1$. This outcome is interesting, because even the most optimistic set of beliefs lead to the low equilibrium when $f(\alpha_0) < 0$. To the extent that Adsera and Ray's analysis is correct, history must always determine outcomes.[42]

Another way that history can re-enter the equation is if we admit some deviation from perfect rationality and perfect information. As was stressed in the introduction, this takes us back to the role of institutions, through which history is transmitted to the present.

It is reasonable to entertain such deviations here for a number of reasons. First and foremost, assumptions of complete information and perfect rationality are usually justified on the basis of experience. Rationality obtains by repeated observation, and by the punishment of deviant behavior through the carrot and stick of economic payoff. Rational expectations are justified by appealing to *induction.* Agents are assumed to have had many observations from a stationary environment. Laws of motion and hence conditional expectations are inferred on the basis of repeated transition sampling *from every relevant state/action pair* (Lucas 1986). When attempting to break free from a poverty

[41] See, for example, Krugman (1991) or Matsuyama (1991).

[42] There are a number of possible criticisms of the result, most of which are discussed in detail by the authors. If, for example, there are congestion costs or first mover advantages, then moving immediately to the modern sector might be rational for some optimistic beliefs and specification of parameters.

trap, however, agents have most likely *never* observed a transition to the high level equilibrium. On the basis of what experience are they to assess its likelihood from each state and action? How will they assess the different costs or benefits?

In a boundedly rational environment with limited information, outcomes will be driven by norms, institutions and conventions. It is likely that *these* factors are among the most important in terms of a society's potential for successful coordination on good equilibria. In fact for some models we discuss below the equilibrium choice is not between traditional technology and the modern sector, but rather is a choice between predation (corruption) and production, or between maintaining kinship bonds and breaking them. In some sense these choices are inseparable from the social norms and institutions of the societies within which they are framed.[43]

The central role of institutions may not prevent rapid, successful coordination on good equilibria. After all, institutions and conventions are precisely how societies solve coordination problems. As was emphasized in the introduction, however, norms, institutions and conventions are path dependent by definition. And, in the words of Matsuyama (1995, p. 724), "coordinating expectations is much easier than coordinating changes in expectations." Because of this, economies that start out in bad equilibria may find it difficult to break free.

Why should a convention that locks an economy into a bad equilibrium develop in the first place? Perhaps this is just the role of historical accident. Or perhaps, as Sugden (1989) claims, conventions tend to spread on the basis of versatility or analogy.[44] If so, the conventions that propagate themselves most successfully may be those which are most versatile or susceptible to analogy—not necessarily those which lead to "good" or efficient equilibria.

Often the debate on historical conditions and coordination is cast as "history versus expectations." We have emphasized the role of history, channeled

[43] More traditional candidates for coordinating roles among the set of institutions include interventionist states promoting industrialization through policy-based financing, or (the cultural component of) large business groups, such as Japan's keiretsu and South Korea's chaebol. In Section 5.2, we discuss the potential for large banks with significant market power to drive "big push" type investments by the private sector.

[44] A versatile convention works reasonably well against many strategies, and hence is advantageous when facing great uncertainty. Analogy means that a rule for a particular situation is suggested by similar rules applied to different but related situations.

through social norms and institutions, but without intending to say that beliefs are not important. Rather, beliefs are no doubt crucial. At the same time, beliefs and expectations are shaped by history. And they in turn combine with value systems and local experience to shape norms and institutions. The latter then determine how successful different societies are in solving the particular coordination problems posed by interactions in free markets.

If beliefs and expectations are shaped by history, then the "history versus expectations" dichotomy is misleading. The argument that beliefs and expectations are indeed formed by a whole variety of historical experiences has been made by many development theorists. In an experiment investigating the effects of the Indian caste system, Hoff and Pandey (2004) present evidence that individuals view the world through their own lens of "historically created social identities," which in turn has a pronounced effect on expectations. Rostow (1990, p. 5) writes that "the value system of [traditional] societies was generally geared to what might be called a long run fatalism; that is, the assumption that the range of possibilities open to one's grandchildren would be just about what it had been for one's grandparents." Ray (2003, p. 1) argues that "poverty and the failure of aspirations may be reciprocally linked in a self-sustaining trap."

Finally, experimental evidence on coordination games with multiple Pareto-ranked equilibria suggests that history is important: Outcomes are strongly path dependent. For example, Van Huyck, Cook and Battalio (1997) study people's adaptive behavior in a generic game of this type, where multiple equilibria are generated by strategic complementarities. In each experiment, eight subjects participated in a sequence of between 15 and 40 plays. The authors find sensitivity to initial conditions, defined here as the median of the first round play. In their view, "the experiment provides some striking examples of coordination failure growing from small historical accidents."

## 4 Empirics of Poverty Traps

Casual observation of the cross-country income panel tends to suggest mechanisms which reinforce wealth or poverty. In Section 4.1 we review the main facts. Section 4.2 considers tests for the empirical relevance of poverty trap models. While the results of the tests support the hypothesis that the map from fundamentals to economic outcomes is not unique, they give no indication as to what forces might be driving multiplicity. Section 4.3 begins the

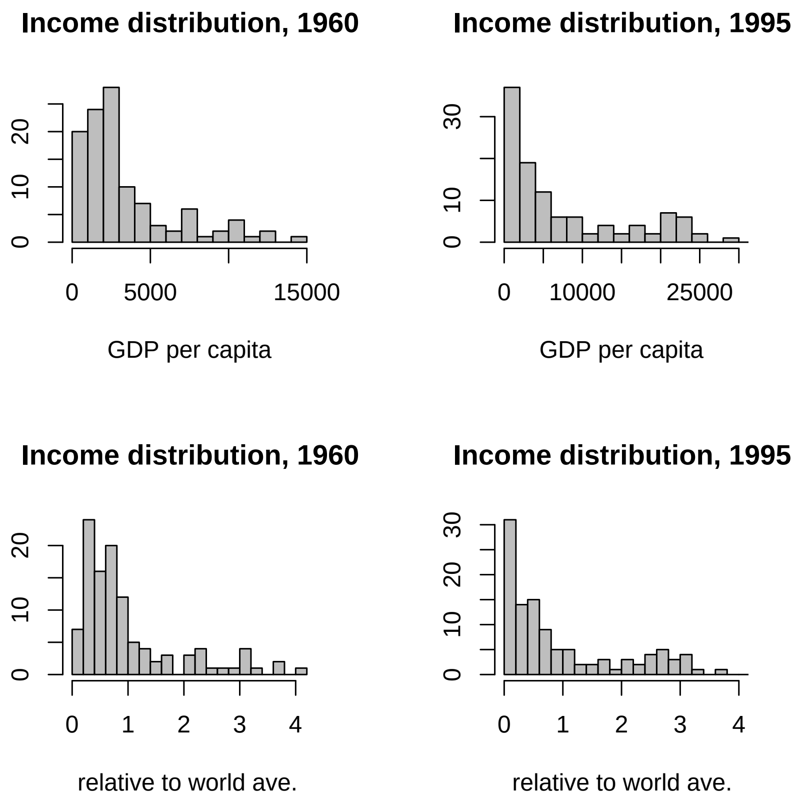


Fig. 14.

difficult task of addressing this issue in a macroeconomic framework. Finally, Section 4.4 gives references to empirical tests of specific microeconomic mechanisms that can reinforce poverty at the individual or group level.

### *4.1 Bimodality and convergence clubs*

A picture of the evolving cross-country income distribution is presented in Figure 14. For both the top and bottom histograms the $y$-axis measures frequency. For the top pair (1960 and 1995) the $x$-axis is GDP per capita in 1996 PPP adjusted dollars. This is the standard histogram of the cross-country income distribution. For the bottom pair the $x$-axis represents income as a fraction of the world average for that year.

The single most striking feature of the absolute income histograms for 1960 and 1995 is that over this period a substantial fraction of poor countries have grown very little or not at all. At the same time, a number of middle income countries have grown rapidly, in some cases fast enough to close in on the rich. Together, these forces have caused the distribution to become somewhat thinner in the middle, with probability mass collecting at the two extremes. Such an outcome is consistent with mechanisms that accentuate differences in initial conditions, and reinforce wealth or poverty (Azariadis and Drazen 1990; Quah 1993, 1996; Durlauf and Johnson 1995; Bianchi 1997; Pritchett 1997; Desdoigts 1999; Easterly and Levine 2000).

As well as observing past and present distributions, Quah (1993) also used the Penn World Tables to estimate a transition probability matrix by discretizing the state space (income per capita), treating all countries as observations from the same Markovian probability law, and measuring transition frequency. This matrix provides information on mobility. Also, by studying the ergodic distribution, and by multiplying iterations of the matrix with the current cross-country income distribution, some degree of inference can be made as to where the income distribution is heading.

In his calculation, Quah uses per capita GDP relative to the world average over the period 1962 to 1984 in a sample of 118 countries. Relative income is discretized into state space $S := \{1, 2, 3, 4, 5\}$ consisting of 5 "bins," with states corresponding to values for relative GDP of 0–0.25, 0.25–0.5, 0.5–1, 1–2 and 2–$\infty$ respectively. The transition matrix $\mathbf{P} = (p_{ij})$ is computed by setting $p_{ij}$ equal to the fraction of times that a country, finding itself in state $i$, makes the transition to state $j$ the next year. The data is assumed to be stationary, so that all of the transitions can be pooled when calculating transition probabilities. The result of this calculation (Quah 1993, p. 431) is

$$\mathbf{P} = \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix}.$$

The Markov chain represented by $\mathbf{P}$ is easily shown to be ergodic, in the sense that there is a unique $\psi^* \in \mathscr{P}(S)$, the distributions on $S$, with the property that $\psi^*\mathbf{P} = \psi^*$, and $\psi\mathbf{P}^t \to \psi^*$ as $t \to \infty$ for all $\psi \in \mathscr{P}(S)$.[45] Quah calculates this ergodic distribution $\psi^*$ to be $(0.24, 0.18, 0.16, 0.16, 0.27)$.

[45] Following Markov chain convention we are treating the distributions in $\mathscr{P}(S)$ as row vectors. Also, $\mathbf{P}^t$ is $t$ compositions of $\mathbf{P}$ with itself. For more discussion of ergodicity see the technical appendix, or Stachurski (2004).

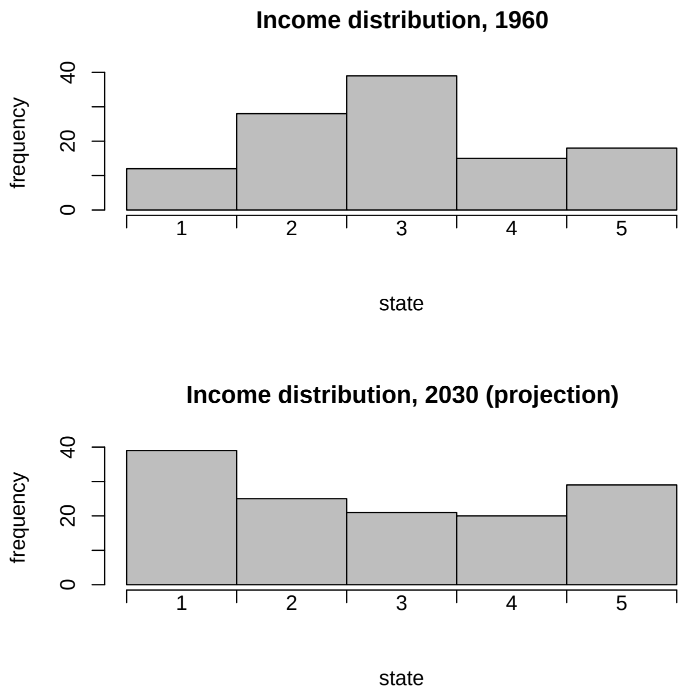


Fig. 15.

The ergodic distribution is quite striking, in that the world is divided almost symmetrically into two convergence clubs of rich and poor at either end of the income distribution.

It is not immediately clear just how long the long run is. To get some indication, we can apply $\mathbf{P}^t$ to the current distribution for different values of $t$. Figure 15 shows the results of applying $\mathbf{P}^{30}$ to the year 2000 income distribution from the Penn World Tables. This gives a projection for the 2030 distribution. Contrasted with the 1960 distribution the prediction is strongly bimodal.

As Quah himself was at pains to emphasize, the projections carried out above are only a first pass at income distribution dynamics, with many obvious problems. One is that the dynamics generated by a discretized version of a continuous state Markov chain can deviate very significantly from the true dynamics generated by the original chain, and error bounds are difficult to quantify.[46] Also, since the estimation of $\mathbf{P}$ is purely nonparametric, the projections do not contain any of the restrictions implied by growth theory.

Quah (1996) addressed the first of these problems by estimating a continuous

[46] Compare, for example, Feyrer (2003) and Johnson (2004).

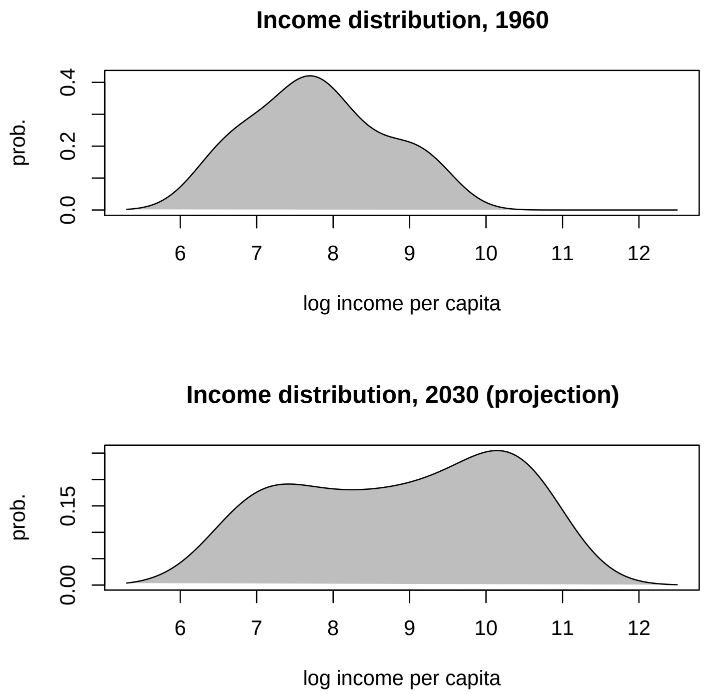


Fig. 16.

state version. In the language of this survey, he estimates a stochastic kernel $\Gamma$, of which $\mathbf{P}$ is a discretized representation. The estimation is nonparametric, using a Parzen-window type density smoothing technique. The kernel is suggestive of considerable persistence.

Azariadis and Stachurski (2004) make some effort to address both the discretization problem and the lack of economic theory simultaneously, by estimating $\Gamma$ *parametrically*, using a theoretical growth model. In essence, they estimate equation (9), where $k \mapsto A(k)$ is represented by a three-parameter logistic function. The logistic function nests a range of growth models, from the convex model in Figure 2 to the nonconvex models in Figure 7, panels (a), (b) and (d). Once the law of motion (9) is estimated, the stochastic kernel $\Gamma$ is calculated via equation (10), and the projection of distributions is computed by iterating (4).

The resulting 2030 prediction is shown in Figure 16, with the 1960 distribution drawn above for comparison. The $x$-axis is log of real GDP per capita in 1996 US dollars. The 1960 density is just a smoothed density estimate using Gaussian kernels, with data from the Penn World Tables. The same data was used to estimate the parameters in the law of motion (9). As in Figure 15, a unimodal distribution gives way to a bimodal distribution.

These findings do lend some support to Quah's convergence club hypothesis. Much work remains to be done. For example, in all of the methodologies discussed above, nonstationary data is being fitted to a stationary Markov chain. This is clearly a source of bias. Furthermore, all of these models are too small, in the sense that the state space used in the predictions are only one-dimensional. [47]

### *4.2 Testing for existence*

Poverty trap models tend to be lacking in testable quantitative implications. Where there are multiple equilibria and sensitive dependence to initial conditions, outcomes are much harder to pin down than when the map from parameters to outcomes is robust and unique. This has led many economists to question the empirical significance of poverty trap models. [48] In this section, we ask whether or not there is any evidence that poverty traps exist.

In answering this question, one must be very careful to avoid the following circular logic: First, persistent poverty is observed. Poverty traps are then offered as the explanation. But how do we know there are poverty traps? Because (can't you see?) poverty persists. [49] This simple point needs to be kept in mind when interpreting the data with a view to assessing the empirical relevance of the models in this survey. Persistent poverty, emergent bimodality and the dispersion of cross-country income are the phenomena we seek to explain. They cannot themselves be used as proof that poverty traps explain the data.

Also, a generalized convex neoclassical model can certainly be the source of bimodality and dispersion if we accept that the large differences in total factor

[47] In fact within each economy there are many interacting endogenous variables, only one of which is income. Even if the process as a whole is stationary and Markov, projection of the system onto one dimension will yield a process which is not generally Markovian. Moreover, there are interactions between countries that affect economic performance, and these interactions are important. A first-best approach would be to treat the world economy as an $N \times M$-dimensional Markov process, where $N$ is the number of countries, and $M$ is the number of endogenous variables in each country. One would then estimate the stochastic kernel $\Gamma$ for this process, a map from $\mathbb{R}_+^{N \times M} \times \mathbb{R}_+^{N \times M} \to [0, \infty)$. Implications for the cross-country income distribution could be calculated by computing marginals.
[48] See Matsuyama (1997) for more discussion of this point.
[49] Recall Karl Popper's famous tale about Neptune and the sea.

productivity residuals across countries are due to some exogenous force, the precise nature of which is still waiting to be explained. In this competing explanation, the map from fundamentals to outcomes is unique, and shocks or historical accidents which perturb the endogenous variables can safely be ignored.

The central question, then, is *whether or not the poverty trap explanation of cross-country income differentials survives if we control for the exogenous forces which determine long run economic performance.* In other words, do self-reinforcing and path dependent mechanisms imply that economies populated by fundamentally similar people in fundamentally similar environments can support very different long run outcomes? What empirical support is there for such a hypothesis?

One particularly interesting study which addresses this question is that of Bloom, Canning and Sevilla (2003). Their test is worth discussing in some detail. To begin, consider again the two multiple equilibria models shown in Figure 8 (page 28), along with their ergodic distributions. As can be seen in the left hand panels, when the shock is suppressed both Country A and Country B have two locally stable equilibria for capital per worker—and therefore two locally stable equilibria for income. Call these two states $y_1^*$ and $y_2^*$, the first of which is interpreted as the poverty trap.

In general, $y_1^*$ and $y_2^*$ will depend on the vector of exogenous fundamentals, which determine the exact functional relationships in the model, and hence become parameters in the law of motion. Let this vector be denoted by $\mathbf{x}$. Consider a snapshot of the economy at some point in time $t$. We can write income per capita as

$$y = \begin{cases} y_1^*(\mathbf{x}) + u_1 & \text{with probability } p(\mathbf{x}); \\ y_2^*(\mathbf{x}) + u_2 & \text{with probability } 1 - p(\mathbf{x}). \end{cases} \tag{R2}$$

Here $p(\mathbf{x})$ is the probability that the country in question is in the basin of attraction for the lower equilibrium $y_1^*(\mathbf{x})$ at time $t$. This probability is determined by the time $t$ marginal distribution of income. The shock $u_i$ represents deviation from the deterministic attractor at time $t$.

Figure 8 (page 28) helps to illustrate how $y_1^*$ and $y_2^*$ might depend on the exogenous variables. Imagine that Countries A and B have characteristics $\mathbf{x}_A$ and $\mathbf{x}_B$ respectively. These different characteristics account for the different shapes of the laws of motion shown in the left hand side of the figure. As drawn, $y_2^*(\mathbf{x}_A)$, the high level attractor for Country A, is less than $y_2^*(\mathbf{x}_B)$, the high level attractor for Country B, while $y_1^*(\mathbf{x}_A)$ and $y_1^*(\mathbf{x}_B)$ are roughly equal.

In addition, we can see how the probability $p(\mathbf{x})$ of being in the poverty trap basin depends on these characteristics. For time $t$ sufficiently large, ergodicity means that the time $t$ marginal distribution—which determines this probability—can be identified with the ergodic distribution. The ergodic distribution in turn depends on the underlying structure, which depends on $\mathbf{x}$. This is illustrated by the different sizes of the distribution modes for Countries A and B in Figure 8. For Country A the left hand mode is relatively large, and hence so is $p(\mathbf{x})$.

Using a maximum likelihood ratio test, the specification (R2) is evaluated against a single regime alternative

$$y = y^*(\mathbf{x}) + u, \tag{R1}$$

which can be thought of for the moment as being generated by a convex Solow model. The great benefit of the specifications (R1) and (R2)—as emphasized by the authors—is that long run output depends only on exogenous factors. The need to specify the precise system of endogenous variables and their interactions is circumvented.[50]

In conducting the test of (R1) against (R2), it is important *not* to include as exogenous characteristics any variable which is in fact endogenously determined. For to do so might result in conditioning on the *outcomes* of the underlying process which generates multiple equilibria. In the words of the authors, "Including such variables may give the impression of a unique equilibrium relationship [for the economic system] when in reality they are a function of the equilibrium being observed. Fundamental forces must be characteristics that determine a country's economic performance but are not determined by it."

In the estimation of Bloom, Canning and Sevilla, only geographic features are included in the set of exogenous variables. These include data on distance from equator, rainfall, temperature, and percentage of land area more than 100km from the sea. For this set of variables, the likelihood ratio test rejects the single regime model (R1) in favor of the multiple equilibria model (R2). They find evidence for a high level equilibrium which does not vary with $\mathbf{x}$, and a low level equilibrium which does. In particular, $y_1^*(\mathbf{x})$ tends to be smaller for hot, dry, land-locked countries (and larger for those with more favorable geographical features). In addition, $p(\mathbf{x})$ is larger for countries with

[50] Ergodicity is critical in this respect, for without it $p$ will depend not just on $\mathbf{x}$ but also on the lagged values of endogenous variables.

unfavorable geographical features. In other words, the mode of the ergodic distribution around $y_1^*(\mathbf{x})$ is relatively large. For these economies escape from the poverty trap is more difficult.

Overall, the results of the study support the poverty trap hypothesis. They also serve to illustrate the importance of distinguishing between variables which are exogenous and those which have feedback from the system. If one conditions on "explanatory" variables which deviate significantly from fundamental forces, the likelihood of observing multiple equilibria in the map from those variables to outcomes will be lower. For example, one theme of this survey is that institutions can be an important source of multiplicity, either directly or indirectly through their interactions with the market. If institutions are endogenous, and if traps in institutions drive the disparities in cross-country incomes, then conditioning on institutions may give spurious convergence results entirely disconnected from long run outcomes generated by the system.

### *4.3 Model calibration*

One of the advantages of the methodology proposed by Bloom, Canning and Sevilla is that estimation and testing can proceed without fully specifying the underlying model. The exacting task of determining the relevant set of endogenous variables and the laws by which they interact is thereby circumvented. But there are two sides to this coin. While the results of the test suggest that poverty traps matter, they give no indication as to their source, or to the appropriate framework for formulating them as models.

Graham and Temple (2004) take the opposite approach. They give the results of a numerical experiment starting from a specific poverty trap model, somewhat akin to the inertial self-reinforcement model of Section 3.4. The question they ask is whether or not the model in question has the potential to explain observed cross-country variation in per capita income for a reasonable set of parameters. We briefly outline their main findings, as well as their technique for calibration, which is of independent interest.

As in Section 3.4, there is both a traditional agricultural sector and a modern sector with increasing social returns due to technical externalities. The agricultural sector has a decreasing returns technology

$$Y_a = A_a L_a^{\gamma}, \quad \gamma \in (0,1), \tag{11}$$

where $Y_a$ is output, $A_a$ is a productivity parameter and $L_a$ is labor employed in the agricultural sector. The $j$-th firm in the modern sector has technology

$$Y_{m,j} = A_m L_{m,j} L_m^\lambda, \quad \lambda > 0, \tag{12}$$

where $Y_{m,j}$ is output of firm $j$, $A_m$ is productivity, $L_{m,j}$ is labor employed by firm $j$, and $L_m$ is total employment in the modern sector. The firm ignores the effect of its hiring decisions on $L_m$, thus setting the stage for multiplicity. We set $L_a + L_m = L$, a fixed constant, and, as usual, $\alpha := L_m/L$.

The relative price of the two goods is fixed in world markets and normalized to one by appropriate choice of units. Wages are determined by marginal cost pricing: $w_a = \gamma A_a L_a^{\gamma-1}$ and $w_m = A_m L_m^\lambda$. Setting these factor payments equal gives the set of equilibrium modern sector shares $\alpha$ as solutions to the equation

$$(1-\alpha)^{1-\gamma}\alpha^\lambda = \frac{A_a \gamma L^{\gamma-1-\lambda}}{A_m}. \tag{13}$$

Regarding calibration, $\gamma$ is a factor share, and the increasing returns parameter $\lambda$ has been calculated in several econometric studies.[51] Relative productivity is potentially more problematic. However, it turns out that (13) has precisely two solutions for reasonable parametric values. Since both solutions $\alpha_1$ and $\alpha_2$ satisfy (13) we have

$$(1-\alpha_1)^{1-\gamma}\alpha_1^\lambda - (1-\alpha_2)^{1-\gamma}\alpha_2^\lambda = 0. \tag{14}$$

In which case, assuming that current observations are in equilibrium, one can take the observed share as $\alpha_1$, calculate $\alpha_2$ as the other solution to (14), and set the poverty trap equilibrium equal to $\alpha_1^* := \min\{\alpha_1, \alpha_2\}$. The high productivity equilibrium is $\alpha_2^* := \max\{\alpha_1, \alpha_2\}$. Figure 17 illustrates this procedure for $\alpha_1 = 0.1$, $\gamma = 0.7$ and $\lambda = 0.3$.

When $\alpha_1^*$, $\alpha_2^*$, $\gamma$ and $\lambda$ are known, a little algebra shows that the ratio of output in the high equilibrium to output in the low equilibrium can also be computed. In this way it is possible to evaluate the relative impact of the poverty trap on individual countries and the cross-country income distribution.

Using this strategy and a more elaborate model (including both capital and land), Graham and Temple's main findings are as follows. First, for reasonable parameter values some 1/4 of the 127 countries in their 1988 data set are in the poverty trap $\alpha_1^*$. Second, after calculating the variance of log income across countries when all are in their high output equilibrium and comparing it to

[51] See, for example, Cabarello and Lyons (1992).

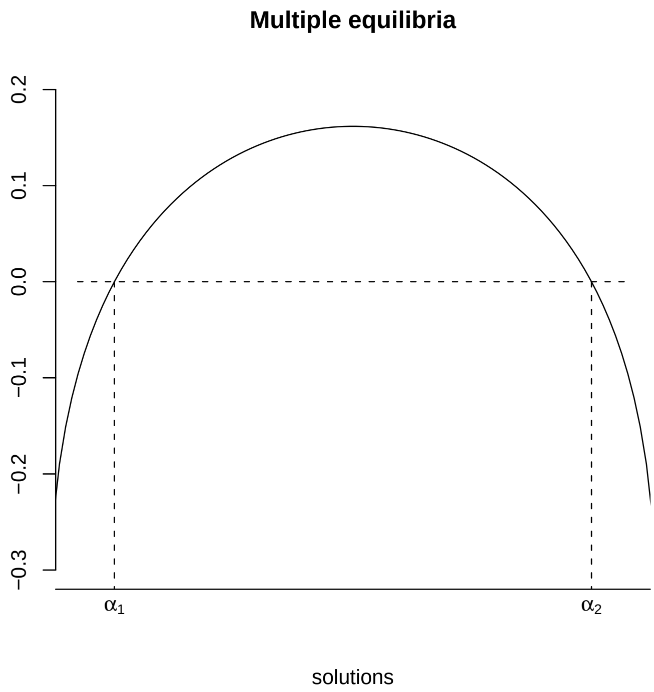


Fig. 17.

the actual variance of log income, they find that the poverty trap model can account for some 2/5 to one half of all observed variation in incomes.

Overall, their study suggests that the model can explain some properties of the data, such as the difference between poor, agrarian economies and low to middle income countries. On the other hand, it cannot account for the huge differences between the very poorest and the rich industrialized countries. In the model, the largest ratios of low to high equilibrium production are in the region of two to three. As we saw in Section 2.1, however, actual per capita output ratios between rich and poor countries are much larger.

### *4.4 Microeconomic data*

There has also been research in recent years on poverty traps that occur at the individual or group level. For example, Jalan and Ravallion (2002) fit a microeconomic model of consumption growth with localized spillovers to farm-household panel data from rural China. Their results are consistent with empirical significance of geographical poverty traps. Other authors have studied particular trap mechanisms. For example, Bandiera and Rasul (2003) and Conley and Udry (2003) consider the effects of positive network externalities on technology adoption in Mozambique and Ghana respectively. Barrett, Bezuneh and Aboud (2001) consider the dynamic impact of credit constraints on the poor in Côte d'Ivoire and Kenya. Morduch (1990) studies the effect of risk on income in India, as does Dercon (1998) for Tanzania.

## 5 Nonconvexities, Complementarities and Imperfect Competition

Increasing returns production under imperfect competition is a natural framework to think about multiple equilibria. Imperfect competition leads directly to externalities transmitted through the price system, because monopolists themselves, rather than Walrasian auctioneers, set prices, and presumably they do so with their own profit in mind. At the same time, their pricing and production decisions impinge on other agents. These general equilibrium effects can be a source of multiplicity.

Section 5.1 illustrates this idea using the big push model of Murphy, Shleifer and Vishny (1989); a model which formalizes an earlier discussion in Rosenstein-Rodan (1943). Rosenstein-Rodan argued that modern industrial technology is freely available to poor countries, but has not been adopted because the domestic market is too small to justify the fixed costs it requires. If all sectors industrialize simultaneously, however, the market may potentially be expanded to the extent that investment in modern technology is profitable.

Thus the big push model of Section 5.1 helps to clarify the potential challenges posed by coordination for the industrialization process. We shall see that the major coordination problem facing monopolists cannot be resolved by the given market structure. In this situation, the ability of a society to successfully coordinate entrepreneurial activity—and thereby realize the social benefits available in modern production technologies—will depend in general on such structures as its institutions, political organizations, the legal framework, and social and business conventions.

In countries such as South Korea, the state has been very active in attempting to overcome coordination problems associated with industrialization. In Western Europe, the state was typically much less active, and the role of the private sector was correspondingly larger. For example, Da Rin and Hellmann (2002) have recently emphasized the important role played by banks in coordinating industrialization. Section 5.2 reviews their model.

A theme of this survey is traps that prevent economies as a whole from adopting modern production technologies. One aspect of this transformation to modernity is the need for human capital. If investment in human capital has a high economic payoff then a skilled work-force should spontaneously arise. Put differently, if the poor are found to invest little in schooling or training then this suggests to us that returns to these investments are relatively low. Section 5.3 reviews Kremer's (1993) matching model, where low investment in schooling sustains itself in a self-reinforcing trap.

Finally, Section 5.4 gives references to notable omissions on the topic of increasing returns.

### *5.1 Increasing returns and imperfect competition*

Murphy, Shleifer and Vishny's (1989) formalization of Rosenstein-Rodan's (1943) big push is something of a watershed in development economics. Their model turns on demand spillovers which create complementarities to investment. They point out that for the economy to generate multiple equilibria, it must be the case that investment simultaneously (i) increases the size of other firms' markets, or otherwise improves the profitability of investment; and (ii) has negative net present value. This means that profits alone cannot be the direct source of the market size effects; otherwise (i) and (ii) would be contradictory.

In the first model they present, higher wages in the modern sector are the channel through which demand spillovers increase market size. Although investment is not individually profitable, it raises labor income, which in turn raises the demand for other products. If the spillovers are large enough, multiple equilibria will occur. In their second model, investment in the modern technology changes the composition of aggregate demand across time. In the first period, the single monopolistic firm in each sector decides whether to invest or not. Doing so incurs a fixed cost $F$ in the first period, and yields output $\omega L$ in the second, where $\omega > 1$ is a parameter and $L$ is labor input. The cost in the second period is just $L$, as wages are the numeraire. If, on the other hand, the monopolist chooses not to invest, production in that sector will take place in a "competitive fringe" of atomistic firms using constant returns to scale technology. For these firms, one unit of labor input yields one unit of output. The price for each unit so produced is unity.

All wages and profits accrue to a representative consumer, who supplies $L$ units of labor in both periods, and maximizes the undiscounted utility of his consumption, that is,

$$\max\left\{\int_0^1 \ln c_1(\alpha)d\alpha + \int_0^1 \ln c_2(\alpha)d\alpha \;\middle|\; c_t\colon [0,1] \ni \alpha \mapsto c_t(\alpha) \in [0,\infty)\right\}$$

subject to the constraints $\int_0^1 c_1 p_1 \leq y_1$ and $\int_0^1 p_2 c_2 \leq y_2$. Here $\alpha \in [0,1]$ indexes the sector, $c_t(\alpha)$ and $p_t(\alpha)$ are consumption and price of good $\alpha$ at time $t$ respectively, and $y_t$ is income (wages plus profits) at time $t$.[52]

In the first period only the competitive fringe produces, and $p_1(\alpha) = 1$ for all $\alpha$. In the second, monopolists face unit elastic demand curves $c_2(\alpha) = y_2/p_2(\alpha)$. Given these curves and the constraints imposed by the competitive fringe, monopolists set $p_2(\alpha) = 1$ for all $\alpha$. Their profits are $\pi = ay_2 - F$, where $a := 1 - 1/\omega$ is the mark-up.

Consider profitability when all entrepreneurs corresponding to sectors $[0,\alpha]$ decide to invest. (The number $\alpha$ can also be thought of as the fraction of the total number of monopolists who invest.) It turns out that for some parameter values both $\alpha = 0$ and $\alpha = 1$ are equilibria. To see this, consider first the case $\alpha = 0$, so that $y_1 = y_2 = L$. It is not profitable for a firm acting alone to invest if $\pi = aL - F \leq 0$. On the other hand, if $\alpha = 1$, then $y_1 = L - F$ and $y_2 = \omega L$, so monopolists make positive profits when $a\omega L - F \geq 0$. Multiple equilibria exist if these inequalities hold simultaneously. In Figure 18 multiple equilibria obtain for all $L \in [L_1, L_2]$.

As was mentioned in the introduction, coordination problems and other mechanisms that reinforce the status quo can interact with each other and magnify their individual impact. Murphy, Shleifer and Vishny (1989) provide a simple example of this in the context of the model outlined above. They point out that the coordination problem for the monopolists is compounded if industrialization requires widespread development of infrastructure and intermediate inputs, such as railways, road networks, port facilities and electricity grids. All of these projects will themselves need to be coordinated with industrialization.

For example, suppose that $n$ infrastructure projects must be undertaken in the first period to permit industrialization in the second. Each project has a fixed cost $R_n$, and operates in the second period at zero marginal cost. Leaving

[52] To simplify the exposition we assume that consumers can neither save nor dissave from current income. For the moment we also abstract from the existence of a financial sector. Firms which invest simply pay all wages in the second period at a zero rate of interest. See the original for a more explicitly general equilibrium formulation.

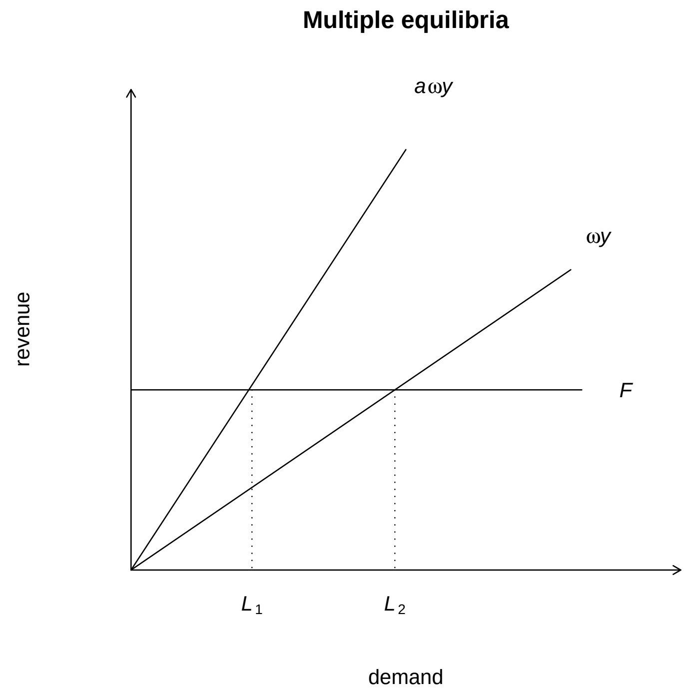


Fig. 18.

aside the issue of how the spoils of industrialization will be divided among the owners of the projects and the continuum of monopolists, it is clear that industrialization has the potential to be profitable for all only when $a\omega L - F$, the profits of the monopolists when $\alpha = 1$, exceed total infrastructure costs $\sum_{i=1}^{n} R_i$.

If the condition $aL - F \leq 0$ continues to hold, however, individual monopolists investing alone will be certain to lose money. Realizing this, investors in infrastructure face extrinsic uncertainty as to whether or not industrialization will actually take place. Given their subjective evaluation, they may choose not to start their infrastructure projects. In turn, the monopolists are aware that investors in infrastructure face uncertainty, and may themselves refrain from starting projects. This makes monopolists even more uncertain as to whether or not the conditions for successful industrialization will eventuate. The fixed point of this infinite regression of beliefs may well be inaction. In either case, the addition of more actors adds to the difficulty of achieving coordination.

As Da Rin and Hellmann (2002) have recently emphasized, one candidate within the private sector for successfully coordinating a big push type industrialization is the banks. Banks are the source of entrepreneurs' funds, and shape the terms and conditions under which capital may be raised. In addition, banks interact directly with many entrepreneurs. Finally, banks can potentially profit from coordinating industrialization if their market power is large.

Da Rin and Hellmann find that the structure and legal framework of the banking sector are important determinants of its ability to coordinate successful industrialization. To illustrate their ideas, consider again the big push model of Section 5.1. In order to make matters a little easier, let us simply define the second period return of monopolists (entrepreneurs) to be $f(\alpha)$, where $\alpha$ is the fraction of entrepreneurs who decide to set up firms and the function $f \colon [0,1] \to \mathbb{R}$ is strictly increasing. As before, there is a fixed cost $F$ to be paid in the first period, which we set equal to 1. The future is not discounted.

It is convenient to think of the number of entrepreneurs as some large but finite number $N$.[53] In addition to these $N$ entrepreneurs, there is now a financial sector, members of whom are referred to as either "banks" or "investors." There are $B \in \mathbb{N}$ banks, the first $B-1$ of which have an intermediation cost of $r$ per unit of investment. The last bank has an intermediation cost of zero, but can lend to only $\ell \leq N$ firms. The number $\ell$ can be thought of as a measure of the last bank's market power.

The equilibrium lending rate at which firms borrow in the first period is determined by the interaction of the monopolists and the banks. In the first stage of the game, each bank $b$ offers a schedule of interest rates to the $N$ firms. This strategy will be written as $\sigma_b := \{i_n^b : 1 \leq n \leq N\}$. The collection of these strategies across banks will be written as $\sigma := \{\sigma_b : 1 \leq b \leq B\}$. Let $\Sigma$ be the set of all such $\sigma$.

In the second stage, each entrepreneur either rejects all offers and does not set up the firm, or selects the minimum interest rate, pays the fixed cost and enters the market. In what follows we write $m_n(\sigma)$ to mean $\min_b i_n^b$, the minimum interest rate offered to firm $n$ in $\sigma$. If a fraction $\alpha$ accepts contracts then firm

[53] In particular, entrepreneurs do not take into account their influence on $\alpha$ when evaluating whether to set up firms or not.

$n$ makes profits

$$\pi(\alpha, m_n(\sigma)) = f(\alpha) - (1 + m_n(\sigma)). \tag{15}$$

For bank $b < B$, profits are given by

$$\Pi_b(\sigma_b) = \sum_{n=1}^{N} (i_n^b - r)\mathbb{1}\{\text{firm } n \text{ accepts}\}, \tag{16}$$

where here and elsewhere $\mathbb{1}\{Q\}$ is equal to one when the statement $Q$ is true and zero otherwise. For $b = B$, profits are

$$\Pi_b(\sigma_b) = \sum_{n=1}^{N} i_n^b \mathbb{1}\{\text{firm } n \text{ accepts}\}. \tag{17}$$

In equilibrium, banks never offer interest rates strictly greater than $r$, because should they do so other banks will always undercut them. As a result, we can and do assume in all of what follows that $m_n(\sigma) \leq r$ for all $n$. Also, to make matters interesting, we assume that $f(0) < 1 + r < f(1)$, or, equivalently,

$$\pi(0, r) < 0 < \pi(1, r). \tag{18}$$

Firms' actions will depend on their beliefs—in particular, on what fraction $\alpha$ of the $N$ firms they believe will enter. Clearly beliefs will be contingent on the set of contracts offered by banks. Thus a belief for firm $n$ is a map $\alpha_n^e$ from $\Sigma$ into $[0, 1]$. Given this belief, firm $n$ enters if and only if

$$\pi(\alpha_n^e(\sigma), m_n(\sigma)) \geq 0. \tag{19}$$

Given $\sigma$, the set of self-supporting equilibria for the second stage subgame is

$$\Omega(\sigma) := \left\{ \alpha \in [0, 1] \ : \ \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\{\pi(\alpha, m_n(\sigma)) \geq 0\} = \alpha \right\}. \tag{20}$$

In other words, $\alpha \in \Omega(\sigma)$ if, given the set of offers $\sigma$ and the belief on the part of all firms that the fraction of firms entering will be $\alpha$, exactly $\alpha \times 100\%$ of firms find it optimal to enter.

Beliefs are required to be consistent in the sense that $\alpha_n^e(\sigma) \in \Omega(\sigma)$ for all $\sigma$ and all $n$. Beliefs are called *optimistic* if $\alpha_n^e = \alpha^{\text{opt}}$ for all $n$, where $\alpha^{\text{opt}}(\sigma) := \max \Omega(\sigma)$ for all $\sigma \in \Sigma$. In other words, all agents believe that as many firms will enter as are consistent with offer $\sigma$, and this is true for every $\sigma \in \Sigma$. Beliefs are defined to be *pessimistic* if the opposite is true; that is, if $\alpha_n^e = \alpha^{\text{pes}}$ for all $n$, where $\alpha^{\text{pes}}(\sigma) := \min \Omega(\sigma)$ for all $\sigma \in \Sigma$.

Da Rin and Hellmann first observe that if $\ell = 0$, then the outcome of the game will be determined by beliefs. In particular, if beliefs are pessimistic, then the low equilibrium $\alpha = 0$ will obtain. If beliefs are optimistic, then the high equilibrium $\alpha = 1$ will obtain. The interpretation is that when $\ell = 0$, so that the market for financial services is entirely competitive (in the sense of Bertrand competition with identical unit costs described above), the existence of the financial sector will not alter the primary role of beliefs in determining whether industrialization will take place.

Let us verify this observation in the case of pessimistic beliefs. To do so, it is sufficient to show that if $\sigma \in \Sigma$ is optimal, then $0 \in \Omega(\sigma)$. The reason is that if $0 \in \Omega(\sigma)$, then by (20) we have $\pi(0, m_n(\sigma)) < 0$ for all $n$. Also, beliefs are pessimistic, so $\alpha_n^e(\sigma) = \min \Omega(\sigma) = 0$. In this case no firms enter by (19).

To see that $0 \in \Omega(\sigma)$ for all optimal $\sigma$, suppose to the contrary that $\sigma \in \Sigma$ is optimal, but $0 \notin \Omega(\sigma)$. Then $\pi(0, m_k(\sigma)) \geq 0$ for some $k$, in which case (18) implies that $m_k(\sigma) < r$. Because firms only accept contracts at rates less than $r$ (that is, $m_n(\sigma) \leq r$ for all $n$), it follows from (16) that the bank which lent to $k$ looses money, and $\sigma$ is not optimal. The intuition is that no bank has market power, and cannot recoup losses sustained when encouraging firms to enter by offering low interest rates.

More interesting is the case where the last bank $B$ has market power. With sufficient market power, $B$ will induce industrialization (the high equilibrium where $\alpha = 1$) even when beliefs are pessimistic:

**Proposition 5.1 (Da Rin and Hellmann)** *Suppose beliefs are pessimistic. In this case, there exists an $\bar{\alpha} \in [0, 1]$ depending on $r$ and $f$ such that industrialization will occur whenever $\ell$, the market power of $B$, satisfies $\ell/N \geq \bar{\alpha}$.*

The result shows that rather than relying on spontaneous coordination of beliefs, financial intermediaries may instead be the source of coordination. The key intuition is that a financial intermediary may have a profit motive for inducing industrialization. But to achieve this, two things are necessary: size and market power. Size (as captured by $\ell$) is necessary to induce a critical mass of entrepreneurs to invest. Market power (as captured by the cost advantage $r$) is necessary to recoup the costs of mobilizing that critical mass. We sketch Da Rin and Hellmann's proof in the appendix.

Until now we have considered only the possibility that the banks offer pure debt contracts. Da Rin and Hellmann also study the case where the banks may hold equity as well (i.e., universal banking). They show that in this case the threshold level at which the lead bank $B$ has sufficient market power to mobilize the critical mass is lower. Industrialization is unambiguously more likely to occur. The reason is that equity permits $B$ to partake in the ex post profits of the critical mass, who benefit from low interest rates on one hand and complete entry ($\alpha = 1$) on the other. With a lower cost of mobilizing firms, $B$ requires less market power to recoup these losses. In Da Rin and Hellmann's words,

> Our model provides a rationale for why a bank may want to hold equity that has nothing to do with the standard reasons of providing incentives for monitoring. Instead, equity allows a bank to participate in the gains that it creates when inducing a higher equilibrium.

In summary, the theory suggests that large universal banks with a high degree of market power can play a central role in the process of industrialization. This theory is consistent with the evidence from countries such as Belgium, Germany and Italy, where a few oligopolist banks with strong market positions played a pivotal role. Some were pioneers of universal banking, and many directly coordinated activity across sectors by participation in management. The theory may also explain why other countries, such as Russia, failed to achieve significant industrialization in the 19th Century. There banks were small and dispersed, their market power severely restricted by the state.

### *5.3 Matching*

The next model we consider is due to Kremer (1993), and has the following features. A production process consists of $n$ distinct tasks, organized within a firm. For our purposes $n$ can be regarded as exogenous. The tasks are undertaken by $n$ different workers, all of whom have their own given skill level $h_i \in [0, 1]$. Here the skill level will be thought of as the probability that the worker performs his or her task successfully. We imagine that if one worker fails in their task the entire process is ruined and output is zero. If all are successful, the outcome of the process is $n$ units of the product.[54] That is,

$$y = n \prod_{i=1}^{n} \mathbb{1}\{\text{worker } i \text{ successful}\}, \quad \mathbb{P}\{\text{worker } i \text{ successful}\} = h_i, \tag{21}$$

[54] Assuming one unit might seem more natural than $n$, but the latter turns out to be more convenient.

where as before $\mathbb{1}\{Q\} = 1$ if the statement $Q$ is true and zero otherwise. All of the success probabilities are independent, so that $\mathbb{E}(y) = n \prod_i h_i$.

Consider an economy with a unit mass of workers. The distribution of skills across workers is endogenous, and will be discussed at length below. Kremer's first point is that in equilibrium, firms will match workers of equal skill together to perform the process. The intuition is that (i) firms will not wish to pair a work-force of otherwise skilled employees with one relatively unskilled worker, who may ruin the whole process; and (ii) firms with a skilled work-force will be able to bid more for skilled workers, because the marginal value of increasing the last worker's skill is increasing in the skill of the other workers. Thus, for each firm,

$$\mathbb{E}(y) = nh^n, \quad h \text{ the firm's common level of worker skill.} \tag{22}$$

The first thing to notice about this technology is that the expected marginal return to skill is increasing. As a result, small differences in skill can have relatively large effects on output. This may go some way to explaining the extraordinarily large wage differentials between countries. Moreover, for economies with such technology, positive feedback dynamics of the kind considered in Section 3.3 may result, even if the technology for creating human capital is concave.

Another channel for positive feedbacks occurs when matching is imperfect, perhaps because it is costly or the population is finite. Exact matches may not be possible. In that case, there are potentially returns to agglomeration: Skilled people clustering together will decrease the cost of matching, and increase the likelihood of good matches. Also, an initial distribution of skills will tend to persist, because workers will choose skills so as to be where the distribution is thickest. This maximizes their chances of finding good matches. But this is self-reinforcing: Their choices perpetuate the current shape of the distribution.

There is yet another channel that Kremer suggests may lead to multiple equilibrium distributions of skill. This is the situation where skill levels are imperfectly *observed*. We present a simple (and rather extremist) version of Kremer's model. In the first period, workers decide whether to undertake "schooling" or not. This education involves a common cost $c \in (0, 1)$. In the second, firms match workers, produce, and pay out wages. Both goods and labor markets are competitive, and total wages exhaust revenue. Specifically, it is assumed that each worker's wage $w$ is $1/n$-th of firm's output.

Not all of those who undertake schooling become skilled. We assume that the educated receive a skill level $h = 1$ with probability $p > 1/2$ and $h = 0$ with probability $1 - p$. Those who do not undertake schooling have the skill level $h = 0$. Further, $h$ is not observable, even for workers. Instead, all workers take a test, which indicates their true skill with probability $p$ and the reverse with probability $1 - p$.[55] That is,

$$t := \text{test score} = \begin{cases} h & \text{with probability } p; \\ 1 - h & \text{with probability } 1 - p. \end{cases} \tag{23}$$

Firms then match workers according to the test score $t$ rather than $h$.

Let $\alpha \in [0, 1]$ denote the fraction of workers who choose to undertake schooling. We will show that for certain values of the parameters $p$ and $c$, both $\alpha = 0$ and $\alpha = 1$ are equilibria. In doing so, we assume that $p$ is known to all. Also, workers and firms are risk neutral.

Consider first the case where $\alpha = 0$. If the worker undertakes schooling, then, regardless of his skill and test score, his expected wage is $1/n$-th of $n \prod_i h_i$, where his co-workers are drawn from a pool in which the skilled workers have measure zero. That is, $\mathbb{P}\{h_i = 0\} = 1$. It follows that expected output and wage are zero. Since $c > 0$, it is optimal to avoid schooling.[56]

Now consider the agent's problem when $\alpha = 1$. In the second period, the agent will be matched with other workers having the same test score. In either case, computing expected wages is a signal extraction problem. First, using the fact that agents in the pool of potential co-workers have chosen schooling with probability one, the agent can calculate probable skills of a co-worker chosen at random from the population, given their test score:

$$\mathbb{P}\{h = 1 \,|\, t = 1\} = \frac{\mathbb{P}\{h = 1 \text{ and } t = 1\}}{\mathbb{P}\{t = 1\}} = \frac{p^2}{p^2 + (1 - p)^2} =: \theta_p, \tag{24}$$

and,

$$\mathbb{P}\{h = 1 \,|\, t = 0\} = \frac{\mathbb{P}\{h = 1 \text{ and } t = 0\}}{\mathbb{P}\{t = 0\}} = \frac{p(1 - p)}{p(1 - p) + p(1 - p)} = \frac{1}{2}. \tag{25}$$

The worker can use these probabilities to compute expected output and hence wages given the different outcomes of his own test score. In particular, $\mathbb{E}(w \,|\, t =$

[55] We are using the same $p$ as before just to simplify notation.

[56] On the other hand, if skills are perfectly observable, workers who acquire skills will be matched with $n$ workers from the measure zero set of agents having $h = 1$. In that case $w = 1$. Since $c < 1$ it is optimal to choose schooling, and $\alpha = 0$ is not an equilibrium. The same logic works for any $\alpha < 1$

$1) = \theta_p^n$ and $\mathbb{E}(w \mid t = 0) = (1/2)^n$. It follows that the expected return to schooling for the agent is

$$\begin{aligned}\mathbb{E}(w \mid \text{schooling}) &= \mathbb{E}(w \mid t=0)\mathbb{P}\{t=0\} + \mathbb{E}(w \mid t=1)\mathbb{P}\{t=1\} \\ &= \frac{1}{2^n}(1-p) + \theta_p^n p.\end{aligned}$$

Conversely, $\mathbb{E}(w \mid \text{no schooling}) = \frac{1}{2^n}p + \theta_p^n(1-p)$. Schooling is optimal if

$$\begin{aligned}c < \mathbb{E}(w \mid \text{schooling}) - \mathbb{E}(w \mid \text{no schooling}) \\ = (2p-1)(\theta_p^n - (1/2)^n) := c^*(p).\end{aligned}$$

It is easy to see that $c^*(p) > 0$ whenever $p > 1/2$, which is true by assumption. As a result, schooling will be optimal for some sufficiently small $c$, and $\alpha = 1$ is an equilibrium too.[57]

What are the sources of multiple equilibria in the model? The first is pecuniary externalities in the labor market: When more agents become educated, the probability that the marginal worker can successful match with a skilled co-worker increases. In turn, this increases the returns to education.[58] Second, there is imperfect information: Skilled workers cannot readily match with other skilled workers. Instead, matching is probabilistic, and depends on the overall distribution of skills. Finally, the increasing expected marginal reward for skill inherent in the production function means that the wage spillovers from the decisions of other agents are potentially large.

Another important model of human capital investment with multiple equilibria is Acemoglu (1997). He shows how labor market frictions can induce a situation where technology adoption is restricted by a lack of appropriately skilled workers. Low adoption in turn reduces the expected return to training, further exacerbating the scarcity of workers who are trained. In other words, poor technology adoption and low capital investment are self-reinforcing, because they cause the very shortage of skilled workers necessary to make such investments profitable.

[57] It may seem that if $p = 1$ and observation is perfect, then $\mathbb{E}(w \mid \text{schooling}) - \mathbb{E}(w \mid \text{no schooling})$ should be zero, so that no multiple equilibria are possible. But under this assumption the above derivation of $c^*(p)$ is not valid, because we would be conditioning on sets with probability zero.

[58] In fact the expected wage is increased for all, but those who become skilled benefit more.

Young's (1928) famous paper on increasing returns notes that not only does the degree of specialization depend on the size of the market, but the size of the market also depends on the degree of specialization. In other words, there are efficiency gains from greater division of labor, primarily due to application of machines. Greater specialization increases productivity, which then expands the market, leading back into more specialization, and so on. As a result, there are complementarities in investment. These complementarities can be the source of poverty traps. A detailed discussion of this process is omitted from the present survey, but only because excellent surveys already exist. See in particular Matsuyama (1995) and Matsuyama (1997). Other references include Matsuyama and Ciccone (1996), Rodríguez-Clare (1996) and Rodrik (1996).

Increasing returns are also associated with geographical agglomeration. Starrett (1978) points out that agglomerations *cannot* form as the equilibria of perfectly competitive economies set in a homogeneous space. Thus all agglomerations must be caused either by exogenous geographical features or by some market imperfection. An obvious candidate is increasing returns. (It is difficult to see what geographical features could explain the extent of concentration witnessed in places such as Tokyo or Hong Kong.) This survey does not treat geography and its possible connections with poverty traps in much detail. Interested readers might start with the review of Ottaviano and Thisse (2004).[59]

Another source of complementarities partly related to geography is positive network externalities in technology adoption. These are often thought to arise from social learning: Local experience with a technology allows the cost of adoption to decrease as the number of adopters in some network gets larger. As well as information spillovers, more adopters of a given technology may lead to the growth of local supply networks for intermediate inputs, repairs and servicing, skilled labor and so on. See, for example, Beath, Katsoulacos and Ulph (1995), Bandiera and Rasul (2003), Conley and Udry (2003), and Baker (2004).

Finally, an area that we have not treated substantially in this survey is *optimal* growth under nonconvexities, as opposed to the fixed savings rate model

[59] See also Limao and Venables (2001) or Redding and Venables (2004) for the empirics of geography and international income variation.

considered in Section 3.3. In other words, how do economies evolve when (i) agents choose investment optimally by dynamic programming, given a set of intertemporal preferences; and (ii) the aggregate production function is nonconvex?

There are two main cases. One is that increasing returns are taken to be external, perhaps as a feedback from aggregate capital stock to the productivity residual, and agents perceive the aggregate production function to be *convex*. In this case there is a subtle issue: In order to optimize, agents must have a belief about how the productivity residual evolves. This may or may not coincide with its actual evolution as a result of their choices. An equilibrium transition rule is a specification of savings and investment behavior such that (a) agents choose this rule given their beliefs; and (b) those choices cause aggregate outcomes to meet their expectations. Existence of such an equilibrium is far from assured. See Mirman, Morand and Reffett (2004) and references therein. Dynamics are still actively being investigated.

The second case is where increasing returns are internal, and agents perceive aggregate production possibilities exactly as they are. These models generate similar poverty traps as were found for fixed savings rates in Section 3.3. The literature is large. An early investigation is Skiba (1978). See also Dechert and Nishimura (1983), who consider a per capita production function $k \mapsto f(k)$ which is convex over a lower region of the state space (capital per worker), and concave over the remainder; and Amir, Mirman and Perkins (1991), who study the same problem using lattice programming. Majumdar, Mitra and Nyarko (1989) study optimal growth for stochastic nonconvex models, as do Nishimura and Stachurski (2004). Dimaria and Le Van (2002) analyze the dynamics of deterministic models with R&D and corruption.[60]

## 6 Credit Markets, Insurance and Risk

In terms of informational requirements necessary for efficient free market operation and low transaction costs, one of the most problematic of all markets is the intertemporal trade in funds. Here information is usually asymmetric, and lenders face the risk of both voluntary and involuntary default (Kehoe and Levine 1993). Voluntary default is strategic default by borrowers who judge

[60] One should be cautious about interpreting these nonconvex models as aggregative studies of development. The Second Fundamental Welfare Theorem does not apply, so decentralization is problematic.

the expected rewards of repayment to be lower than those of not repaying the loan. Involuntary default occurs when ex post returns are insufficient to cover total loans.

Facing these risks, a standard response of lenders is to make use of collateral (Kiyotaki and Moore 1997). But the poor lack collateral almost by definition; as a result they are credit constrained. Credit constraints in turn restrict participation by the poor in activities with substantial set up costs, as well as those needing large amounts of working capital. For the poor, then, the range of feasible income-generating activities is reduced. Thus, the vicious circle of poverty: Income determines wealth and low wealth restricts collateral. This trap is discussed in Section 6.1.[61]

The market for insurance is similar to the market for credit, in that information is asymmetric and transaction costs are high. This can lead to poverty traps in several ways. In Section 6.2, we study a model where poor agents, lacking access to insurance or credit, choose low risk strategies at the cost of low mean income. These choices reinforce their poverty.

In Section 6.3 we review Matsuyama's (2004) world economy model, where all countries must compete for funds in a global financial market. On one hand, diminishing returns imply that rewards to investment in the poor countries are large. High returns attract funds and investment, and high investment provides a force for convergence. On the other hand, credit markets are imperfect, and rich countries have more collateral. This puts them in a strong position vis-a-vis the poor when competing for capital. The inability of the poor to guarantee returns with collateral is a force for divergence.

### *6.1 Credit markets and human capital*

Consider an economy producing only one good and facing a risk free world interest rate of zero. Agents live for one period. Each has one and only one child. From their parent, the child receives a bequest $x$. At the beginning of life, each agent chooses between two occupations. The first is to work using a constant returns technology $Y = \bar{w}L$, where $Y$ is output, $L$ is total labor input in this sector, and $\bar{w}$ is a productivity parameter. The agent supplies all

[61] See also Tsiddon (1992) for a poverty trap model connected to the market for credit. In his model, asymmetric information leads to a moral hazard problem, which restricts the ability of investors to raise money. The market solution involves quantity constraints on loans, the severity of which depends on the level of income.

of his or her labor endowment $\ell_t$, and we define $w_t := \bar{w}\ell_t$ as the return to this choice of occupation. We admit the possibility that $\ell_t$ varies stochastically, so $w_t$ may be random.

Alternatively, the agent may set up a project at cost $F$. The gross payoff from the project is equal to $Q_t$. Agents with wealth $x_t < F$ may borrow to cover the costs of the project beyond which they are able to self-finance. They face interest rate $i > 0$, where the excess of the borrowing rate over the risk free rate reflects a credit market imperfection. In this case we have in mind costs imposed on lenders due to the need for supervision and contract enforcement (c.f., e.g., Galor and Zeira 1993, p. 39). These costs are then passed on to the borrower.

The two stochastic productivity parameters $w_t$ and $Q_t$ are draws from joint distribution $\varphi$. We assume that $\mathbb{E}\ell_t = 1$, and that $\mathbb{E}w_t = \bar{w} < \mathbb{E}Q_t - F$. Thus, the net return to setting up the project is higher on average than the wage. However, the agent may still choose to work at wage rate $w_t$ if his or her income is relatively low. The reason is that for the poor setting up a project requires finance at the borrowing rate $i > 0$, which may offset the differential return between the two occupations.

Consider the employment decisions and wealth dynamics for each dynasty. Omitting time subscripts, an agent with bequest $x$ has

$$y := \text{lifetime income } = \begin{cases} x + w & \text{if do not set up project;} \\ (x - F)(1 + i) + Q & \text{if set up project, } x < F; \\ (x - F) + Q & \text{if set up project, } x \geq F. \end{cases}$$

Preferences are given by $u(c, b) = (1 - \theta)\ln c + \theta \ln b$, where $\theta \in (0, 1)$ is a parameter, $c$ is consumption and $b$ is bequest to the child. As a result, each agent bequeaths a fraction $\theta$ of $y$; the remainder is consumed. Indirect utility is $v(y) = \gamma + \delta \ln y$, where $\gamma, \delta > 0$ are constants.

To abstract temporarily from the issue of risk aversion let us suppose that each agent can observe his or her idiosyncratic shocks $(w, Q)$ prior to choosing a field of employment. As a result, agents with $x \geq F$ will choose to set up projects iff $Q - F \geq w$. Agents with $x < F$ will choose the same iff $(x - F)(1 + i) + Q \geq x + w$; in other words, iff

$$x \geq \hat{x} := \frac{w - Q + F(1 + i)}{i}.$$

It follows that dynamics for each dynasty's wealth in this economy are given

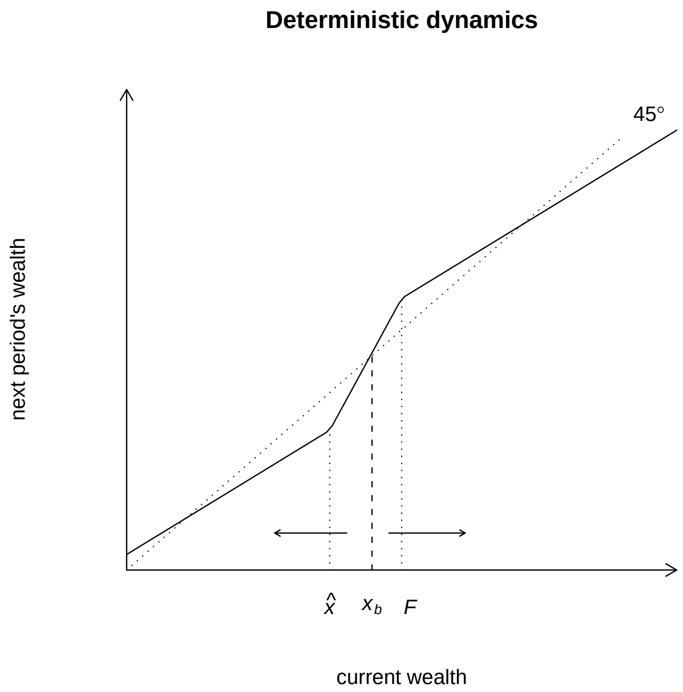


Fig. 19.

by the transition rule

$$x_{t+1} = S_t(x_t); \quad S_t(x_t) = \theta \times \begin{cases} x_t + w_t & \text{if } x_t \leq \hat{x}_t; \\ (x_t - F)(1+i) + Q_t & \text{if } x_t \in (\hat{x}_t, F); \\ x_t - F + Q_t & \text{if } x_t \geq F. \end{cases}$$

Figure 19 illustrates a transition rule $S$ and hence the dynamics of this economy when the two rates of return are constant and equal to their means.[62] For this particular parameterization there are multiple equilibria. Agents with initial wealth less than the critical value $x_b$ will converge to the lower attractor, while those with greater wealth will converge to the high attractor. Given any initial distribution $\psi_0$ of wealth in the economy, the fraction of agents converging to the lower attractor will be $\int_0^{x_b} d\psi_0$. If this fraction is large, average long run income in the economy will be small.

A more realistic picture can be obtained if the productivity parameters are permitted to vary stochastically around their means. This will allow at least some degree of income mobility—perhaps very small—which we tend to observe over time in almost all societies. To this end, suppose that for each agent

[62] The parameters here are set to $\theta = 0.7$, $w = 0.06$, $Q = 1.05$, $i = 2$ and $F = 0.65$.

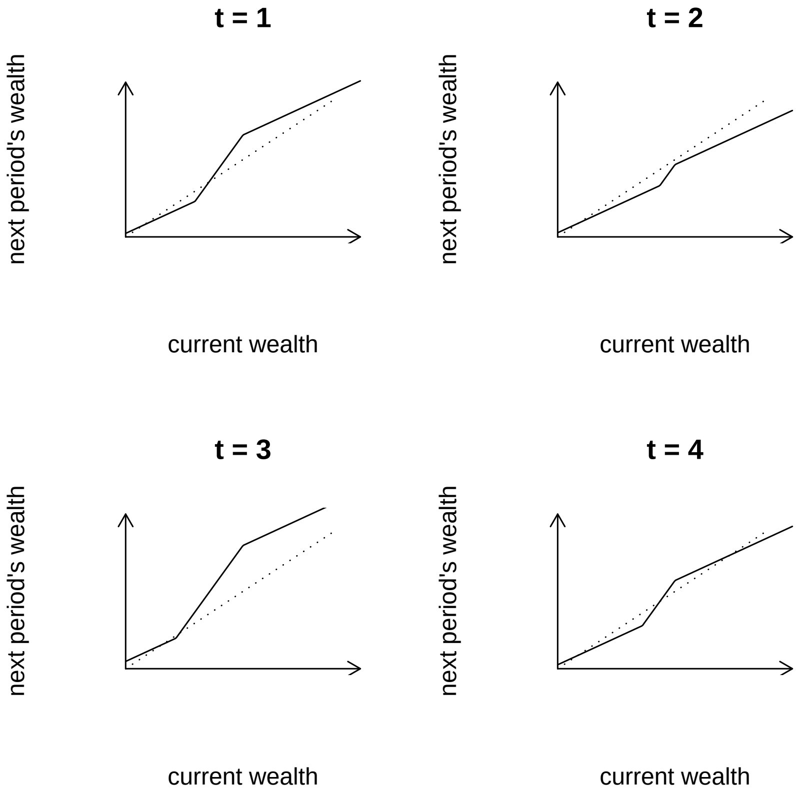


Fig. 20.

and at each point in time the parameters $w_t$ and $Q_t$ are drawn independently across time and agents from a bivariate lognormal distribution. In this case the transition law is itself random, and varies for each agent at each point in time.

Figure 20 shows a simulated sequence of transition rules facing a given agent starting at $t = 1$. At $t = 2$ a negative shock to the project return $Q$ causes the high level attractor to disappear. A *series* of such negative shocks would cause a rich dynasty to loose its wealth. In this case, however, the shocks are iid and such an outcome is unlikely. It turns out that the time 3 shocks are strongly positive.

If the number of agents is large, then the sequence of cross-sectional distributions for wealth over time can be identified with the sequence of marginal probability laws $(\psi_t)_{t \geq 0}$ generated by the Markov process $x_{t+1} = S_t(x_t)$. It is not difficult to prove that this Markov process is ergodic. The intuition and the dynamics are more or less the same as for the nonconvex growth model of

Section 3.3.[63] We postpone further details on dynamics until the next section, which treats another version of the same model.

There are several interpretations of the two sector story with fixed costs described above. One is to take the notion of a project or business literally, in which case $F$ is the cost of set up and working capital which must be paid up before the return is received. Alternatively $F$ might be the cost of schooling, and $Q$ is the payoff to working for skilled individuals.[64] As emphasized by Loury (1981) and others, human capital is particularly problematic for collateral-backed financing, because assets produced by investment in human capital cannot easily be bonded over to cover the risk of default.

Whatever the precise interpretation, the "project" represents an opportunity for the poor to lift themselves out of poverty, while the fixed cost $F$ and the credit market imperfection captured here by $i$ constitute a barrier to taking it. Microeconometric studies suggests that the effects of this phenomenon are substantial. For example, Barrett, Bezuneh and Aboud (2001) analyze the effects of a large devaluation of the local currency that occurred in Côte d'Ivoire in 1994 on rural households. They find that "A macro policy shock like an exchange rate devaluation seems to create real income opportunities in the rural sector. But the chronically poor are structurally impeded from seizing these opportunities due to poor endowments and liquidity constraints that restrict their capacity to overcome the bad starting hand they have been dealt." (Barrett et al. 2001, p. 12)

The same authors also study a local policy shock associated with food aid distribution in Keyna. According to this study, "The wealthy are able to access higher-return niches in the non-farm sector, increasing their wealth and reinforcing their superior access to strategies offering better returns. Those with

---

[63] As we discussed at length in that section, it would be a mistake to claim that this ergodicity result in some way overturns the poverty trap found in the deterministic version.

[64] For these and related stories see Ray (1990), Ray and Streufert (1993), Banerjee and Newman (1993), Galor and Zeira (1993), Ljungqvist (1993), Freeman (1996), Quah (1996), Aghion and Bolton (1997), Piketty (1997), Matsuyama (2000) Mookherjee and Ray (2003) and Banerjee (2003). Yet another possible interpretation of the model is that $F$ is the cost of moving from a rural to an urban area in order to find work. In the presence of imperfect capital markets, such costs—interpreted broadly to include any extra payments incurred when switching to the urban sector—may help to explain the large and growing differentials between urban and rural incomes in some modernizing countries.

weaker endowments ex ante are, by contrast, unable to surmount liquidity barriers to entry into or expansion of skilled non-farm activities and so remain trapped in lower return...livelihood strategies." (Barrett et al. 2001, p. 15).

### *6.2 Risk*

For the poor another possible source of historical self-reinforcement is risk. In the absence of well-functioning insurance and credit markets, the poor find ways to mitigate adverse shocks and to smooth out their consumption. One way to limit exposure is to pass up opportunities which might seem on balance profitable but are thought to be too risky. Another strategy is to diversify activities; and yet another is to keep relatively large amounts of assets in easily disposable form, rather than investing in ventures where mean return is high. All of these responses of the poor to risk have in common the fact that they tend to lower mean income and reinforce long run poverty.

A simple variation of the model from the previous section illustrates these ideas.[65] Let the framework of the problem be the same, but current shocks are no longer assumed to be previsible. In other words, each agent must decide his or her career path before observing the shocks $w_t$ and $Q_t$ which determine individual returns in each sector. Given that preferences are risk averse (indirect utility is $v(y) = \gamma + \delta \ln y$), the agent makes these decisions as a function not only of mean return but of the whole joint distribution. Regarding this distribution, we assume that both shocks are lognormal and may be correlated.

Lenders also cannot observe these variables at the start of time $t$, and hence the borrowing rate $i = i(x)$ reflects the risk of default, which in turn depends on the wealth $x$ of the agent. In particular, default occurs when $Q_t$ is less than the debtor's total obligations $(F - x)(1 + i(x))$. In that case the debtor pays back what he or she is able. Lifetime income is therefore

$$y = \begin{cases} x + w & \text{if do not set up project;} \\ \max\{0, (x - F)(1 + i(x)) + Q\} & \text{if set up project, } x < F; \\ (x - F) + Q & \text{if set up project, } x \geq F. \end{cases}$$

It turns out that in our very simplistic environment agents will never borrow, because when shocks are lognormal agents with $x < F$ who borrow will have $\mathbb{P}\{y = 0\} > 0$, in which case $\mathbb{E}v(y) = -\infty$. (If $x \geq F$ agents may still

[65] What follows is loosely based on Banerjee (2003).

choose to work for a wage, depending on the precise joint distribution.) The result that agents never borrow is clearly unrealistic. For more sophisticated versions of this model with similar dynamics see Banerjee (2003) or Checchi and García-Peñalosa's (2004).

Because agents never borrow, the dynamics for the economy are just

$$x_{t+1} = \theta(x_t + w_t) \cdot \mathbb{1}\{x_t \in D\} + \theta(x_t - F + Q_t) \cdot \mathbb{1}\{x_t \notin D\},$$

where $D := \{x : \mathbb{E}\, v(x + w_t) \geq \mathbb{E}\, v(x - F + Q_t)\}$. (As before, $\mathbb{1}$ is the indicator function.) The stochastic kernel $\Gamma$ for this process can be calculated separately for the two cases $x \in D$ and $x \notin D$ using the same change-of-variable technique employed in Section 3.1. The calculation gives

$$\Gamma(x, x') = \varphi_w\left(\frac{x' - \theta x}{\theta}\right)\frac{1}{\theta} \cdot \mathbb{1}\{x \in D\} + \varphi_Q\left(\frac{x' - \theta(x - F)}{\theta}\right)\frac{1}{\theta} \cdot \mathbb{1}\{x \notin D\},$$

where $\varphi_w$ and $\varphi_Q$ are the marginal densities of $w$ and $Q$ respectively.

A two-dimensional plot of the kernel is given in Figure 21, where the parameters are $F = 1$, $\theta = 0.45$, $\ln w \sim N(0.1, 1)$, and $\ln Q \sim N(1.4, 0.2)$. The dark unbroken line is the $45^o$ line. Lighter areas indicate greater elevation, in this case associated with a collection of probability mass. For the parameters chosen, agents work precisely when $x < F$, and set up projects when $x \geq F$ (so that $D = [0, F]$), despite the fact that mean returns to the project are higher than those of working. The concentration of probability mass along the $45^o$ line in the region $D = [0, F]$ implies that poverty will be strongly self-reinforcing.

Nevertheless, lognormal shocks give poor individuals a non-zero probability of becoming rich at every transition; and the rich can eventually become poor, although it might take a sequence of negative shocks. The rate of mixing depends on the parameters that make up the law of motion and the variance of the shock. Usually some small degree of mixing is a more natural assumption than none. The mixing causes the corresponding Markov chain to be ergodic. This is the case regardless of how small the tails of the shocks are made.[66] For more details on ergodicity see the technical appendix.

To summarize, the poor are not wealthy enough to self-insure, and as a result choose income streams that minimize risk at the expense of mean earnings. The effect is to reinforce poverty. A number of country studies provide evidence

[66] But not necessarily so if the shocks have bounded support.

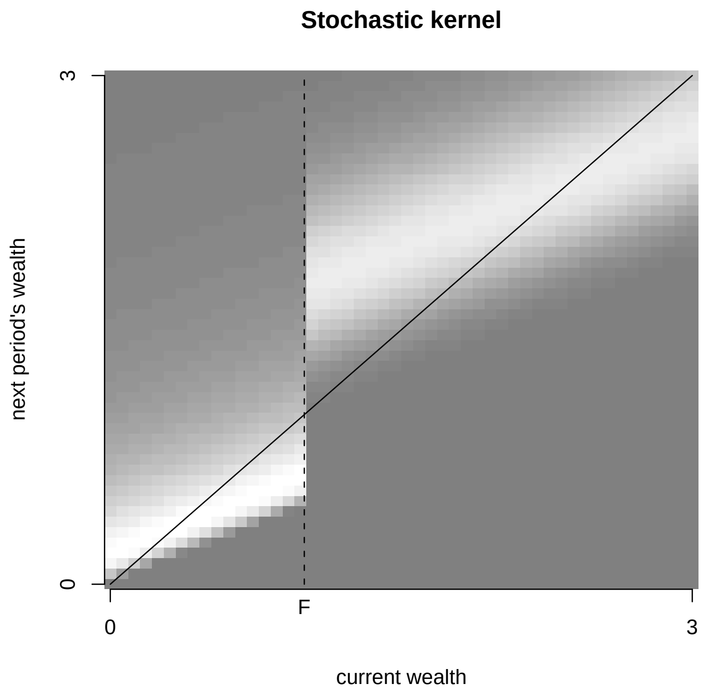


Fig. 21.

of this behavior.[67] Dercon (2003) finds that the effects on mean income are substantial. In a review of the theoretical and empirical literature, he estimates that incomes of the poor could be 25–50% higher on average if they had the same protection against shocks that the rich had as a result of their wealth (Dercon 2003, p. 14).

A more sophisticated model of the relationship between risk and development is Acemoglu and Zilibotti (1997). In their study, indivisibilities in technology imply that diversification possibilities are tied to income. An increase in investment raises output, which then improves the extent of diversification. Since agents are risk averse, greater diversification encourages more investment. In the decentralized outcome investment is too small, because agents do not take into account the effect of their investment on the diversification opportunities of others.

[67] See, for example, Morduch (1990) and Dercon (1998).

Next we consider a world economy model with credit market imperfections due to Matsuyama (2004). For an individual country, the formulation of the problem is as follows. A unit mass of agents live for two periods each, supplying one unit of labor in the first period of life and consuming all their wealth in the second. Per capita output of the consumption good is given by $y_t = f(k_t)\xi_t$, where $f$ is a standard concave production function, $k_t$ is the capital stock and $(\xi_t)_{t\geq 0}$ is a noise process. Once the current shock $\xi_t$ is realized production then takes place. Factor markets are competitive, so that labor and capital receive payments $w_t = [f(k_t) - k_t f'(k_t)]\xi_t =: w(k_t, \xi_t)$ and $\varrho_t = f'(k_t)\xi_t$ respectively.

Current wages $w_t$ are invested by young agents to finance consumption when old. Funds can be invested in a competitive capital market at gross interest rate $R_{t+1}$, or in a project which transforms one unit of the final good into $Q$ units of the capital good at the start of next period. It is assumed that projects are discrete and indivisible: Each agent can run one and only one project.[68] They will need to borrow $1 - w_t$, the excess cost of the project over wages.

Our agents are risk neutral. Time $t$ information is summarized by the information set $\mathscr{F}_t$, and we normalize $\mathbb{E}[\xi_{t+1} \,|\, \mathscr{F}_t] = 1$. In the absence of borrowing constraints, agents choose to start a project if $\mathbb{E}[\varrho_{t+1}Q - R_{t+1}(1 - w_t) \,|\, \mathscr{F}_t] \geq \mathbb{E}[R_{t+1}w_t \,|\, \mathscr{F}_t]$. This is equivalent to

$$\mathbb{E}[R_{t+1} \,|\, \mathscr{F}_t] \leq \mathbb{E}[\varrho_{t+1}Q \,|\, \mathscr{F}_t]. \tag{26}$$

However, it is assumed that borrowers can credibly commit to repay only a fraction $\lambda$ of revenue $\varrho_{t+1}Q$. Thus $\lambda \in [0, 1]$ parameterizes the degree of credit market imperfection faced by borrowers in this economy. As a result, agents can start a project only when $\mathbb{E}[\lambda\varrho_{t+1}Q \,|\, \mathscr{F}_t]$ exceeds $\mathbb{E}[R_{t+1}(1 - w_t) \,|\, \mathscr{F}_t]$, the cost of funds beyond those which the agent can self-finance. In other words, when $w_t = w(k_t, \xi_t) < 1$, we must have

$$\mathbb{E}[R_{t+1} \,|\, \mathscr{F}_t] \leq \Lambda(k_t, \xi_t)\mathbb{E}[\varrho_{t+1}Q \,|\, \mathscr{F}_t], \tag{27}$$

where $\Lambda(k_t, \xi_t) := \lambda/(1 - w_t)$ Given the profitability constraint (26), the borrowing constraint (27) binds only when $\Lambda(k_t, \xi_t) < 1$.[69]

[68] Put differently, we imagine that output is $Q$ units of capital good for all investment levels greater than or equal to one. See the original model for a more general technology.

[69] Of course if $w_t \geq 1$ then all agents can self-finance and the borrowing constraint never binds.

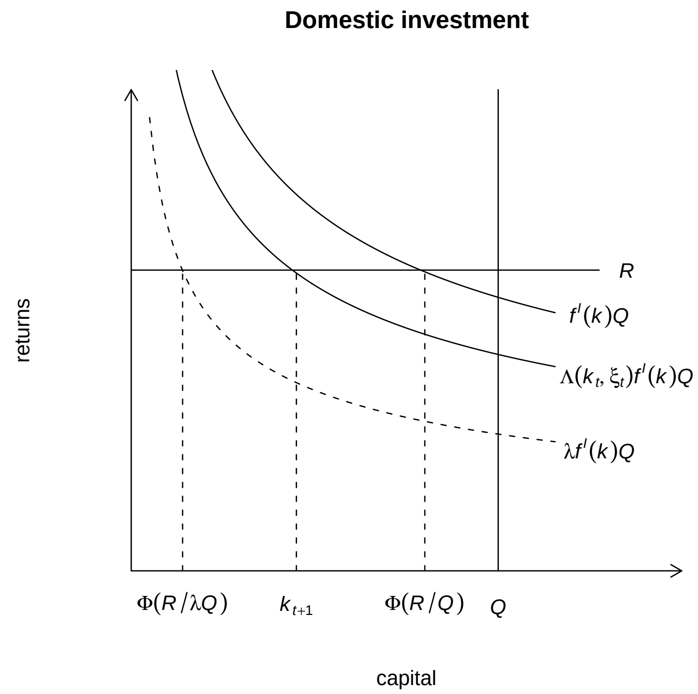


Fig. 22.

In the case of autarky it turns out that adjustment of the domestic interest rate can always equilibrate domestic savings and domestic investment. Since each generation of agents has unit mass, total domestic savings is just $w_t$. If $w_t \geq 1$, then all agents run projects and total output of the capital good is $Q$. If $w_t < 1$, then $w_t$ is equal to the fraction of agents who can start projects. Output of the capital good is $w_t Q$. Assuming that capital depreciates totally in each period, we get $k_{t+1} = \min\{w(k_t, \xi_t)Q, Q\}$. If, for example, technology in the final good sector is Cobb-Douglas, so that $f(k) = Ak^\alpha$, where $\alpha < 1$, then $w(k_t, \xi_t) = (1 - \alpha)Ak^\alpha \xi_t$. For $\xi_t \equiv 1$ there is a unique and globally stable steady state $k^*$.

A more interesting case for us is the small open economy. Here a world interest rate of $R$ is treated as fixed and given. The final good is tradable, so international borrowing and lending are allowed. However, the project must be run in the home country (no foreign direct investment) and factors of production are nontradable.

In the open economy setting there is a perfectly elastic supply of funds at the world interest rate $R$. The effective demand for funds on the part of domestic projects is determined by (26) and (27). The right hand side of (26) is the expected marginal product of capital in this sector, $\mathbb{E}[\varrho_{t+1}Q \mid \mathscr{F}_t]$. Since

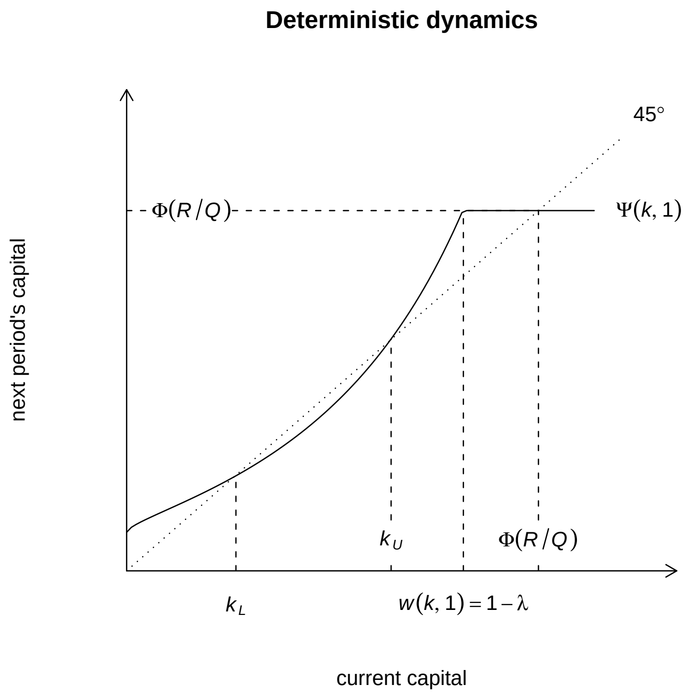


Fig. 23.

$\mathbb{E}[\xi_{t+1}|\mathscr{F}_t] = 1$ we have $\mathbb{E}[\varrho_{t+1}Q \,|\, \mathscr{F}_t] = f'(k_{t+1})Q$. Absent borrowing constraints, investment adjusts to equalize $f'(k_{t+1})Q$ with $R$. Figure 22 shows the intersection of the curve $k \mapsto f'(k)Q$ with the horizontal supply curve $R$ at $\Phi(R/Q)$, where $\Phi$ is the inverse function of $f'$.

As the figure is drawn, however, $\Lambda(k_t, \xi_t) < 1$, perhaps because the capital stock is small, or because of an adverse productivity shock. As a result, the borrowing constraint is binding, and next period's capital stock $k_{t+1}$ is given by the intersection of the effective demand curve $k \mapsto \Lambda(k_t, \xi_t)f'(k)Q$ and the supply curve $R$.

Assuming that $\Phi(R/Q) < Q$ as drawn in the figure, the law of motion for the capital stock is $k_{t+1} = \Psi(k_t, \xi_t)$, where

$$\Psi(k, \xi) := \begin{cases} \Phi[R/\Lambda(k,\xi)Q] & \text{if } w(k,\xi) < 1 - \lambda; \\ \Phi(R/Q) & \text{if } w(k,\xi) \geq 1 - \lambda. \end{cases} \tag{28}$$

For $w(k_t, \xi_t) < 1 - \lambda$ we have $\Lambda(k_t, \xi_t) < 1$ and the borrowing constraint binds. Domestic investment is insufficient to attain the unconstrained equilibrium $\Phi(R/Q)$. In this region of the state space, the law of motion $k \mapsto \Psi(k, \xi)$ is increasing in $k$. Behind this increase lies a credit multiplier effect: Greater domestic investment increases collateral, which alleviates the borrowing con-

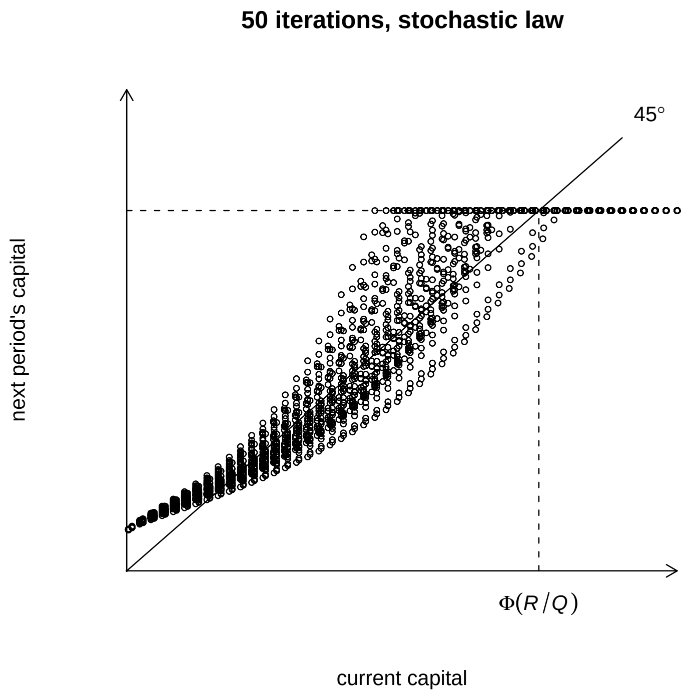


Fig. 24.

straint. This in turn permits more domestic investment, which increases collateral, and so on. Individual agents do not take into account the effect of their actions on the borrowing constraint.

Figure 23 shows the law of motion when $\xi_t \equiv 1$. As drawn, there is a poverty trap at $k_L$ and another attractor at $\Phi(R/Q)$. Countries with $k_t > k_U$ tend to $\Phi(R/Q)$, while those with $k_t < k_U$ tend to $k_L$. Figure 24 shows stochastic dynamics by superimposing the first 50 laws of motion from a simulation on the $45^o$ diagram. The shocks $(\xi_t)_{t \geq 0}$ are independent and identically distributed.[70] Notice that for particularly good shocks the lower attractor $k_L$ disappears, while for particularly bad shocks the higher attractor at $\Phi(R/Q)$ vanishes.

Figure 25 shows a simulated time series for the same parameters as Figure 24 over 500 periods. At around $t = 300$ the economy begins a transition to the higher attractor $\Phi(R/Q)$. Subsequent fluctuations away from this equilibrium are due to shocks so negative that $\Phi(R/Q)$ ceases to be an attractor (see Figure 24).

The story does not end here. What is particularly interesting about Mat-

[70] The production function is $f(k) = k^\alpha$. The shock is lognormal. The parameters are $\alpha = 0.59$, $Q = 2.4$, $\lambda = 0.40$, $R = 1$ and $\ln \xi \sim N(0.01, 0.08)$.

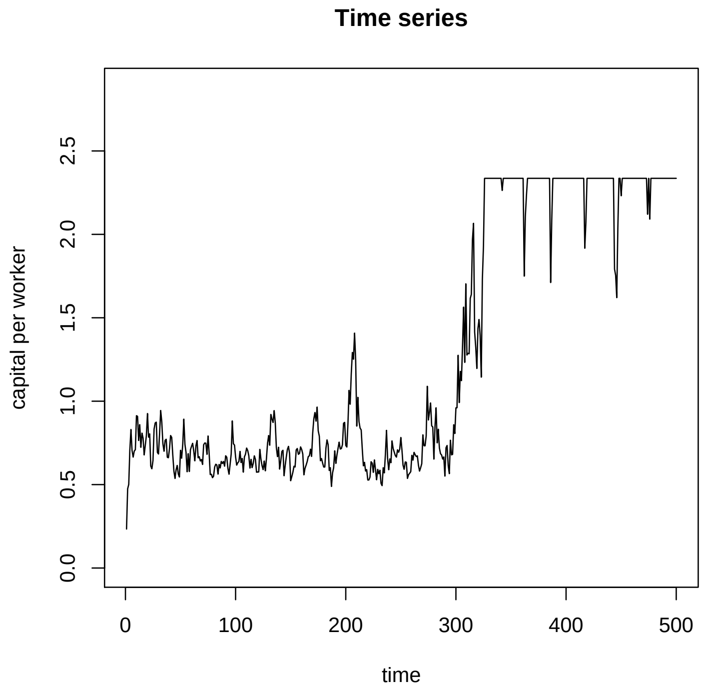


Fig. 25.

suyama's study is his analysis of symmetry-breaking. He shows the following for a large range of parameter values: For a world economy consisting of a continuum of such countries, the deterministic steady state for autarky, which is $k^*$ defined by $k^* = w(k^*, 1)Q$, is precisely $k_U$, the *unstable* steady state for each country under open international financial markets and a world interest rate that has adjusted to equate world savings and investment. Figure 26 illustrates the situation.

Thus, the symmetric steady state after liberalization, where each country has capital stock $k^*$, is unstable and cannot be maintained under any perturbation. The reason is that countries which suffer from bad (resp., good) shocks are weakened (resp., strengthened) in terms of their ability to guarantee returns on loans, and therefore to compete in the world financial market. This leads to a downward (resp. upward) spiral. Under these dynamics the world economy is polarized *endogenously* into rich and poor countries.

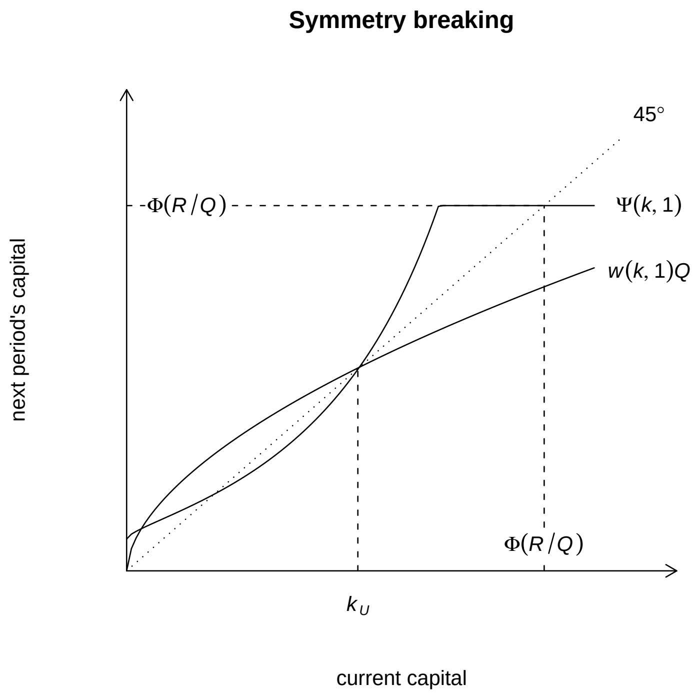


Fig. 26.

## 7 Institutions and Organizations

The fundamental economic problem is scarcity. Since the beginning of life on earth, all organisms have engaged in competition for limited resources. The welfare outcomes of this competition have ranged from efficient allocation to war, genocide and extinction. It is the *rules of the game* which determine the social welfare consequences. More precisely, it is the long run interaction between the rules of the game and the agents who compete.

Institutions—which make up the rules of the game—were at one time thought to have strong efficiency properties in equilibrium. To a large extent, this is no longer the case (for an introduction to the literature, see, for example North 1993, 1995; or Hoff 2000). Institutions can either reinforce market failure or themselves be the source of inefficiency. Moreover, institutions are path dependent, so that bad equilibria forming from historical accident may be locked in, causing poverty to persist.

Among the set of institutions, the state is one of the most important determinants of economic performance; and one of the most common kinds of "government failure" is corruption.[71] In Section 7.1 we review why corruption is thought to be not only bad for growth and development, but also self-reinforcing.

Section 7.2 then looks at the kinship system, a kind of institution that arises spontaneously in many traditional societies to address such market problems as lack of formal insurance. We consider how these systems may potentially form a local poverty trap, by creating hurdles to adoption of new techniques of production. Although the aggregate outcome is impoverishing, it is shown that the kinship system may nevertheless fail to be dismantled as a result of individual incentives.

## *7.1 Corruption and rent-seeking*

Corruption is bad for growth. A number of ways that corruption retards development have been identified in the literature. First, corruption tends to reduce the incentive to invest by decreasing net returns and raising uncertainty. This effect impacts most heavily on increasing returns technologies with large fixed costs. Once costs are sunk, investors are subject to hold-up by corrupt officials, who can extort large sums. Also, governments and officials who have participated in such schemes find it difficult to commit credibly to new infrastructure projects.

Second, corruption diverts public expenditure intended for social overhead capital. At the same time, the allocation of such capital is distorted, because officials prefer infrastructure projects where large side payments are feasible. Corruption also hinders the collection of tax revenue, and hence the resource base of the government seeking to provide public infrastructure. Again, a lack of social overhead capital such as transport and communication networks tends to impact more heavily on the modern sector.

Third, innovators suffer particularly under a corrupt regime, because of their higher need for such official services as permits, patents and licenses (De Soto 1989; Murphy, Shleifer and Vishny 1993). The same is true for foreign investors, who bring in new technology. Lambsdorff (2003) finds that on average

[71] Following the excellent survey of Bardhan (1997), we define corruption to be "the use of public office for private gains, where an official (the agent) entrusted with carrying out a task by the public (the principal) engages in some sort of malfeasance for private enrichment which is difficult to monitor for the principal" (Bardhan 1997, p. 1321).

a 10% worsening in an index of transparency and corruption he constructs leads to a fall of 0.5 percentage points in the ratio of foreign direct investment to GDP.

Not only is corruption damaging to growth, but it also tends to breed more corruption. In other words, there are complementarities in corruption and other rent-seeking activities. It is this increasing returns nature of corruption which may serve to lock in poverty. Some equilibria will be associated with high corruption and low income, where many rent-seekers prey on relatively few producers. Others will have the reverse.

The decision of one official to seek bribes will increase expected net rewards to bribe taking in several ways. The most obvious of these complementarities is that when many agents are corrupt, the probability of detection and punishment for the marginal official is lowered. A related point is that if corruption is rampant then detection will not entail the same loss of reputation or social stigma as would be the case in an environment where corruption is rare. In other words, corruption is linked to social norms, and is one of the many reasons why they matter for growth.[72] Third, greater corruption tends to reduce the search cost for new bribes.

Murphy, Shleifer and Vishny (1993) point out yet another source of potential complementarities in rent-seeking. Their idea is that even if returns to predation are decreasing in an absolute sense, they may still be increasing *relative* to production. This would occur if the returns to productive activities—the alternative when agents make labor supply decisions—fall *faster* than those to rent-seeking as the number of rent-seekers increases. The general equilibrium effect is that greater rent-seeking decreases the (opportunity) cost of an additional rent-seeker.

In their model there is a modern sector, where output by any individual is equal to $a$, and a subsistence technology with which agents can produce output $c < a$. Alternatively, agents can prey on workers, obtaining for themselves an amount no more than $b$ per person, but limited by the amount of output available for predation. This in turn depends on the number of people working in the productive sectors. The authors assume, in addition, that only modern

[72] Transparency International's 2004 Global Corruption Report cites a statement by the president of the Government Action Observatory in Burundi that "corruption has spread, openly and publicly, to such an extent that those who practice it have become stronger than those who are fighting against it. This has led to a kind of *reversal of values*." (Emphasis added.)

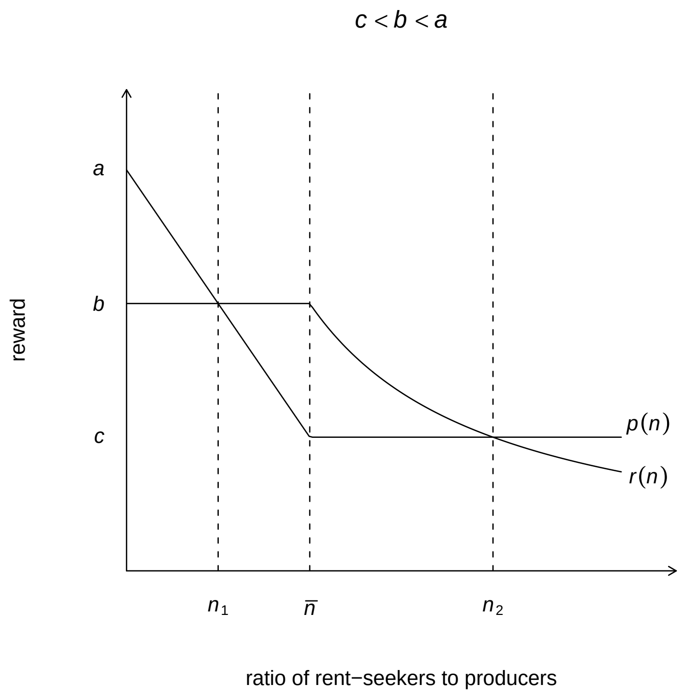


Fig. 27.

sector output can be appropriated by rent-seekers, so returns to subsistence farming are always equal to $c$.

An equilibrium is an allocation of labor across the different occupations such that returns to all are equal, and no individual agent can increase their reward by acting unilaterally. To locate equilibria, we now discuss returns to working in the different sectors as a function of $n$, which is defined to be the number of rent seekers for each modern sector producer.

Returns to employment in the subsistence sector are always given by $c$. Rent-seekers all take a slice $b$ of the pie until their ratio to modern sector producers $n$ satisfies $a - bn = c$. At this ratio, which we denote $\bar{n}$, the earnings of the modern sector producers fall to that of the subsistence producers, and the rent-seekers must reduce the size of their take (or earn nothing). After $\bar{n}$, the rent-seekers each take $(a - c)/n$, exactly equalizing returns to modern sector production and subsistence.

Let $p(n)$ and $r(n)$ be returns to modern sector production and rent-seeking respectively, so that $p(n) = (a-bn)\mathbb{1}\{n < \bar{n}\} + c\mathbb{1}\{n \geq \bar{n}\}$ and $r(n) = b\mathbb{1}\{n < \bar{n}\} + \frac{a-c}{n}\mathbb{1}\{n \geq \bar{n}\}$. These curves are drawn in Figure 27. The figure shows that there are multiple equilibria whenever the parameters satisfy $c < b < a$.

One is where all work in the modern sector. Then $n = 0$, and $p(n) = p(0) = a > r(n) = b > c$. This allocation is an equilibrium, where all agents earn the relatively high revenue available from modern sector production. In addition, because $b > c$, the payoff functions $n \mapsto p(n)$ and $n \mapsto r(n)$ intersect above $\bar{n}$, at $n_2$. This is again an equilibrium, where the payoffs to working in the subsistence sector, the modern sector and the rent-seeking sector are all equal and given by $c$.

Notice that $b$ does not affect income in either of these two equilibria. However, it does affect which one is likely to prevail. If $b$ declines below $c$, for example, then only the good equilibrium will remain. If it increases above $a$, then the bad equilibrium will be unique. When there are two equilibria, higher $b$ increases the basin of attraction for the bad equilibrium under myopic Marshallian dynamics.

In summary, the model exhibits a general equilibrium complementarity to corruption, which helps illustrate why corruption tends to be self-reinforcing, therefore causing poverty to persist. These kind of stories are important, because in practice corruption and related crimes tend to show a great deal of variation across time and space, often without obvious exogenous characteristics that would cause such variation.

There are many other models which exhibit self-reinforcement and path dependence in corruption. One is Tirole (1996), who studies the evolution of individual and group reputation. In his model, past behavior provides information about traits, such as honesty, ability and diligence. However, individual behavior is not perfectly observed. As a result, actions of the group or cohort to which the individual belongs have predictive power when trying to infer the traits of the individual. It follows that outcomes and hence incentives for the individual are affected by the actions of the group.

In this case we can imagine the following scenario. Young agents progressively joint an initial cohort of workers, a large number of whom are known to be corrupt. Because the behavior of new agents is imperfectly observed, they inherit the suspicion which already falls on the older workers. As a result, they may have little incentive to act honestly, and drift easily to corruption. This outcome in turn perpetuates the group's reputation for corrupt action.

One can contemplate many more such feedback mechanisms. For example, it is often said that the low wages of petty officials drive them to corruption. But if corruption lowers national output and hence income, then this will reduce the tax base, which in turn decreases the amount of resources with which to pay wages. For further discussion of corruption and poverty traps see Bardhan (1997).[73]

### *7.2 Kinship systems*

All countries and economies are made up of people who at one time were organized in small tribes with their own experiences, customs, taboos and conventions. Over time these tribes were united into cities, states and countries; and the economies within which they operated grew larger and more sophisticated. Some of these economies became vibrant and strong. Others have stagnated. According to North (1993, p. 4),

> The reason for differing success is straightforward. The complexity of the environment increased as human beings became increasingly interdependent, and more complex institutional structures were necessary to capture the potential gains from trade. Such evolution required that the society develop institutions that will permit anonymous, impersonal exchange across time and space. But to the extent that "local experience" had produced diverse mental models and institutions with respect to the gains from such cooperation, the likelihood of creating the necessary institutions to capture the gains from trade of more complex contracting varied.

North and other development thinkers have emphasized that success depends on institutions rewarding efficient, productive activity; and having sufficient flexibility to cope with the structural changes experienced in the transition to modernity. The degree of flexibility and ability to adapt determines to what extent an economy can take advantage of the application of science, of new techniques, and of specialization and the effective division of labor.

To illustrate these ideas, in this section we review recent analysis of the "kin" system, an institution found in many traditional societies, usually defined as an informal set of shared rights and obligations between extended family and friends for the purpose of mutual assistance.[74] Where markets and state institutions are less developed, the kin system replaces formal insurance and social security by implementing various forms of community risk sharing, and

[73] For other kinds of poverty traps arising through interactions between the state and markets, see, for example, Hoff and Stiglitz (2004), or Gradstein (2004).

[74] A related form of local poverty traps is those generated by neighborhood effects. See Durlauf (2004).

by the provision of other social services (Hoff and Sen 2004). The question we ask in the remainder of this section is how, in the process of development, the kin system interacts with the nascent modern sector, and whether or not it may serve to *impede* the diffusion of new technologies and the exploitation of gains from trade.

An interesting example of such analysis is Baker (2004), who interprets Africa's lack of robust growth as a failure of technology diffusion caused by institutional barriers. She presents a model of a rural African village, and suggests two path dependent mechanisms related to the kin system which may serve to retard growth. Both of them involve community risk sharing, and indicate how technology adoption may have positive network externalities beyond simple social learning.

The first mechanism concerns risk sharing among kin members in the form of interest free "loans" with no fixed repayment schedule. Kin members in need can expect to receive these transfers from the better off, who in turn must comply or face various social sanctions (including, in the countries Baker studied, accusations of witchcraft as the source of their good fortune). Beyond the obvious incentive effects on those who might seek to improve their circumstances by using new technology, Baker suggests that a kin member who adopts new techniques may face significant additional uncertainty vis-a-vis income net of transfers if the kin group makes mistakes in estimating his or her true profits. Such a miscalculation may lead to excessive demands for "gifts" or other transfers.

As Baker points out, the uncertainty effect of the transfers will be larger for those who adopt new technology, where costs and revenue are harder for the other kin to estimate. For example, the kin may have difficulty in measuring the real costs of new techniques, such as fertilizer or more expensive seed, causing them to overestimate true profits. (New techniques are often associated with higher revenues combined with higher costs.)

On the other hand, cost and net profit will be easier to estimate if more kin members have experience of the new techniques. In other words, uncertainty will be mitigated for the marginal adopter if more of his or her fellow kin members adopt the same technology. As a result there are positive network externalities in terms of expected cost. This mechanism generates a coordination problem, whereby a critical mass of co-adopters may be necessary to make the new technology more attractive than the old. This need for coordination may present a barrier to adoption.

At the same time, the coordination barrier would not seem to be insurmountable. Perhaps a kin group can negotiate to a better equilibrium when the gains are genuinely large? Baker suggests that in fact this will not be easy, because the risk sharing problem interacts with other path dependent institutions.

One of these concerns the nature of old age insurance among self-employed African farmers. Given the lack of state pensions and the difficulty of accumulating assets, support in old age may be contingent on the old providing some form of useful service to the household from which resources are to be acquired. And the most likely candidate for productive service from elderly farmers is the benefit of their experience. The problem here is that the value of this service provided by the old depends on a stagnant technology which does not change from generation to generation. Under new techniques the experience of old farmers may become redundant. If old farmers are able to resist the introduction of new techniques then it will be in their interests to do so. Once again, this is a source of multiple equilibria. The reason is that if the newer technology were already adopted then presumably it would be supported by old farmers, because this is then the methodology in which they have experience.

Another interesting study of the kin system has been conducted recently by Hoff and Sen (2004). They analyze the migration of kin members from rural areas to modern sector jobs, and show how network externalities arise in the migration decision. Even if kin members can coordinate on simultaneous migration, Hoff and Sen suggest that the kin group may put up barriers to prevent the loss of their most productive members. It is shown that even when the kin decisions are made by a majority, the barriers can be inefficient in terms of aggregate group welfare.

A simplified version of their story runs as follows. Kin members who do migrate may find themselves besieged by their less fortunate brethren. The latter come seeking not only "gifts" of cash transfers, but also help in finding jobs in the modern sector for themselves. Realizing this, employers will find it profitable to restrict employment of kin members. Here we assume these barriers are so high that migration while maintaining kin ties is never optimal. As a result, kin members choose between remaining in the rural sector or migrating while breaking their kin ties.

The kin group is thought of as a continuum of members with total mass of one. A fraction $\bar{\alpha} \in (0, 1)$ of the kin receive job offers in the modern sector.

The utility of remaining in the rural sector is

$$u_s(\alpha) = s_0 + b(1 - \alpha), \tag{29}$$

where here and elsewhere $\alpha \leq \bar{\alpha}$ is the fraction of the kin who break ties and move. The constant $s_0$ is a stand-alone payoff to rural occupation. The constant $b$ is positive, so that utility of staying is higher when more kin members remain. On the other hand, the utility of moving to the modern sector is

$$u_m(\alpha) = m_0 - c(1 - \alpha), \tag{30}$$

where $m_0$ is a payoff to working in the modern sector and $c$ is a positive constant. The function $\alpha \mapsto c(1 - \alpha)$ is the cost of ending kin membership (measured in the utility equivalent of various social sanctions which we will not describe). It is assumed that the cost of breaking kin ties for the marginal kin member decreases as more members leave the kin group and shift to the modern sector.[75]

Consider the interesting case, where $u_m(0) < u_s(0)$ and $u_m(\bar{\alpha}) > u_s(\bar{\alpha})$. A pair of curves for (29) and (30) which fit this pattern are depicted in Figure 28. If no kin members take modern sector jobs then it is not optimal to do so for an individual member. On the other hand, if all those with offers take up jobs, then their utility payoff will be higher than the payoff of those who remain.

If, as in the figure, we also have $u_m(\bar{\alpha}) > u_s(0)$, then it seems plausible that the kin members with job offers will coordinate their way to the equilibrium where all simultaneously move to modern sector jobs. Kin groups are not as diffuse as some other groups of economic actors, and coordination should prove correspondingly less problematic.

However, Hoff and Sen show that when kin members are heterogenous, a majority may take steps to forestall coordination by the productive critical mass on movement to the modern sector. Moreover, they may do so even when this choice is inefficient in terms of the kin's aggregate group payoff. In doing so, the kin group becomes a "dysfunctional institution," responsible for enforcing an inefficient status quo.

[75] Hoff and Sen cite Platteau (2000), who writes that to leave and enter the modern sector, a kin member "needs the protection afforded by the deviant actions of a sufficient number of other innovators in his locality. Rising economic opportunities alone will usually not suffice to generate dynamic entrepreneurs in the absence of a *critical mass* of cultural energies harnessed towards countering social resistance..." (Emphasis added.)

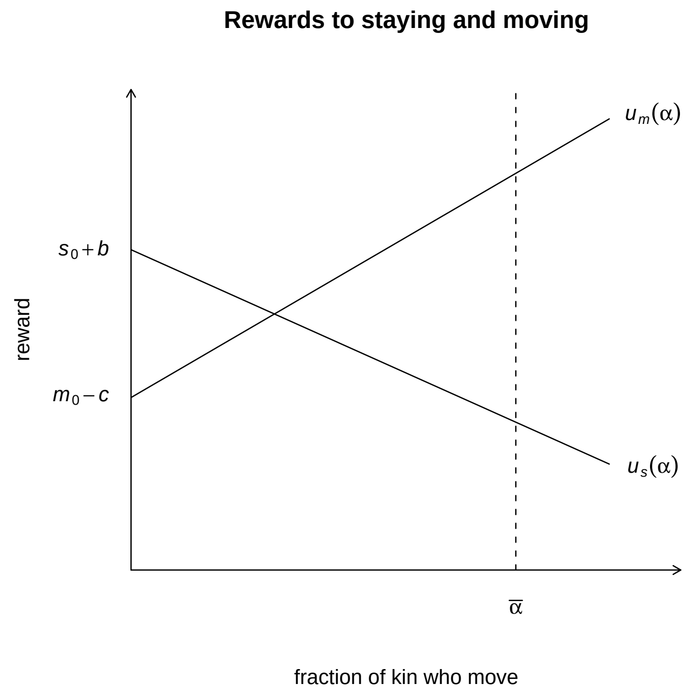


Fig. 28.

Their example works as follows. Consider a two stage game. First, the kin set the exit cost parameter $c$ by majority vote. The two possible values are $c^a$ and $c^b$, where $c^a < c^b$. Next, job offers are received, and kin members decide whether or not to move. Coordination always takes place in the situation where those with job offers together have a higher payoff in the modern sector.

There are now two types of kin members, those with high "ability" and those with low. The first type are of measure $\gamma$, and have probability $\alpha_H$ of getting a job offer from the modern sector. The second type are of measure $1-\gamma$, and have probability $\alpha_L$ of getting a job offer from the modern sector, where $0 < \alpha_L < \alpha_H < 1$. We assume that $\gamma < 1/2$, so high ability types are in the minority. Also, we assume that $\gamma\alpha_H + (1-\gamma)\alpha_L = \bar{\alpha}$. Ex post, the law of large numbers implies that the fraction of kin members who get job offers will again be $\bar{\alpha}$.

Regarding parameters, we assume that $u_m^a(\alpha) := m_0 - c^a(1-\alpha)$ satisfies $u_m^a(\bar{\alpha}) > u_s(0)$, but $u_m^b(\alpha) := m_0 - c^b(1-\alpha)$ satisfies $u_m^b(\bar{\alpha}) < u_s(0)$. The first inequality says that under the low cost regime, the payoff to working in the modern sector is greater than that of staying if all with job offers move. The second inequality says that under the high cost regime the opposite is true.

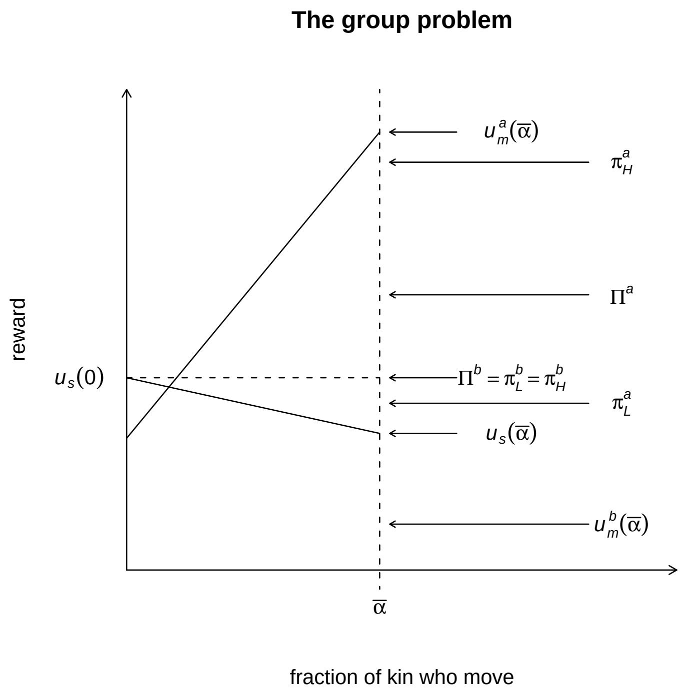


Fig. 29.

Because of coordination, under $c^a$ all of those with job offers will move. The ex ante payoff of the high ability types is

$$\pi_H^a := \alpha_H u_m^a(\bar{\alpha}) + (1 - \alpha_H) u_s(\bar{\alpha}),$$

while that of low ability types is

$$\pi_L^a := \alpha_L u_m^a(\bar{\alpha}) + (1 - \alpha_L) u_s(\bar{\alpha}).$$

Under $c^b$ all remain in the traditional sector, so the payoffs are $\pi_L^b := u_s(0) =: \pi_H^b$. Ex ante aggregate welfare measured as the sum of total payoffs is given under $c^a$ by

$$\Pi^a := \bar{\alpha} u_m^a(\bar{\alpha}) + (1 - \bar{\alpha}) u_s(\bar{\alpha}).$$

Under $c^b$ it is $\Pi^b := u_s(0)$.

What Hoff and Sen point out is that under some parameters it is possible to have

$$\pi_L^a < u_s(0) = \pi_L^b = \pi_H^b = \Pi^b < \Pi^a < \pi_H^a. \tag{31}$$

In this case $\pi_L^a < u_s(0) = \pi_L^b$, and since those with low ability are in the majority they will choose to set $c = c^b$. But then aggregate welfare is reduced, because $\Pi^b < \Pi^a$. This situation is illustrated in Figure 29.[76] Incentives are

[76] The parameters are $s_0 = 0.8$, $b = 0.2$, $m_0 = 2$, $c^a = 1.1$, $c^b = 2.3$, $\alpha_H = 0.9$,

such that the kinship institution perpetuates a low average income status quo.

## 8 Other Mechanisms

The poverty trap literature is vast, and even in a survey of this size many models must be neglected. A few of the more egregious omissions are listed in this section.

One of the earliest streams of literature on poverty traps is that related to endogenous fertility. A classic contribution is Nelson (1956), who shows how persistent underdevelopment can result from demographics. In his model, any increase in income lowers the death rate, which increases population and lowers capital stock per worker. If the population effect is stronger than diminishing returns then capital per worker cannot rise. See Azariadis (1996, Section 3.4) for other mechanisms and more references.

Other kinds of traps that arise in convex economies with complete markets include impatience traps and technology traps. Impatience traps typically involve subsistence levels of consumption, and sensitivity of consumption to income at low levels. See Magill and Nishimura (1984) or Azariadis (1996, Section 3.1). Technology traps are associated with low degrees of substitutability between capital and labor. See Azariadis (1996, Section 3.2).

See Dasgupta and Ray (1986) or Dasgupta (2003) for an introduction to the literature on malnutrition and underdevelopment. See also Basu and Van (1998) for a model of child labor with multiple equilibria.

## 9 Conclusions

The poor countries are not rich because they have failed to adopt the modern techniques of production which first emerged in Britain during the Industrial Revolution and then spread to some other nations in Western Europe and elsewhere. As a result, their economies have stagnated. By contrast, the rich countries possess market environments where the same techniques have been continuously refined, upgraded and extended, leading to what are now striking disparities between themselves and the poor.

$\alpha_L = 0.1$ and $\gamma = 0.45$.

Why would techniques not be adopted even when they are more efficient? Is it not the case that more efficient techniques are more profitable? The main objective of this survey has been to review a large number of studies which show why self-reinforcing traps may prevent the adoption of new technologies. For example, Section 5 showed how increasing returns can generate an incentive structure whereby agents avoid starting modern sector businesses, or invest little in their own training. Section 6 focused on credit market imperfections. Poor individuals lack collateral, which restricts their ability to raise funds. As a result, projects with large fixed costs are beyond the means of the poor, leaving them locked in low return occupations such as subsistence farming.

Recently many economists have highlighted the role of institutions in perpetuating poverty. Section 7 looked at why rent-seeking is both bad for growth and yet strongly self-reinforcing. Essentially similar societies may exhibit very different levels of predation simply as a result of historical accident, or some spontaneous coordination of beliefs. In addition, the role of kinship systems was analyzed as representative of the kinds of social conventions which may potentially harm formation of the modern sector.

Together, these mechanisms add up to a very different picture of development than the convex neoclassical benchmark model on which so much of modern growth theory has been based. Growth is not automatic. Small initial differences are magnified and then propagated through time. Poverty coexists with riches, much as it is observed to do in the cross-country income panel.

### *9.1 Lessons for economic policy*

There is a real sense in which poverty trap models are optimistic. Poverty is not the result of some simple geographic or cultural determinism. The poor are not condemned to poverty by a set of unfavorable exogenous factors, or even a lack of resources. Temporary policy shocks will have large and permanent effects if one-off interventions can cause the formation of new and better equilibria.

In practice, however, engineering the emergence of more efficient equilibria seems problematic for a number of reasons. First, we have seen many examples of how bad equilibria can be stable and self-reinforcing. In this case small policy changes are not enough to escape from their grip. Large changes must be made to the environment that people face, and the structure of their incentives. Such changes may be resisted by the forces that have perpetuated the inefficient equilibrium, such as a corrupt state apparatus fighting to preserve the status quo.

Second, coordinating changes in expectations and the status quo is difficult because norms and conventions are highly persistent. While it is possible to change policy and legislation almost instantaneously, it needs to be remembered that informal norms and conventions are often more important in governing behavior than the formal legalistic ones. Informal norms cannot be changed in the manner of interest rates, say, or tariffs. Rather they are determined within the system, and perpetuated by those forces that made them a stable part of the economy's institutional framework.

Third, policies can create new problems as a result of perverse incentives.[77] Successful policies will need to be carefully targeted, and operate more on the level of incentives than compulsion. These kinds of policies require a great deal of information. Traps which prevent growth and prosperity cannot be overcome without proper understanding and the careful design of policy.

## A Technical Appendix

Section A.1 gives a general discussion of Markov chains and ergodicity. The proof of Proposition 3.1 is outlined. Section A.2 gives remaining proofs.

### *A.1 Markov chains and ergodicity*

In the survey we repeatedly made use of a simple framework for treating Markov chains and ergodicity. The following is an elementary review. Our end objective is to sketch the proof of Proposition 3.1, but the review is intended to be more generally applicable.

Consider first a discrete time dynamical system evolving in state space $S \subset \mathbb{R}^n$.

[77] For example, in South Korea the state is generally credited with solving many of the coordination problems associated with industrialization in that country through their organization and support of large industrial conglomerates, and through active policy-based lending. However, these actions also led to a moral hazard problem, as the industrial groups became highly leveraged with government-backed loans. In the 1970s, investment was increasingly characterized by a costly combination of duplication and poor choices. Losses were massive, and motivated subsequent liberalization.

Just as for deterministic systems on $S$, which are represented by a *transition rule* associating each point in $S$ with another point in $S$—the value of the state next period—a Markov chain is represented by a rule associating each point in $S$ with a probability distribution over $S$. From this conditional distribution (i.e., distribution conditional on the current state $x \in S$) the next period state is drawn. In what follows the conditional distribution will be denoted by $\Gamma(x, dy)$, where $x \in S$ is the current state.

Because for Markov chains points in $S$ are mapped into probability distributions rather than into individual points, it seems that the analytical methods used to study the evolution of these processes must be fundamentally different to those used to study deterministic discrete time systems. But this is not the case: Markov chains can always be reduced to deterministic systems.

To see this, note that since the state variable $x_t$ is now a random variable, it must have some (marginal) distribution on $S$, which we call $\psi_t$. Suppose, as is often the case in economics, that $\psi_t$ is a density on $S$, and that the distribution $\Gamma(x, dy)$ is in fact a density $\Gamma(x, y)dy$ for every $x \in S$. In that case the marginal distribution for $x_{t+1}$ is a density $\psi_{t+1}$, and $\psi_{t+1}(y) = \int_S \Gamma(x, y)\psi_t(x)dx$. This last equality is just a version of the law of total probability: The probability of ending up at $y$ is equal to the probability of going to $y$ via $x$, weighted by the probability of being at $x$ now, summed over all $x \in S$.

Now define map $\mathbf{M}\colon \mathscr{D} \to \mathscr{D}$, where $\mathscr{D} := \{\varphi \in L_1(S) \mid \varphi \geq 0 \text{ and } \int \varphi = 1\}$ is the space of densities on $S$, by

$$\mathbf{M}\colon \mathscr{D} \ni \psi \mapsto (\mathbf{M}\psi)(\cdot) := \int_S \Gamma(x, \cdot)\psi(x)dx \in \mathscr{D}. \tag{A.1}$$

With this definition our law of total probability rule for linking $\psi_{t+1}$ and $\psi_t$ can be written simply as $\psi_{t+1} = \mathbf{M}\psi_t$. Since the map $\mathbf{M}$ is deterministic, we have succeeded in transforming our stochastic system into a deterministic system to which standard methods of analysis may be applied. The only difficulty is that the state space is now $\mathscr{D}$ rather than $S$. The latter is finite dimensional, while the former clearly is not.

The map $\mathbf{M}$ is usually called the stochastic operator or Markov operator associated with $\Gamma$. There are many good expositions of Markov operators in economics, including Stokey, Lucas and Prescott (1989) and Futia (1982). However those expositions treat the more general case, where $\Gamma(x, dy)$ does not necessarily have a density representation. Here it does, and it turns out that this extra structure is *very* useful for treating the models in this survey.

We wish to know when the difference equation $\psi_{t+1} = \mathbf{M}\psi_t$ has fixed points, and, more specifically, whether the system is globally stable in the sense that there is a unique fixed point $\psi^*$, and $\psi_t = \mathbf{M}^t\psi_0 \to \psi^*$ as $t \to \infty$ for all $\psi_0 \in \mathscr{D}$.[78] This is just ergodicity in the sense of Definition 3.1 on page 18.

Let $\|\cdot\|$ be the $L_1$ norm. Were $\mathbf{M}$ a uniform (Banach) contraction on $\mathscr{D}$, which is to say that $\exists\, \lambda < 1$ with $\|\mathbf{M}\psi - \mathbf{M}\psi'\| \leq \lambda\|\psi - \psi'\|$ for all $\psi, \psi' \in \mathscr{D}$, then ergodicity would hold because $\mathscr{D}$ is a closed subset of the complete metric space $L_1(S)$. Sadly, for continuous state Markov chains this uniform contraction property rarely holds. However it is often the case that $\mathbf{M}$ satisfies a weaker contraction condition:

**Definition A.1** *Let $T\colon X \to X$, where $(X, d)$ is a metric space. The map $T$ is called a T2 contraction if $d(Tx, Tx') < d(x, x')$ for every $x \neq x'$ in $X$.*

T2 contractions maps distinct points strictly closer together. A sufficient condition for $\mathbf{M}\colon \mathscr{D} \to \mathscr{D}$ to satisfy the T2 property is given below. The essential requirement is communication across all regions of the state space. Although T2 contractions do not always have fixed points (examples in $\mathbb{R}$ are easy to construct), they do if the state space is compact! In fact if $X$ is a compact set and $T\colon X \to X$ is a T2 contraction then $T$ has unique fixed point $x^* \in X$ and $T^t x \to x^*$ as $t \to \infty$ for all $x \in X$. This is just what we require for ergodicity when $\mathbf{M}$ is thought of as a map on $\mathscr{D}$.

Now $\mathscr{D}$ is not itself a compact set in the $L_1$ norm topology, but it may be the case that every orbit $(\mathbf{M}^t\psi_0)_{t\geq 0}$ of $\mathbf{M}$ is compact when taken with its closure. (From now on, call a set with compact closure *precompact*). Such a property is called *Lagrange stability.*[79] And it turns out that Lagrange stability can substitute for compactness of the state space $\mathscr{D}$: If $\mathbf{M}$ is (a) a T2 contraction, and (b) Lagrange stable, then the associated Markov chain is ergodic.[80]

How to establish Lagrange stability? To check precompactness of orbits it seems we must look at characterizations of compactness in $L_1$ (there is a famous one due to Kolmogorov), but Lasota (1994, Theorem 4.1) has proved that one need only check *weak* precompactness.[81] In fact it is sufficient to

[78] Here $\mathbf{M}^t$ is $t$ compositions of $\mathbf{M}$ with itself, and $\psi_0$ is the marginal distribution of $x_0$, so iterating the difference equation backwards gives $\psi_t = \mathbf{M}^t\psi_0$.

[79] That is, a self-mapping $T$ on topological space $X$ is called Lagrange stable if the set $\{T^t x \mid t \geq 0\}$ is precompact for every $x \in X$.

[80] The proof that Lagrange stability is sufficient is not hard. See Stachurski (2002, Theorem 5.2).

[81] Here is where the density structure is crucial. The operator $\mathbf{M}$ inherits nice prop-

check weak precompactness of orbits starting from $\psi \in \mathscr{D}_0$, where $\mathscr{D}_0$ is a (norm) dense subset of $\mathscr{D}$. Weak compactness is much easier to work with than norm compactness. Several well-known conditions are available.

Using one such condition due to Dunford and Pettis, Mirman, Reffett and Stachurski (2004) show that Lasota's criterion for Lagrange stability is satisfied when (i) there exists a continuous "norm-like" function $V \colon S \to \mathbb{R}$ and constants $\alpha, \beta \in [0, \infty)$, $\alpha < 1$, such that

$$\int \Gamma(x, y) V(y) dy \leq \alpha V(x) + \beta, \quad \forall x \in S; \tag{A.2}$$

and (ii) there exists a continuous function $h \colon S \to \mathbb{R}$ such that $\sup_{x \in S} \Gamma(x, y) \leq h(y)$ for all $y \in S$. By $V$ being norm-like is meant that $V$ is nonnegative, and that the sets $\{x \in S : V(x) \leq a\}$ are precompact for all $a$. (For example, when $S = \mathbb{R}^n$ it is easy to convince yourself that $x \mapsto \|x\|$ is norm-like. Note that when $S$ is a proper subset of $\mathbb{R}^n$ precompactness of sublevel sets refers to the *relative* Euclidean topology on $S$.)

Condition (i) is a standard drift condition, which pushes probability mass towards the center of the state space. This implies that orbits of the Markov process will be "tight." Tightness is a component of Dunford and Pettis' criterion for weak precompactness. Condition (ii) is just a technical condition which combines with (i) to fill out the requirements of the Dunford-Pettis criterion.

In the case of Proposition 3.1, we can take $S = (0, \infty)$, where $0 \notin S$ so that any stationary distribution we find is automatically nontrivial. One can then show that $V(x) = |\ln x|$ is norm-like on $S$, and a little bit of algebra shows that condition (i) holds for $\Gamma$ given in (10). Also, one can show that (ii) holds when $h(y) := 1/y$.[82]

This takes care of Lagrange stability. Regarding T2 contractiveness, one can show that $\mathbf{M}$ is a T2 contraction whenever the set $\operatorname{supp} \mathbf{M}\psi \cap \operatorname{supp} \mathbf{M}\psi'$ has positive measure for all $\psi, \psi' \in \mathscr{D}$, where $\operatorname{supp} f := \{x \in S \mid f(x) \neq 0\}$. This basically says that probability mass is mixed across the state space—all areas of $S$ communicate. In the case of (10) it is easy to show that $\operatorname{supp} \mathbf{M}\psi = (0, \infty) = S$ for every $\psi \in \mathscr{D}$. This is clearly sufficient for the condition.

---

erties from the fact that $\Gamma(x, dy)$ has a density representation. Also, we can work in $L_1$ rather than a space of measures. The former has a nice norm-dual space in $L_\infty$—helpful when dealing with weak precompactness.

[82] For more details see Stachurski (2004).

The proof of Proposition 5.1 in Section 5.2 is now given. The first point is that the banks $b = 1, \dots, B-1$ are equal-cost Bertrand competitors, and as a result always offer the interest rate $r$ to all firms in equilibrium. The main issue is the optimal strategy of the last bank $B$. So consider the following strategy $\sigma_B^*$ for $B$, which is illustrated with the help of Figure A.1. To firm $n$ the bank offers $i_n^*$ defined by $i_n^* = f[(n-1)/N] - 1$ if $n \leq \alpha_C N$. To the remaining firms $B$ offers the interest rate $r$. (Without loss of generality, we suppose that the index of firms from 1 to $N$ and the ranking of the offers made by $B$ always coincide.) Let $\sigma^* \in \Sigma$ be the strategy where $B$ offers $\sigma_B^*$ and all other banks offer $r$.

For the strategy $\sigma^*$ we have $\Omega(\sigma^*) = \{1\}$. The reason is that for $\alpha = n/N \leq \alpha_C$, firms $j = 1, \dots, n+1$ all satisfy

$$\pi(n/N, m_j(\sigma^*)) \geq \pi(n/N, i_j^*) \geq \pi(n/N, i_{n+1}^*) = 0.$$

In which case $\alpha \notin \Omega(\sigma^*)$ by (20). Also, for $\alpha \in (\alpha_C, 1)$ we have $\pi(\alpha, m_n(\sigma^*)) \geq \pi(\alpha, r) \geq 0$ for all $n \in \{1, \dots, N\}$, so again $\alpha \notin \Omega(\sigma^*)$. For the same reason, $1 \in \Omega(\sigma^*)$, because $\pi(1, m_n(\sigma^*)) \geq \pi(1, r) \geq 0$.

It follows that under this strategy $\alpha^{\text{pes}}(\sigma^*) = 1$. By (19) all firms enter. The profits of bank $B$ are given by the sum of the regions $P$, $Q$ and $R$, minus the region $O$, in Figure A.1. Here $Q$ and $\bar{\alpha}$ are chosen so that $P + Q - O = 0$. Thus, $\bar{\alpha}$ is the break-even point for the bank, where it recoups all losses made by offering cheap loans to firms in the "critical mass" region $[0, \alpha_C]$. If $\ell/N \geq \bar{\alpha}$ and hence $R \geq 0$, the bank $B$ makes positive profits.

It is not too hard to see that $\sigma^*$ is indeed the optimal strategy in $\Sigma$ for the banks. The banks $b = 1, \dots, B-1$ always offer $r$. For $B$, strategy $\sigma_B^*$ is optimal for the following reasons. First, if $B$ offers interest rates to $n \in \{1, \dots, N\}$ which are all less than or equal to those in $\sigma_B^*$, then all firms will enter as above, but $B$ will make lower profits by (17). So suppose that $B$ offers a schedule of rates $\{i_1^{**}, \dots, i_N^{**}\}$ where $i_n^{**} > i_n^*$ for at least one $n$, and let $k$ be the first such $n$. It is not difficult to see that the chain of logic whereby all firms enter now unravels: It must be that $k/N \leq \alpha_C$, because to other firms $B$ offers the rate $r$, which cannot be exceeded due to $B$'s competitors. One can now check that $(k-1)/N \in \Omega(\sigma^{**})$, and in fact $(k-1)/N = \min \Omega(\sigma^{**})$. As a result, $\alpha^{\text{pes}}(\sigma^{**}) = (k-1)/N$, and precisely $k-1$ firms enter. Clearly the profits of $B$ are lower for $\sigma^{**}$ than for $\sigma^*$.

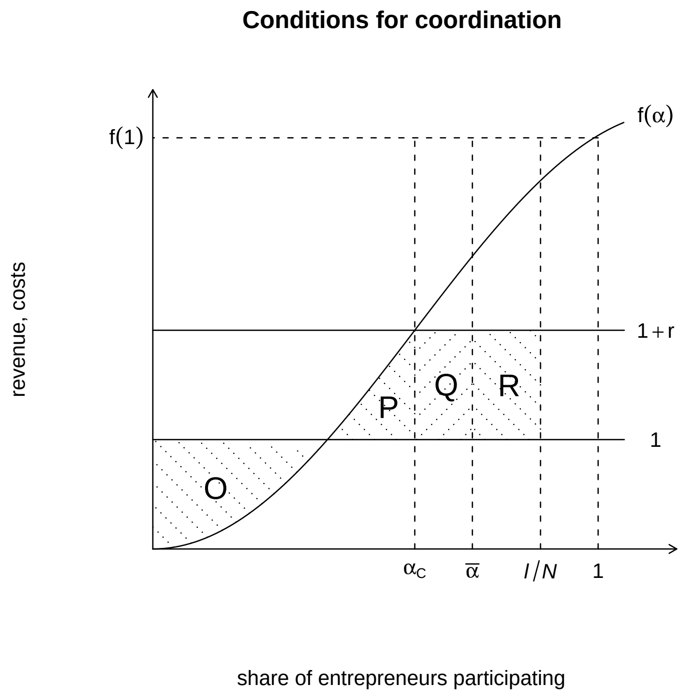


Fig. A.1.

## References

[1] Acemoglu, D. (1997): "Training and innovation in an imperfect labor market," *Review of Economic Studies,* 64 (3), 445–464.

[2] Acemoglu, D., S. Johnson and J. Robinson (this volume): "Institutions as the fundamental cause of long-run growth."

[3] Acemoglu, D. and F. Zilibotti (1997): "Was Prometheus unbound by chance? risk, diversification and growth," *Journal of Political Economy,* 105 (4), 709–751.

[4] Adsera, A. and D. Ray (1997): "History and coordination failure," *Journal of Economic Growth,* 3, 267–276.

[5] Aghion, P. and P. Bolton (1997): "A theory of trickle-down growth and development," *Review of Economic Studies,* 64, 151–172.

[6] Amir, R., L.J. Mirman and W.R. Perkins (1991): "One-sector nonclassical optimal growth: optimality conditions and comparative dynamics, *International Economic Review,* 32 (3), 625–644.

[7] Arthur, W. B. (1994): *Increasing Returns and Path Dependence in the Economy.* Ann Arbor: The University of Michigan Press.

[8] Azariadis, C. (1996): "The economics of poverty traps," *Journal of Economic Growth,* 1, 449–486.

[9] Azariadis, C. (2004): "The theory of poverty traps: what have we learned?" in *Poverty Traps,* (S. Bowles, S. Durlauf and K. Hoff, eds), in press.

[10] Azariadis, C. and A. Drazen (1990): "Threshold externalities in economic development," *Quarterly Journal of Economics,* 105, 501–526.

[11] Azariadis, C. and J. Stachurski (2004): "A forward projection of the cross-country income distribution," mimeo, Université Catholique de Louvain.

[12] Baker, E. (2004): "Institutional barriers to technology adoption in rural Africa," mimeo, Stanford University.

[13] Bandiera, O. and I. Rasul (2003): "Complementarities, social networks and technology adoption in northern Mozambique," mimeo.

[14] Banerjee, A. (2003): "The two poverties," mimeo, Massachusetts Institute of Technology.

[15] Banerjee, A. and A. Newman (1993): "Occupational choice and the process of development," *Journal of Political Economy,* 101, 274–298.

[16] Bardhan, P. (1997): "Corruption and development: a review of the issues," *Journal of Economic Literature,* 35, 1320–1346.

[17] Barrett, C. B., M. Bezuneh and A. Aboud (2001): "Income diversification, poverty traps and policy shocks in Côte d'Ivoire and Kenya," *Food Policy*, in press.

[18] Barrett, C. B. and B. M. Swallow (2003): "Fractal poverty traps," Strategies and Analysis for Growth and Access Working Paper, Cornell and Clark Atlanta Universities.

[19] Basu, K. and P. H. Van (1998): "The economics of child labor," *American Economic Review,* 88 (3), 412–427.

[20] Beath, J., Y. Katsoulacos and D. Ulph (1995): "Game-theoretic approaches to the modeling of technological change," in *Handbook of the Economics of Innovation and Technological Change,* (P. Stoneman, ed) Blackwell.

[21] Bianchi, M. (1997): "Testing for convergence: evidence from nonparametric multimodality tests," *Journal of Applied Econometrics,* 12, 393–409.

[22] Bloom, D. E., D. Canning and J. Sevilla (2003): "Geography and poverty traps," *Journal of Economic Growth,* 8, 355–378.

[23] Bowles, S., S. N. Durlauf and K. Hoff, eds. (2004): *Poverty Traps,* in press.

[24] Brock, W. A. and L. Mirman (1972): "Optimal economic growth and uncertainty: the discounted case," *Journal of Economic Theory,* 4, 479–513.

[25] Cabarello, R. and R. K. Lyons (2002): "The case for external economies," in *Political Economy, Growth and Business Cycles* (A. Cukierman et al., eds.) MIT Press, Cambridge.

[26] Cass, D. (1965): "Optimum growth in an aggregative model of capital accumulation," *Review of Economic Studies,* 32, 233–240.

[27] Chen, S. and M. Ravallion (2001): "How did the world's poor fare in the 1990s?" *Review of Income and Wealth,* 47 (3), 238–300.

[28] Checchi, D. and C. García-Peñalosa (2004): "Risk and the distribution of human capital," *Economics Letters,* 82 (1), 53–61.

[29] Conley, T. G. and C. R. Udry (2003): "Learning about a new technology: pinapple in Ghana," mimeo

[30] Da Rin, M. and T. Hellmann (2002): "Banks as catalysts for industrialization," *Journal of Financial Intermediation,* 11, 366–397.

[31] Dasgupta, P. (2003): "World poverty: causes and pathways," mimeo, University of Cambridge.

[32] Dasgupta, P. and D. Ray (1986): "Inequality as a determinant of malnutrition and unemployment," *The Economic Journal,* 96, 1011–1034.

[33] David, P. A. (1994) "Why are institutions the 'carriers of history'?: Path dependence and the evolution of conventions, organizations and institutions," *Structural Change and Economic Dynamics,* 5 (2), 205–220

[34] Dechert, W. D. and K. Nishimura (1983): "A complete characterization of optimal growth paths in an aggregated model with non-concave production function, *Journal of Economic Theory,* 31, 332–354.

[35] Den Haan, W. J. (1995): "Convergence of stochastic growth models: the importance of understanding why income levels differ," *Journal of Monetary Economics,* 35, 65–82.

[36] Dercon, S. (1998): "Wealth, risk and activity choice: cattle in western Tanzania," *Journal of Development Economics,* 55, 1–42.

[37] Dercon, S. (2003) "Risk and poverty: a selective review of the issues," mimeo, Oxford University.

[38] Desdoigts, A. (1999): "Patters of economic development and the formation of clubs," *Journal of Economic Growth,* 4 (3), 305–330.

[39] De Soto, H. (1989): *The Other Path: The Invisible Revolution in the Third World,* New York: Harper and Row.

[40] Dimaria, C. H. and C. Le Van (2002): "Optimal growth, debt, corruption and R&D," *Macroeconomic Dynamics,* 6, 597–613.

[41] Durlauf, S. N. (1993): "Nonergodic economic growth," *Review of Economic Studies,* 60, 349–366.

[42] Durlauf, S. N. and P. A. Johnson (1995): "Multiple regimes and cross-country growth behavior", *Journal of Applied Econometrics,* 10 (4), 365-384.

[43] Durlauf, S.N. (2004): "Neighborhood effects," in *Handbook of Urban and Regional Economics,* (J.V. Henderson and J.F. Thisse, eds), in press.

[44] Easterly, W. (2001): *The Elusive Quest for Qrowth: Economists' Adventures and Misadventures in the Tropics,* The MIT Press, Cambridge, Massachusetts.

[45] Easterly, W., M. Kremer, L. Pritchett and L. H. Summers (1993): "Good policy or good luck: country growth performance and temporary shocks," *Journal of Monetary Economics,* 32, 459–484.

[46] Easterly, W. and R. Levine (2000): "It's not factor accumulation: stylized facts and growth models," *World Bank Economic Review,* 15 (2), 177–219.

[47] Engerman, S. and K. L. Sokoloff (2004): "The persistence of poverty in the Americas: the role of institutions," in *Poverty Traps,* (S. Bowles, S. Durlauf and K. Hoff, eds), in press.

[48] Feyrer, J. (2003): "Convergence by parts," mimeo, Dartmouth College.

[49] Freeman, S. (1996): "Equilibrium income inequality among identical agents," *Journal of Political Economy,* 104 (5), 1047–1064.

[50] Futia, C. A. (1982): "Invariant distributions and the limiting behavior of Markovian economic models," *Econometrica,* 50, 377–408.

[51] Galor, O. and J. Zeira (1993): "Income distribution and macroeconomics," *Review of Economic Studies,* 60, 35–52.

[52] Glynn, P. W. and S. G. Henderson (2001): "Computing densities for Markov chains via simulation," *Mathematics of Operations Research,* 26, 375–400.

[53] Gradstein, M. (2004): "Governance and growth," *Journal of Development Economics,* 73, 505–518.

[54] Graham, B. S. and J. Temple (2004): "Rich nations poor nations: how much can multiple equilibria explain?" mimeo (revised version of CEPR Discussion Paper 3046).

[55] Greif, A., P. Milgrom and B.R. Weingast (1994): "Coordination, commitment and enforcement: the case of merchant guilds," *Journal of Political Economy,* 102 (4), 745–776.

[56] Hall, R. E. and C. I. Jones (1999): "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics,* 114 (1), 84–116.

[57] Heston, A., R. Summers and B. Aten (2002): "Penn World Table version 6.1." Center for International Comparisons, University of Pennsylvania.

[58] Hoff, K. (2000): "Beyond Rosenstein-Rodan: the modern theory of coordination problems in development," *Proceedings of the World Bank Annual Conference on Development Economics 2000*, in press.

[59] Hoff, K. and A. Sen (2004): "The kin system as a poverty trap," in *Poverty Traps,* (S. Bowles, S. Durlauf and K. Hoff, eds), in press.

[60] Hoff, K. and Pandey (2004): "Belief systems and durable inequalities: an experimental investigation of Indian caste," World Bank Policy Research Working Paper 3351.

[61] Hoff, K. and J. Stiglitz (2004): "After the Big Bang? Obstacles to the emergence of the rule of law in post-communist societies," *American Economic Review,* 94 (3) 753–763.

[62] Ihaka, R. and R. Gentleman (1996): "R: a language for data analysis and graphics," *Journal of Computational and Graphical Statistics,* 5 (3), 299–314.

[63] Jalan, J. and M. Ravallion (2002): "Geographic poverty traps? A micro model of consumption growth in rural China," *Journal of Applied Econometrics,* 17, 329–346.

[64] Johnson, P. A. (2004): "A continuous state space approach to convergence by parts," mimeo, Vassar College.

[65] Kehoe, T. and D. Levine (1993): "Debt constrained asset markets," *Review of Economic Studies,* 60, 865–888.

[66] King, R. G. and T. Rebelo (1993): "Transitional dynamics and economic growth in the neoclassical model," *American Economic Review,* 83 (4), 908–931.

[67] Kiyotaki, N. and J. H. Moore (1997): "Credit cycles," *Journal of Political Economy,* 105 (2), 211–248.

[68] Krugman, P. (1991): "History versus expectations," *The Quarterly Journal of Economics,* 106 (2), 651–667.

[69] Koopmans, T. (1965): "On the concept of optimal economic growth," *Pontificae Academiae Scientiarum Scripta Varia,* 28, 225–300.

[70] Kremer, M. (1993): "The O-ring theory of economic development," *Quarterly Journal of Economics,* August, 551–575.

[71] Lambsdorff, J. G. (2003): "How corruption affects persistent capital flows," *Economics of Governance,* 4 (3), 229–243.

[72] Lasota, A. (1994), "Invariant principle for discrete time dynamical systems," *Universitatis Iagellonicae Acta Mathematica* 31, 111–127.

[73] Limao, N. and A.J. Venables (2001): "Infrastructure, geographical disadvantage, transport costs and trade," *World Bank Economic Review,* 15, 451–479.

[74] Ljungqvist, L. (1993): "Economic underdevelopment: the case of the missing market for human capital," *Journal of Development Economics,* 40, 219–239.

[75] Loury, G. C. (1981): "Intergenerational transfers and the distribution of earnings," *Econometrica,* 49, 843–867.

[76] Lucas, R. E. Jr, (1986): "Adaptive behavior and economic theory," *Journal of Business,* 59 (4), 401–426.

[77] Lucas, R. E. (1990): "Why doesn't capital flow from rich to poor countries?" *American Economic Review,* 80 (2), 92-96

[78] Maddison, A. (1995): *Monitoring the World Economy,* OECD Development Center, Paris.

[79] Magill, M. and K. Nishimura (1984): "Impatience and accumulation," *Journal of Mathematical Analysis and Applications,* 98, 270–281.

[80] Majumdar, M., T. Mitra and Y. Nyarko (1989): "Dynamic optimization under uncertainty: non-convex feasible set" in *Joan Robinson and Modern Economic Theory,* (G. R. Feiwel, ed) MacMillan Press, New York.

[81] Matsuyama, K. (1991): "Increasing returns, industrialization and indeterminacy of equilibrium," *Quarterly Journal of Economics,* 106, 617–650.

[82] Matsuyama, K. (1995): "Complementarities and cumulative processes in models of monopolistic competition," *Journal of Economic Literature,* 33, 701–729.

[83] Matsuyama, K. (1997): "Complementarity, instability and multiplicity," *Japanese Economic Review,* 48 (3), 240–266.

[84] Matsuyama, K. (2000): "Endogenous inequality," *Review of Economic Studies,* 67, 743–759.

[85] Matsuyama, K. (2004): "Financial market globalization, symmetry-breaking, and endogenous inequality of nations," *Econometrica,* 72, 853–884.

[86] Matsuyama, K. and A. Ciccone (1996): "Start-up costs and pecuniary externalities as barriers to economic development," *Journal of Development Economics,* 49, 33–59.

[87] Mirman, L.J., O.F. Morand and K. Reffett (2004), "A qualitative approach to Markovian equilibrium in infinite horizon economies with capital," mimeo, Arizona State University.

[88] Mirman, L. J., K. Reffett and J. Stachurski (2004): "Some stability results for Markovian economic semigroups," *International Journal of Economic Theory,* in press.

[89] Mokyr, J. (2002): *The Gifts of Athena: Historical Origins of the Knowledge Economy,* Princeton University Press, Princeton New Jersey.

[90] Mookherjee, D. and D. Ray (2001) *Readings in the Theory of Economic Development,* Blackwell Publishing, New York.

[91] Mookherjee, D. and D. Ray (2003): "Persistent inequality," *Review of Economic Studies,* 70, 369–393.

[92] Morduch, J. (1990): "Risk, production and saving: Theory and evidence from Indian households," Harvard University, mimeo.

[93] Murphy, K. M., A. Shleifer and R. W. Vishny (1989): "Industrialization and the big push," *Journal of Political Economy,* 97, 1003–1026.

[94] Murphy, K. M., A. Shleifer and R. W. Vishny (1993): "Why is rent-seeking so costly to growth?" *American Economic Review,* 83 (2), 409–414.

[95] Nelson, R. R. (1956): "A theory of the low level equilibrium trap", *American Economic Review,* 46, 894–908.

[96] Nishimura, K. and J. Stachurski (2004): "Stability of stochastic optimal growth models: a new approach," *Journal of Economic Theory,* in press.

[97] North (1993): "The new institutional economics and development," WUSTL Economics Working Paper Archive.

[98] North (1995): "Some fundamental puzzles in economic history/development," WUSTL Economics Working Paper Archive.

[99] Nurkse, R (1953): *Problems of Capital-Formation in Underdeveloped Countries,* 1962 edition, New York: Oxford Univeristy Press.

[100] Ottaviano, A. and J.F. Thisse (2004): "Agglomeration and economic geography," in *Handbook of Urban and Regional Economics,* (J.V. Henderson and J.F. Thisse, eds), in press.

[101] Parente, S. L. and E. C. Prescott (this volume): "A unified theory of the evolution of international income levels."

[102] Piketty, T. (1997): "The dynamics of the wealth distribution and the interest rate with credit rationing," *Review of Economic Studies,* 64, 173–189.

[103] Platteau, J-P. (2000): *Institutions, Social Norms and Economic Development* Amsterdam: Harwood Publishers

[104] Prescott, E. C. (1998): "Needed: a theory of total factor productivity," *International Economic Review,* 39, 529–549.

[105] Pritchett, L. (1997): "Divergence, big time," *Journal of Economic Perspectives,* 11 (3), 3–17.

[106] Quah, D. T. (1993): "Empirical cross-section dynamics in economic growth," *European Economic Review,* 37, 426-434.

[107] Quah, D. (1996): "Convergence empirics across economies with (some) capital mobility," *Journal of Economic Growth,* 1, 95–124.

[108] Radlet, S. (2004): "Aid effectiveness and the Millenium Development Goals," manuscript prepared for the Millenium Project Task Force, United Nations Development Group.

[109] Ray, D. (1990): "Income distribution and macroeconomic behavior," mimeo, New York University.

[110] Ray, D. (2003): "Aspirations, poverty and economic change," mimeo, New York University.

[111] Ray, D. and P. Streufert (1993): "Dynamic equilibria with unemployment due to undernourishment," *Economic Theory,* 3, 61–85.

[112] Redding, S. and A. J. Venables (2004): "Economic geography and international inequality," *Journal of International Economics,* 62, 53–82.

[113] Rodríguez-Clare, A. (1996): "The division of labor and economic development," *Journal of Development Economics,* 49, 3–32.

[114] Rodrik, D. (1996): "Coordination failures and government policy: A model with applications to East Asia and Eastern Europe," *Journal of International Economics,* 40, 1–22.

[115] Romer, P. M. (1986): "Increasing returns and long-run growth," *Journal of Political Economy,* 94 (5), 1002-1037.

[116] Romer, P. M. (1990): "Are nonconvexities important for understanding growth?," *American Economic Review,* 80 (2), 97–103.

[117] Rosenstein-Rodan, P. (1943): "The problem of industrialization of eastern and south-eastern Europe," *Economic Journal,* 53, 202–211

[118] Rostow, W. W. (1975): *How it All Began: Origins of the Modern Economy,* McGraw-Hill, New York.

[119] Rostow, W. W. (1990): *The Stages of Economic Growth: A Noncommunist Manifesto,* 3rd edition, Cambridge University Press, Cambridge.

[120] Sachs, J. D., J. W. McArthur, G. Schmidt-Traub, M. Kruk, C. Bahadur, M. Faye and G. McCord, "Ending Africa's poverty trap," mimeo

[121] Simon, H. A. (1986): "Rationality in psychology and economics," *Journal of Business,* 59 (4), 209–224.

[122] Skiba, A. K. (1978): "Optimal growth with a convex-concave production function," *Econometrica,* 46, 527–539.

[123] Solow, R. M. (1956): "A contribution to the theory of economic growth." *Quarterly Journal of Economics,* 70, 65–94.

[124] Stachurski, J. (2002): "Stochastic optimal growth with unbounded shock," *Journal of Economic Theory,* 106, 45–60.

[125] Stachurski, J. (2004): "Stochastic economic dynamics," mimeo, the University of Melbourne.

[126] Starrett, D. (1978): "Market allocations of location choice in a model with free mobility," *Journal of Economic Theory,* 17, 21–37.

[127] Stokey, N. L., R. E. Lucas and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics,* Harvard University Press, Massachusetts.

[128] Sugden, R. (1989): "Spontaneous order," *Journal of Economic Perspectives,* 3 (4), 85–97.

[129] Tirole, J. (1996): "A theory of collective reputations (with applications to persistence of corruption and to firm quality)," *Review of Economic Studies,* 63, 1–22.

[130] Tsiddon, D. (1992): "A moral hazard trap to growth," *International Economic Review,* 33 (2), 299–321.

[131] Van Huyck, J. B., J. P. Cook and R. C. Battalio (1997): "Adaptive behavior and coordination failure," *Journal of Economic Behavior and Organization,* 32 (4), 483–503

[132] Young, A. A. (1928): "Increasing returns and economic progress," *Economic Journal,* 28, 527–542.

[133] Zilibotti, F. (1995): "A Rostovian model of endogenous growth and underdevelopment traps," *European Economic Review,* 39, 1569–1602.