Institutions, Development, and Economic Growth

edited by Theo S. Eicher
and Cecilia García-Peñalosa

CESifo Seminar Series

Institutions, Development, and Economic Growth

CESifo Seminar Series
Edited by Hans-Werner Sinn

Inequality and Growth: Theory and Policy Implications
Theo S. Eicher and Stephen J. Turnovsky, editors

Public Finance and Public Policy in the New Century
Sijbren Cnossen and Hans-Werner Sinn, editors

Spectrum Auctions and Competition in Telecommunications
Gerhard Illing and Ulrich Klüh, editors

Managing European Union Enlargement
Helge Berger and Thomas Moutos, editors

European Monetary Integration
Hans-Werner Sinn, Mika Widgrén, and Marko Köthenbürger, editors

Measuring the Tax Burden on Capital and Labor
Peter Birch Sørensen, editor

A Constitution for the European Union
Charles B. Blankart and Dennis C. Mueller, editors

Labor Market Institutions and Public Regulation
Jonas Agell, Michael Keen, and Alfons J. Weichenrieder, editors

Venture Capital, Entrepreneurship, and Public Policy
Vesa Kanniainen and Christian Keuschnigg, editors

Exchange Rate Economics: Where Do We Stand?
Paul De Grauwe, editor

Prospects for Monetary Unions after the Euro
Paul De Grauwe and Jacques Mélitz, editors

Reasons and Remedies for Western Europe's Structural Unemployment
Martin Werding, editor

Institutions, Development, and Economic Growth
Theo S. Eicher and Cecilia García-Peñalosa, editors

Institutions, Development, and Economic Growth

Theo S. Eicher and
Cecilia García-Peñalosa, editors

CESifo Seminar Series

The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Palatino on 3B2 by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Institutions, development, and economic growth / Theo S. Eicher and Cecilia García-Peñalosa, editors.
p. cm. — (CESifo seminar series)
Includes bibliographical references and index.
ISBN 0-262-05081-1 (hc : alk. paper)
1. Economic development. 2. Economic policy. 3. Social institutions. I. Eicher, Theo S. II. García-Peñalosa, Cecilia. III. Series.

HD75.I57 2006
338.9—dc22 2006042010

10 9 8 7 6 5 4 3 2 1

Contents

Series Foreword vii
Preface ix

I Institutions and Economic Performance 1

1 On Institutions and Growth 3
Philippe Aghion

2 Financial Institutional Reform, Growth, and Equality 33
Costas Azariadis and David de la Croix

3 How Do Institutions Lead Some Countries to Produce So Much More Output per Worker than Others? 65
Theo S. Eicher, Cecilia García-Peñalosa, and Utku Teksoz

4 Regulation and Economic Performance: Product Market Reforms and Productivity in the OECD 81
Giuseppe Nicoletti and Stefano Scarpetta

II The Impact of Institutions on Innovation and Entrepreneurship 121

5 Institutions and Technological Innovation during Early Economic Growth: Evidence from Great Inventors in the United States, 1790–1930 123
B. Zorina Khan and Kenneth L. Sokoloff

6 On the Efficacy of Reforms: Policy Tinkering, Institutional Change, and Entrepreneurship 159
Murat Iyigun and Dani Rodrik

III Endogenous Institutions 189

7 The Role of Higher Education Institutions: Recruitment of Elites and Economic Growth 191
Elise S. Brezis and François Crouzet

8 Growth and Endogenous Political Institutions 215
Matteo Cervellati, Piergiuseppe Fortunato, and Uwe Sunde

9 The Road from Agriculture 249
Thorvaldur Gylfason and Gylfi Zoega

Contributors 281
Index 283

Series Foreword

This book is part of the CESifo Seminar Series in Economic Policy, which aims to cover topical policy issues in economics from a largely European perspective. The books in this series are the products of the papers presented and discussed at seminars hosted by CESifo, an international research network of renowned economists supported jointly by the Center for Economic Studies at Ludwig-Maximilians-Universität, Munich, and the Ifo Institute for Economic Research. All publications in this series have been carefully selected and refereed by members of the CESifo research network.

Hans-Werner Sinn

Preface

Growth theory has identified a plethora of determinants that are crucial for successful development. To explain differences in economic performance, economists focused for decades on physical/human capital and technical change as sources of the wealth of nations. Failed transition experiments and financial crises in the 1990s revealed that even the basic prerequisites for development are incapable of delivering desired living standards in the absence of functioning institutions that support and enable economic incentives.

What are the institutions that seem fundamental to economic performance in developing and developed countries alike? A review in the *Handbook of Economic Growth* (Acemoglu, Johnson, and Robinson 2005) points to a distinguished history of the subject, including works by John Locke, Adam Smith, and John Stuart Mill. Nevertheless, economics still lacks a robust, general framework that provides guidelines for why and how institutions influence the surprisingly large and unexplained differences in per capita incomes across countries.

The past ten years have provided an abundance of empirical studies on the influence of institutions. Trailblazers were researchers who based their empirical analyses on subjective indices provided by private country risk assessment companies; these data were first used by Knack and Keefer (1995) and subsequently by Hall and Jones (1999) and Acemoglu, Johnson, and Robinson (2001) to establish the influence of institutions on per capita income. Since then a hunt has begun to uncover ever better measures of institutions, as well as the mechanisms by which institutions influence development.

This volume provides an overview of the current state of the literature regarding the impact of institutions on growth. The book opens with a chapter by Philippe Aghion that highlights some of the key arguments linking institutions and growth. Institutions, Aghion argues, have many facets, each impacting growth and development differently. His approach emphasizes three aspects. First, convergence depends crucially on the quality of financial institutions. It is often argued that countries that are further from the technological frontier benefit from a catching-up process and hence grow faster than those closer to the frontier. Aghion maintains that because technological catch-up requires investment in imitation, the quality of financial institutions becomes an essential element in the catch-up process. As a result, underdeveloped financial markets can totally offset the advantages of technological backwardness and result in slower growth in backward countries than in those closer to the technological frontier. Second, Aghion examines how, contrary to common wisdom, product market competition may not be detrimental to growth. The basic idea is that a fierce competition will force firms to innovate if they want to remain ahead of other producers in the sector and make positive profits. Aghion introduces the concept of "appropriate institutions," by which he means that certain institutional setups will be suitable at some levels of development but not at others. For example, in the early stages of industrialization, when capital accumulation is important, institutions that favor long-term relationships between firms and banks are optimal. However, as an economy moves into the phase in which growth is driven by innovation, more flexible institutional arrangements that foster entrepreneurship and risk taking are preferable. As a result, the institutions that promoted growth at one stage are precisely those that retard it at another stage. Third, Aghion discusses the difference between academic institutions and private firms in promoting innovation.

There is general consensus that financial institutions might be among the most important in development, next to secure property rights. In "Financial Institutional Reform, Growth, and Equality" (chapter 2), Costas Azariadis and David de la Croix explore the consequences of liberalized credit markets for growth and inequality. The key insight in this chapter is that premature liberalization in the least developed countries (low total factor productivity or capital intensity) may redirect economic growth toward a poverty trap. This highlights the importance of understanding the exact contribution of institutions to growth. Reforms for the sake of reforms may actually harm growth if they are not sequenced correctly.

The next two chapters turn to the empirical evidence on the effect of institutions on economics performance. In chapter 3, Theo S. Eicher, Cecilia García-Peñalosa, and Utku Teksoz ask "How Do Institutions

Lead Some Countries to Produce So Much More Output per Worker than Others?" Their purpose is to examine the mechanisms by which institutions might affect economic growth. They first combine the two most influential approaches to explaining differences in per capita income across countries, the growth accounting approach of Mankiw, Romer, and Weil (1991), in which physical and human capital stocks determine output, and the methodology of Hall and Jones (1999), in which institutional quality has a direct influence on GDP. Their analysis hence seeks to understand the degree to which institutions actually enhance the productivity of skilled labor and investment. The surprising result is that while physical capital and institutions are complements in development, human capital and institutions are shown to be substitutes. That is, in countries with weak institutions, human capital is critical to development; in those with high levels of human capital, institutional quality has a much weaker impact on output.

Differences in institutions are clearly a major source of income gaps between developed and developing countries, as highlighted by the first three chapters of the book. However, institutional reform can also be a source of growth in industrial economies. For example, in chapter 1, Aghion argues that the degree of product market competition can influence the amount of innovation taking place in industrial economies. In "Regulation and Economic Performance: Product Market Reforms and Productivity in the OECD" (chapter 4), Giuseppe Nicoletti and Stefano Scarpetta examine the role of institutional reforms in the OECD. The last two decades have witnessed substantial institutional and regulatory reforms in OECD countries. The differences in these reforms across countries have provided a suitable natural experiment to assess the macroeconomic impact of such reforms. Nicoletti and Scarpetta review the literature on the effect of these reforms on investment, productivity, and employment. The evidence suggests that strengthening private governance and increased competition in product and labor markets have had a major positive impact on labor productivity, and can help understand differences across countries and over time.

The new growth theories have emphasized the role of innovation and entrepreneurship on growth, and this is the focus of part II. We start with chapter 5 by B. Zorina Khan and Ken Sokoloff, "Institutions and Technological Innovations during Early Economic Growth: Evidence from Great Inventors in the United States, 1790–1930," on the impact of patent legislation on patenting activity in the United States during the period from 1790 to 1930. Their analysis emphasizes that a well-functioning system of intellectual property rights turns patents into tradable assets. The authors highlight the contrast between the U.S. patent system and that prevailing in Europe at the time. The major difference concerned the use of an examination system in the United States. In Europe, the inventor would obtain a patent upon payment of a fee, but this patent could be challenged in court implying that property rights could not be considered to be fully established until the case had been assessed in court. In the United States, a patent application was subject to examination, and only once rightful property rights over the innovation were established would the payment be made. This system established ownership in a way that could not be challenged. Khan and Sokoloff show that the use of the examination system had two important implications. First, it encouraged innovation by individuals of all education levels. Second, it resulted in extensive patent selling and licensing, with double benefits in the form of ensuring that the goods were produced and ensuring that the innovator had access to funds permitting the continuation of innovation.

Chapter 6, "On the Efficacy of Reforms: Policy Tinkering, Institutional Change, and Entrepreneurship" by Murat Iyigun and Dani Rodrik, focuses on the relationship between entrepreneurship and reform. Iyigun and Rodrik examine how policy affects entrepreneurship. Two alternatives are considered, "policy tinkering" and institutional reform. The authors argue that growth is largely due to an increase in the number of available products, and that product diversification requires entrepreneurs who invest and discover new products. The central insight in their model is that low growth is due to an insufficient level of entrepreneurship. Policy tinkering can improve entrepreneurial rewards marginally; and deep institutional reforms can make substantial changes in the reward structure but at a cost to incumbent entrepreneurs. As a result, the efficacy of one policy or the other will depend on the current state of the economy in terms of entrepreneurship. Their empirical evidence supports this hypothesis: major reforms have worked in countries with a low level of entrepreneurial activity, and failed otherwise. The chapter hence captures one of the key messages developed in chapter 1, namely, that some types of institutions are appropriate at certain stages of development but not at others.

Clearly the role, functioning, and quality of institutions is itself endogenous to level of development. The third part of the book examines the implications of this endogeneity in terms of education and the political process. In "The Role of Higher Education Institutions: Recruitment of Elites and Economic Growth" (chapter 7), Elise S. Brezis and François Crouzet examine a specific mechanism by which institutions influence the fortunes of an economy. The authors analyze the evolution of the recruitment of elites over time and highlight how recruitment institutions subsequently impact the economy. The key result is that meritocratic recruitment actually leads to class stratification and auto-recruitment. Auto-recruitment is then shown to lead to a stratification of the economy, which may be the most dramatic impact of meritocratic institutions on economic growth.

In chapter 8, "Growth and Endogenous Political Institutions," Matteo Cervellati, Piergiuseppe Fortunato, and Uwe Sunde study the dynamics of political institutions and the implied differences in public policies. They highlight the circular nature of institutions: political institutions are thought to be influenced by economic development, and economic development in turn influences the political institutions. The chapter highlights that economic development increases the likelihood of transitions from oligarchy to democracy. Moreover, the authors show that democratic regimes tend to provide more efficient public policies, and more redistribution, than oligarchic regimes.

In "The Road from Agriculture" (chapter 9), Thorvaldur Gylfason and Gylfi Zoega seek to explain economic backwardness not in terms of history or mentality but rather in terms of rational agents' maximizing behavior. They show that observed technology adoption in agriculture does not need to coincide with the "frontier" technology at all stages of development. Instead, Gylfason and Zoega show that countries may feature an "optimal technology gap." The size of this gap is shown to depend on factors exogenous to most economic models and seldom subject to change, such as farm size (geography), land productivity, and the ability of farmers to digest and adopt new technologies.

This volume is based on contributions presented at the 2004 CESifo Venice Summer Institute on Institutions and Growth. We thank CESifo for supporting the research project and Roisin Hearn and Nicola Papaphilippou for outstanding logistic support throughout. The quality of the conferences and papers was assured by extensive comments from discussants including Matteo Cervellati, Alain Desdoigts, Thorvaldur Gylfason, Piergiuseppe Fortunato, B. Zorina Khan, Andreas Leukert, Omar Licandro, Stefano Scarpetta, Utku Teksoz, Joachim Voth, and a number of anonymous referees. Finally, we would like to thank our respective spouses, Nello Dolgetta and Regina Lyons, for their loving support and our respective kids, Claudia, Diego, and Luis, for making our work on this volume so enjoyable.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91, no. 5 (December): 1369–1401.

Acemoglu, Daron, Simon Johnson, and James Robinson. 2005. "Institutions as the Fundamental Cause of Long-run Growth." Forthcoming in *Handbook of Economic Growth*, ed. Philippe Aghion and Steven Durlauf. Amsterdam: North-Holland.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114: 83–116.

Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-country Tests Using Alternative Institutional Measures." *Economics and Politics* 7, no. 3: 207–227.

Mankiw, N. G., D. Romer, and D. N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107: 407–437.

I

Institutions and Economic Performance

1 On Institutions and Growth

Philippe Aghion

1.1 Introduction

A main development in growth economics in the recent years has been to point to the fundamental role of institutions in the growth process, although few studies have led so far to precise policy recommendations beyond the general claims about the importance of property right enforcement. This is largely due to the difficulty of defining the term "institutions." North and Thomas (1973) developed the notion that "social infrastructure" reduces uncertainty and diminishes transaction costs. Some authors have emphasized the importance of property right protection and its impact on entrepreneurship; others have concentrated on regulatory institutions in financial, labor, or product markets, but never with a detailed modeling of how those institutions impact on the growth process, which could then be confronted to data.

In this chapter, I use the Aghion-Howitt model of growth with quality-improving innovations, to look at more specific aspects of the relationship between institutions, institutional change, and productivity growth.[1] The first section is devoted to the relationship between financial development and convergence, and argues that financial development is a main determinant of a country's ability to converge in growth rates and/or in levels of GDP per capita toward the technological frontier. Section 1.3 looks at the relationship between productivity growth and product market competition. It spells out the opposite effects that competition can have on innovation incentives in different types of sectors, and the implication for the overall effect of competition on productivity growth when one considers the economy as a whole. Section 1.4 develops the notion of appropriate institutions, showing how different types of institutions or policies maximize growth at different stages of technological development. Finally, section 1.5 discusses the role of academic institutions in the innovation process.

1.2 Financial Institutions and Convergence

The history of cross-country income differences exhibits mixed patterns of convergence and divergence. The most striking pattern over the long run is the "great divergence"—the dramatic widening of the distribution that has taken place since the early nineteenth century. Pritchett (1997) estimates that the proportional gap in living standards between the richest and poorest countries grew more than fivefold from 1870 to 1990, and according to the tables in Maddison 2001 the proportional gap between the richest group of countries and the poorest[2] grew from 3 in 1820 to 19 in 1998. But over the second half of the twentieth century, this widening seems to have stopped, at least among a large group of nations. In particular, the results of Barro and Sala-i-Martin (1992), Mankiw, Romer and Weil (1992), and Evans (1996) seem to imply that most countries are converging to parallel growth paths.

However, the recent pattern of convergence is not universal. In particular, the gap between the leading countries as a whole and the very poorest countries as a whole has continued to widen. The proportional gap in per capita income between Mayer-Foulkes's (2002) richest and poorest convergence groups grew by a factor of 2.6 between 1960 and 1995, and the proportional gap between Maddison's (2001) richest and poorest groups grew by a factor of 1.75 between 1950 and 1998. Thus as various authors[3] have observed, the history of income differences since the mid twentieth century has been one of "club-convergence"; that is, all rich and most middle-income countries seem to belong to one group, or "convergence club," with the same long-run growth rate, whereas all other countries seem to have diverse long-run growth rates, all strictly less than that of the convergence club. In this section, I develop an explanation for this phenomenon.

1.2.1 A Model of Technology Transfer

Consider one country in a world of h different countries. We assume that whenever an innovation takes place in any given sector in any country, the productivity parameter attached to the new product will match the global leading-edge technology. That is, let $\bar{A}_t$ be the maximum productivity parameter over all countries in the sector at the end of period t—in other words, the "frontier" productivity at t—and suppose that the frontier grows at a constant rate $\bar{g}$ that here we take as exogenous for simplicity.

Then domestic productivity in the sector evolves according to

$$A_t = \begin{cases} \bar{A}_t & \text{with probability } \mu \\ A_{t-1} & \text{with probability } 1-\mu \end{cases}, \tag{1}$$

where μ is the country's innovation rate. Let

$$a_t = \frac{A_t}{\bar{A}_t}$$

denote the country's proximity to the technological frontier. Then, it follows immediately from (1) that the distance variable a_t evolves over time according to

$$a_t = \mu + \frac{1-\mu}{1+g} a_{t-1}.$$

If $\mu > 0$, which in turn will depend upon underlying characteristics of the economy such as property right protection or the productivity of R&D, this difference equation has a unique fixed point

$$a^* = \frac{\mu(1+g)}{\mu+g}.$$

That is, as long as the country continues to perform R&D at a positive constant intensity, its distance to the frontier will stabilize, meaning that its productivity growth rate will converge to that of the global frontier. But if $\mu = 0$, the difference equation has no stable rest point and a_t diverges to zero. That is, if the country stops innovating it will have a long-run productivity growth rate of zero because innovation is a necessary condition for the country to benefit from technology transfer.

1.2.2 *The Role of Financial Development in Convergence*

The framework can be further developed by assuming that while the size of innovations increases with the distance to the technological frontier (due to technology transfer), the frequency of innovations depends upon the ratio between the distance to the technological frontier and the current stock of skilled workers. This enriched framework (see Howitt and Mayer-Foulkes 2002) can explain not only why some countries converge while other countries stagnate but also why different countries may display positive yet divergent growth patterns in the long run. Benhabib and Spiegel (2002) develop a similar account of divergence and show the importance of human capital in the process. The rest of this section presents a summary of the related model of Aghion, Howitt, and Mayer-Foulkes (2005) and discusses their empirical results showing the importance of financial development in the convergence process.

Suppose that the world is as portrayed in the previous section, but that research aimed at making an innovation in t must be done at period $t - 1$. If we assume perfectly functioning financial markets, then nothing much happens to the model except that the returns to research are discounted at a factor β to reflect the fact that the expected returns to R&D occur one period later than the expenditure.[4] But when credit markets are imperfect, Aghion, Howitt, and Mayer-Foulkes show that an entrepreneur may face a borrowing constraint that limits her investment to a fixed multiple (which we refer to as the credit multiplier) of her accumulated net wealth. In their model the multiple comes from the possibility that the borrower can, at a cost that is proportional to the size of her investment, decide to defraud her creditors by making arrangements to hide the proceeds of the R&D project in the event of success.[5] They also assume a two-period overlapping-generations structure in which the accumulated net wealth of an entrepreneur is her current wage income, and in which there is just one entrepreneur per sector in each country. This means that the further behind the frontier the country falls the less will any entrepreneur be able to invest in R&D relative to what is needed to maintain any given frequency of innovation. What happens in the long run to the country's growth rate depends upon the interaction between this disadvantage of backwardness, which reduces the frequency of innovations, and the advantage of backwardness described earlier, which increases the size of innovations. The lower the cost of defrauding a creditor the more likely it is that the disadvantage of backwardness will be the dominant force, preventing the country from converging to the frontier growth rate even in the long run. Generally speaking, the greater the degree of financial development of a country the more effective are the institutions and laws that make it difficult to defraud a creditor. Hence the link between financial development and the likelihood that a country will converge to the frontier growth rate.

More formally, the convergence equation becomes

$$a_t = \tilde{\mu}(a_{t-1}) + \frac{1 - \tilde{\mu}(a_{t-1})}{1+g} a_{t-1},$$

where $\tilde{\mu}(a_{t-1})$ is the innovation probability of credit-constrained firms in a country at proximity a_{t-1} from the technological frontier. That $\tilde{\mu}(a_{t-1})$ should increase with a_{t-1}, stems from the fact that the innovation cost is proportional to the frontier productivity (recall that innovations bring sectors all the way to the frontier), whereas the amount of wage resources firms in the country can use as a basis for borrowing are proportional to the country's current level of productivity. Therefore the further below the frontier a country currently is, the tighter credit constraints are on innovative firms. This, in turn, captures what we call the "disadvantage of backwardness." In addition, for given a_{t-1} the innovation probability $\tilde{\mu}(a_{t-1})$ increases with the borrowing/resource ratio (credit multiplier), which in turn increases with the cost of defrauding.

Aghion, Howitt, and Mayer-Foulkes (2005) test this effect of financial development on convergence by running the following cross-country growth regression:

$$g_i - g_1 = \beta_0 + \beta_f F_i + \beta_y \cdot (y_i - y_1) + \beta_{fy} \cdot F_i \cdot (y_i - y_1) + \beta_x X_i + \varepsilon_i \qquad (2)$$

where g_i denotes the average growth rate of per capita GDP in country i over the period 1960–1995, F_i the country's average level of financial development, y_i the initial (1960) log of per capita GDP, X_i a set of other regressors, and ε_i a disturbance term with mean zero. Country 1 is the technology leader, which Aghion, Howitt, and Mayer-Foulkes take to be the United States.

Define $\hat{y}_i \equiv y_i - y_1$, country i's initial relative per capita GDP. Under the assumption that $\beta_y + \beta_{fy} F_i \neq 0$, we can rewrite (2) as

$$g_i - g_1 = \lambda_i \cdot (\hat{y}_i - \hat{y}_i^*),$$

where the steady-state value $\hat{y}_i^*$ is defined by setting the right-hand side of (2) to zero as follows:

$$\hat{y}_i^* = -\frac{\beta_0 + \beta_f F_i + \beta_x X_i + \varepsilon_i}{\beta_y + \beta_{fy} F_i}, \qquad (3)$$

and λ_i is a country-specific convergence parameter

$$\lambda_i = \beta_y + \beta_{fy} F_i \tag{4}$$

that depends on financial development.

A country can converge to the frontier growth rate if and only if the growth rate of its relative per capita GDP depends negatively on the initial value $\hat{y}_i$; that is if and only if the convergence parameter λ_i is negative. Thus the likelihood of convergence will increase with financial development, as implied by the previous theory, if and only if

$$\beta_{fy} < 0. \tag{5}$$

The results of running this regression using a sample of seventy-one countries are shown in table 1 of Aghion, Howitt, and Mayer-Foulkes (2005), which indicates that the interaction coefficient β_{fy} is indeed significantly negative for a variety of different measures of financial development and a variety of different conditioning sets X. The estimation is by instrumental variables, using a country's legal origins, and its legal origins[6] interacted with the initial GDP gap $(y_i - y_1)$ as instruments for F_i and $F_i(y_i - y_1)$. The data, estimation methods, and choice of conditioning sets X are all taken directly from Levine, Loayza, and Beck (2000), who found a strongly positive and robust effect of financial intermediation on short-run growth in a regression identical to (2) but without the crucial interaction term $F_i(y_i - y_1)$ that allows convergence to depend upon the level of financial development. Aghion, Howitt, and Mayer-Foulkes (2005) show that these results are surprisingly robust to different estimation techniques, to discarding outliers, and to including possible interaction effects between the initial GDP gap and other right-hand-side variables.

1.2.3 Concluding Remark

Thus one sees how Schumpeterian growth theory and the quality improvement model can naturally explain club convergence patterns, the so-called twin peaks pointed out by Quah (1996). The Schumpeterian growth framework can deliver an explanation for cross-country differences in growth rates and/or in convergence patterns based upon *institutional considerations*. No one can deny that such considerations are close to what development economists have been concerned with. However, some may argue that the quality improvement paradigm, and new growth theories in general, remain of little help for development policy, that they merely formalize platitudes regarding the growth-enhancing nature of good property right protection, sound education systems, stable macroeconomy, without regard to specifics such as a country's current stage of development. In sections 1.3 and 1.4 we will argue on the contrary that the Schumpeterian growth paradigm can be used to understand (1) why liberalization policies (in particular an increase in product market competition) should affect productivity growth differently in sectors or countries at different stages of technological development as measured by the distance variable *a;* and (2) why the organizations or institutions that maximize growth, or that are actually chosen by societies, also vary with distance to the frontier.

1.3 Competition and Growth

One particularly unappealing feature of most existing endogenous growth models is the prediction that product market competition is unambiguously detrimental to growth because it reduces the monopoly rents that reward successful innovators and thereby discourages R&D investments. Not only does this prediction contradict a common wisdom that goes back to Adam Smith, but it has also been shown to be (partly) counterfactual (e.g., by Geroski (1995), Nickell (1996), and Blundell, Griffith, and Van Reenen (1999)).[7]

However, as I argue in this section, a simple modification reconciles the Schumpeterian paradigm with the evidence on product market competition and innovation, and also generates new empirical predictions that can be tested with firm- and industry-level data. In this respect, the paradigm can meet the challenge of seriously putting industrial organization into growth theory. The theory developed in this section is based on Aghion, Harris, and Vickers 1997 and on Aghion et al. 2001, but cast in the discrete-time framework introduced earlier.

As before, there is a global technological frontier that is common to all sectors, and that is drawn on by all innovations. The model takes as given the growth rate of this global frontier, so that the frontier $\bar{A}_t$ at the end of period t obeys

$$\bar{A}_t = \gamma \bar{A}_{t-1},$$

where $\gamma = 1 + \bar{g} > 1$.

In each country, the general good is produced using the same kind of technology as in the previous sections, but here for simplicity I assume a continuum of intermediate inputs, and I normalize the labor supply at $L = 1$, so that

$$y_t = \int_0^1 A_{it}^{1-\alpha} x_{it}^{\alpha}\, di,$$

where, in each sector i, only one firm produces intermediate input i using general good as capital according to a one-for-one technology.

In each sector, the incumbent firm faces a competitive fringe of firms that can produce the same kind of intermediate good, although at a higher unit cost. More specifically, we assume that at the end of period t, at unit cost χ, where we assume $1 < \chi < 1/\alpha < \gamma\chi$, a competitive fringe of firms can produce one unit of intermediate input i of a quality equal to $\min(A_{it}, \bar{A}_{t-1})$, where A_{it} is the productivity level achieved in sector i after innovation has had the opportunity to occur in sector i within period t.

In each period t, there are three types of sectors, which we refer to as type-j sectors, with $j \in \{0, 1, 2\}$. A type-j sector starts up at the beginning of period t with productivity $A_{j,t-1} = \bar{A}_{t-1-j}$, that is, j steps behind the current frontier $\bar{A}_{t-1}$. The profit flow of an incumbent firm in any sector at the end of period t, will depend upon the technological position of that firm with regard to the technological frontier at the end of the period.

Between the beginning and the end of the current period t, the incumbent firm in any sector i has the possibility to innovate with positive probability. Innovations occur step-by-step: in any sector an innovation moves productivity upward by the same factor γ. Incumbent firms can affect the probability of an innovation by investing more in R&D at the beginning of the period. Namely, by investing the quadratic R&D effort $\frac{1}{2}\gamma A_{i,t-1}\mu^2$ an incumbent firm i in a type-0 or type-1 sector, innovates with probability μ.[8] However, innovation is assumed to be automatic in type-2 sectors, which in turn reflects a knowledge externality from more advanced sectors that limits the maximum distance of any sector to the technological frontier.

Now, consider the R&D incentives of incumbent firms in the different types of sectors at the beginning of period t. Firms in type-2 sectors have no incentive to invest in R&D since innovation is automatic in such sectors. Thus

$$\mu_2 = 0,$$

where μ_j is the equilibrium R&D choice in sector j.

Firms in type-1 sectors, which start one step behind the current frontier at $A_{i,t-1} = \bar{A}_{t-2}$ at the beginning of period t, end up with productivity $A_t = \bar{A}_{t-1}$ if they successfully innovate, and with productivity $A_t = \bar{A}_{t-2}$ otherwise. In either case, the competitive fringe can produce intermediate goods of the same quality but at cost χ instead of 1. Then, as in Acemoglu et al. 2003, the equilibrium profit can be shown to be equal to[9]

$$\pi_t = A_t \delta(\chi),$$

with

$$\delta(\chi) = (\chi - 1)(\chi/\alpha)^{1/(\alpha-1)}$$

increasing in χ.

Thus the net rent from innovating for a type-1 firm is equal to

$$(\bar{A}_{t-1} - \bar{A}_{t-2})\delta(\chi)$$

and therefore a type-1 firm will choose its R&D effort to solve

$$\max_{\mu} \left\{ (\bar{A}_{t-1} - \bar{A}_{t-2})\delta(\chi)\mu - \frac{1}{2}\gamma \bar{A}_{t-2}\mu^2 \right\},$$

which yields

$$\mu_1 = \left(1 - \frac{1}{\gamma}\right)\delta(\chi).$$

In particular, an increase in product market competition, measured as a reduction in the unit cost χ of the competitive fringe, will reduce the innovation incentives of a type-1 firm. This I refer to as the *Schumpeterian effect* of product market competition: competition reduces innovation incentives and therefore productivity growth by reducing the rents from innovations of type-1 firms that start below the technological frontier. This is the dominant effect, both in IO models of product differentiation and entry, and in basic endogenous growth models as the one analyzed in the previous sections. Note that type-1 firms cannot escape the fringe by innovating: whether they innovate or not, these firms face competitors that can produce the same quality as theirs at cost χ. As we shall now see, things become different in the case of type-0 firms.

Firms in type-0 sectors, that start at the current frontier, end up with productivity $\bar{A}_t$ if they innovate, and stay with their initial productivity $\bar{A}_{t-1}$ if they do not. But the competitive fringe can never get beyond producing quality $\bar{A}_{t-1}$. Thus, by innovating, a type-0 incumbent firm produces an intermediate good which is γ times better than the competing good the fringe could produce, and at unit cost 1 instead of χ for the fringe. Our assumption $\frac{1}{\alpha} < \gamma\chi$ then implies that competition by the fringe is no longer a binding constraint for an innovating incumbent, so that its equilibrium profit post-innovation, will simply be the profit of an unconstrained monopolist, namely,

$$\pi_t = \bar{A}_t \delta(1/\alpha).$$

On the other hand, a type-0 firm that does not innovate, will keep its productivity equal to $\bar{A}_{t-1}$. Since the competitive fringe can produce up to this quality level at cost χ, the equilibrium profit of a type-0 firm that does not innovate, is equal to

$$\pi_t = \bar{A}_{t-1}\delta(\chi).$$

A type-0 firm will then choose its R&D effort to

$$\max_{\mu}\left\{\left[\bar{A}_t\delta(1/\alpha) - \bar{A}_{t-1}\delta(\chi)\right]\mu - \frac{1}{2}\gamma\bar{A}_{t-1}\mu^2\right\},$$

so that in equilibrium

$$\mu_0 = \delta(1/\alpha) - \frac{1}{\gamma}\delta(\chi).$$

In particular, an increase in product market competition, namely, a reduction in χ, will now have a fostering effect on R&D and innovation. This, we refer to as the *escape competition effect*: competition reduces pre-innovation rents of type-0 incumbent firms, but not their post-innovation rents since by innovating these firms have escaped the fringe. This in turn induces those firms to innovate in order to escape competition with the fringe.

1.3.1 Composition Effect and the Inverted-U Relationship between Competition and Innovation

I have just shown that product market competition tends to have opposite effects on frontier and lagging sectors, fostering innovation by the former and discouraging innovation by the latter. In this section, I consider the impact of competition on the steady-state aggregate innovation intensity

$$I = q_0\mu_0 + q_1\mu_1, \tag{6}$$

where q_j is the steady-state fraction of type-j sectors (recall that type-2 sectors do not perform R&D).

To get a nontrivial steady-state fraction of type-0 firms, we need that the net flows out of state 0 (which corresponds to type-0 firms that fail to innovate in the current period) be compensated by a net flow into state 0. I simply postulate such a flow into state 0, by assuming that at the end of any period t, with exogenous probability ε entry at the new frontier, that is by a type-0 firm with productivity level $\bar{A}_t$, occurs in a type-2 sector after the incumbent firm has produced. I then have the following flow equations describing the net flows into and out of states 0, 1, and 2:

$$q_2\varepsilon = q_0(1 - \mu_0);$$

$$q_0(1 - \mu_0) = q_1(1 - \mu_1);$$

$$q_1(1 - \mu_1) = q_2\varepsilon;$$

in which the left-hand sides represent the steady-state expected flow of sectors that move into a state j and the right-hand sides represent the expected outflow from the same state, for $j = 0, 1,$ and 2. This, together with the identity

$$q_0 + q_1 + q_2 = 1,$$

implies that

$$I = 1 - q_2(1 + 2\varepsilon),$$

where

$$q_2 = \frac{1}{1 + \dfrac{\varepsilon}{1 - \mu_0} + \dfrac{\varepsilon}{1 - \mu_1}}.$$

In particular, one can see that the overall effect of increased product market competition on I is ambiguous since it produces opposite effects on innovation probabilities in type-0 and type-1 sectors (i.e., on μ_0 and μ_1). In fact, one can say more than that, and show that (1) the

Schumpeterian effect always dominates for γ sufficiently large; (2) the escape competition effect always dominates for γ sufficiently close to one; (3) for intermediate values of γ, the escape competition effect dominates when competition is initially low (with χ close to $1/\alpha$) whereas the Schumpeterian effect dominates when competition is initially high (with χ close to one). In this latter case, the relationship between competition and innovation is inverted-U shaped.

This inverted-U pattern can be explained as follows: at low initial levels of competition (i.e., high initial levels of $\delta(\chi)$), type-1 firms have strong reason to innovate; it follows that many intermediate sectors in the economy will end up being type-0 firms in steady state (this we refer to as the *composition effect* of competition on the relative equilibrium fractions of type-0 and type-1); but then the dominant effect of competition on innovation is the escape competition effect whereby more competition fosters innovation by type-0 firms. On the other hand, at high initial levels of competition, innovation incentives in type-1 sectors are so low that a sector will remain of type-1 for a long time, and therefore many sectors will end up being of type-1 in steady state, which in turn implies that the negative Schumpeterian appropriability effect of competition on innovation should tend to dominate in that case.

1.3.2 Empirical Predictions

The preceding analysis generates several interesting predictions:

1. Innovation in sectors in which firms are close to the technology frontier react positively to an increase in product market competition.

2. Innovation reacts less positively, or negatively, in sectors in which firms are further below the technological frontier.

3. The average fraction of frontier sectors decreases, namely, the average technological gap between incumbent firms and the frontier in their respective sectors increases, when competition increases.

4. The overall effect of competition on aggregate innovation, is inverted-U shaped.[10]

These predictions have been confronted by Aghion, Bloom, Blundell, Griffith, and Howitt (2002) with U.K. firm-level data on competition and patenting, and I briefly summarize their findings. The prediction I want to emphasize here is that the escape competition effect should be strongest in industries in which firms are closest to the technological frontier.

Aghion et al. (2002) consider a U.K. panel of individual companies during the period 1968–1997. This panel includes all companies quoted on the London Stock Exchange over that period, and whose names begin with a letter from A to L. To compute competition measures, the study uses firm-level accounting data from Datastream; product market competition is in turn measured by one minus the Lerner index (ratio of operating profits minus financial costs over sales), controlling for capital depreciation, advertising expenditures, and firm size. Furthermore, to control for the possibility that variations in the Lerner index be mostly due to variations in fixed costs, we use policy instruments such as the implementation of the Single Market Program (SMP) or lagged values of the Lerner index as instrumental variables. Innovation activities, in turn, are measured both by the number of patents weighted by citations and by R&D spending. Patenting information comes from the U.S. Patent Office where most firms that engage in international trade register their patents; in particular, this includes 461 companies on the London Stock Exchange with names starting by A to L, for which we already had detailed accounting data. Finally, technological frontier is measured as follows: suppose a U.K. firm (call it i) belongs to some industry A; then we measure technological distance by the difference between the maximum total factor productivity (TFP) in industry A across all OECD countries (we call it TFP_{F}, where the subscript "F" refers to the technological frontier) and the TFP of the U.K. firm, divided by the former:

$$m_i = \frac{TFP_{\mathrm{F}} - TFP_i}{TFP_{\mathrm{F}}}.$$

Aghion et al. (2002) find that the effect of product market competition on innovation is all the more positive that firms are closer to the technological frontier (or equivalently are more "neck-and-neck"). Another interesting finding is that the Schumpeterian effect is also at work, and that it dominates at high initial levels of product market competition. This in turn reflects the "composition effect" pointed out in section 1.3.1: namely, as competition increases and neck-and-neck firms therefore engage in more intense innovation to escape competition, the equilibrium fraction of neck-and-neck industries tends to decrease (equivalently, any individual firm spends less time in neck-and-neck competition with its main rivals) and therefore the average impact of the escape competition effect decreases at the expense of the counteracting Schumpeterian effect. The paper indeed shows that the average distance to the technological frontier increases with the degree of product market competition. The Schumpeterian effect was missed by previous empirical studies, mainly as a result of their being confined to linear estimations. Instead, more in line with the Poisson technology that governs the arrival of innovations, both in Schumpeterian and in patent race models, Aghion et al. (2002) use a semi-parametric estimation method in which the expected flow of innovations is a piecewise polynomial function of the Lerner index.

1.4 Appropriate Institutions

1.4.1 From Schumpeter to Gerschenkron

By linking growth to innovation and entrepreneurship, and innovation incentives in turn to characteristics of the economic environment, new growth theories made it possible to analyze the interplay between growth and the design of policies and institutions. For example, the basic model developed in section 1.2 suggested that long-run growth would be best enhanced by a combination of good property right protection (to protect the rents of innovators against imitation), a good education system (to increase the efficiency of R&D activities and/or the supply of skilled manufacturing labor), and a stable macroeconomy to reduce interest rates (and thereby increase the net present value of innovative rents). Our discussion of convergence clubs in section 1.3 then suggested that the same policies or institutions would also increase a country's ability to join the convergence club.

Now, new growth theories may be criticized by development economists and policymakers, precisely because of the universal nature of the policy recommendations that appear to follow from them: no matter how developed a country or sector currently is, it seems that one should prescribe the same medicines (legal reform to enforce property rights, investment climate favorable to entrepreneurship, education, macrostability, etc.) to maximize the growth prospects of that country or sector.

However, in his essay "Economic Backwardness in Historical Perspective," Gerschenkron (1962) argues that relatively backward economies could more rapidly catch up with more advanced countries by introducing "appropriate institutions" that are growth-enhancing at an early stage of development but may cease to be so at a later stage. Thus, countries like Japan or Korea managed to achieve very high growth rates from 1945 up until the 1990s with institutional arrangements involving long-term relationships between firms and banks, the predominance of large conglomerates, and strong government intervention through export promotion and subsidized loans to the enterprise sector, all of which depart from the more market-based and laissez-faire institutional model pioneered and promoted by the United States.

That growth-enhancing institutions or policies might change with a country's or sector's distance to the technological frontier, should not come as a total surprise to our readers at this point: in section 1.3.2, we saw that competition could have opposite effects on innovation incentives depending on whether firms were initially closer to or farther below the fringe in the corresponding industry (it would enhance innovations in neck-and-neck industries, and discourage it in industries where innovating firms are far below the frontier). The same type of conclusion turns out to hold true when one looks at the interplay between countries' distance to the world technology frontier and "openness." Using a cross-country panel of more than one hundred countries over the 1960–2000 period, Acemoglu, Aghion, and Zilibotti (2002) regress the average growth rate over a five-year period on a country's distance to the U.S. frontier (measured by the ratio of GDP per capita in that country to per capita GDP in the United States) at the beginning of the period. Then, splitting the sample of countries in two groups, corresponding respectively to a high and a low openness group according to Frankel-Romer's openness indicator, Acemoglu, Aghion, and Zilibotti show that average growth decreases more rapidly as a country approaches the world frontier when openness is low. Thus, while a low degree of openness does not appear to be detrimental to growth in countries far below the world frontier, it becomes increasingly detrimental to growth as the country approaches the frontier. Acemoglu, Aghion, and Zilibotti repeat the same exercise using entry costs to new firms (measured as in Djankov et al. 2001) instead of openness, and they obtain a similar conclusion, namely, that high entry costs are most damaging to growth when a country is close to the world frontier, unlike in countries far below the frontier.

In this section, I argue that Gerschenkron's idea of "appropriate institutions" can be easily embedded into a growth framework, in a way that can help substantiate the following claims:

1. different institution or policy design affects productivity growth differently depending upon a country's or sector's distance to the technological frontier;

2. a country's distance to the technological frontier affects the type of organizations we observe in this country (e.g., bank versus market finance, vertical integration versus outsourcing, etc.).

The remaining part of section 1.4 is organized as follows. I first describe the growth equation that Acemoglu, Aghion, and Zilibotti (2002) introduce to embed the notion of "appropriate institutions" into the above growth framework. I then focus on the first question about the effects of institution design on productivity growth, by concentrating on the relationship between growth and the organization of education. Finally, I briefly discuss the effects of distance on equilibrium institutions in a concluding subsection.

1.4.2 A Simple Model of Appropriate Institutions

Consider the following variant of the multi-country growth model of section 1.3. In each country, a unique general good that also serves as numéraire is produced competitively using a continuum of intermediate inputs according to

$$y_t = \int_0^1 (A_t(i))^{1-\alpha} x_t(i)^\alpha \, di, \tag{7}$$

where $A_t(i)$ is the productivity in sector i at time t, $x_t(i)$ is the flow of intermediate good i used in general good production again at time t, and $\alpha \in [0, 1]$.

As before, ex post each intermediate good producer faces a competitive fringe of imitators that forces her to charge a limit price $p_t(i) = \chi > 1$. Consequently, equilibrium monopoly profits (gross of the fixed cost) are simply given by

$$\pi_t(i) = \delta A_t(i),$$

where $\delta \equiv (\chi - 1)\chi^{-1/(1-\alpha)}$.

We still let

$$A_t \equiv \int_0^1 A_t(i) \, di$$

denote the average productivity in the country at date t, $\bar{A}_t$ the productivity at the world frontier that we assume to grow at the constant rate g from one period to the next, and $a_t = A_t/\bar{A}_t$ the (inverse) measure of the country's distance to the technological frontier at date t.

The main departure from the convergence model in section 1.3, lies in the equation for productivity growth. Suppose that intermediate firms have two ways to generate productivity growth: (1) they can imitate existing world frontier technologies; (2) they can innovate upon the previous local technology. More specifically, we assume

$$A_t(i) = \eta\bar{A}_{t-1} + \gamma A_{t-1}, \tag{8}$$

where $\eta\bar{A}_{t-1}$ and γA_{t-1} refer respectively to the imitation and innovation components of productivity growth. Imitations use the existing frontier technology at the end of period $(t-1)$, thus they multiply $\bar{A}_{t-1}$, whereas innovations build on the knowledge stock of the country, and therefore they multiply A_{t-1}.

Now dividing both sides of (8) by $\bar{A}_t$, using the fact that

$$\bar{A}_t = (1+g)\bar{A}_{t-1},$$

and integrating over all intermediate sectors i, we immediately obtain the following linear relationship between the country's distance to frontier a_t at date t and the distance to frontier a_{t-1} at date $t-1$:

$$a_t = \frac{1}{1+g}(\eta + \gamma a_{t-1}). \tag{9}$$

This equation clearly shows that the relative importance of innovation for productivity growth, increases as (1) the country moves closer to the world technological frontier, namely, as a_{t-1} moves closer to 1, whereas imitation is more important when the country is far below the frontier, namely, when a_{t-1} is close to zero; (2) a new technological revolution (e.g., the ITC revolution) occurs that increases the importance of innovation, namely, increases γ.

This immediately generates a theory of "appropriate institutions" and growth: suppose that imitation and innovation activities do not require the same institutions. Typically, imitation activities (i.e., η in equation (9)) will be enhanced by long-term investments within (large) existing firms, which in turn may benefit from long-term bank finance and/or subsidized credit as in Japan or Korea since 1945. On the other hand, innovation activities (i.e., γ) require initiative, risk taking, and also the selection of good projects and talents and the weeding out of projects that turn out not to be profitable. This in turn calls for more market-based and flexible institutions, in particular for a higher reliance on market finance and speculative monitoring, higher competition and trade liberalization to weed out the bad projects, more flexible labor markets for firms to select the most talented or best matched employees, non-integrated firms to increase initiative and entrepreneurship downstream, and so forth. It then follows from equation (9) that the growth-maximizing institutions will evolve as a country moves toward the world technological frontier. Far below the frontier, a country will grow faster if it adopts what Acemoglu, Aghion, and Zilibotti (2002) refer to as *investment-based* institutions or policies, whereas closer to the frontier growth will be maximized if the country switches to *innovation-based* institutions or policies.

A natural question is of course whether institutions actually change when they should from a growth- (or welfare-) maximizing point of view; in other words, how do equilibrium institutions at all stages of development compare with the growth-maximizing institutions? This question is addressed in detail in Acemoglu, Aghion, and Zilibotti 2002, and we will come back to it briefly in section 1.6.

1.4.3 Appropriate Education Systems

In his seminal paper on economic development, Lucas (1988) emphasized the *accumulation* of human capital as a main engine of growth; thus, according to the analysis in that paper, cross-country differences in growth rates across countries would be primarily attributable to differences in *rates of accumulation* of human capital. An alternative approach, pioneered by Nelson and Phelps (1966), revived by the Schumpeterian growth literature,[11] would instead emphasize the combined effect of the *stock* of human capital and of the innovation process in generating long-run growth and fostering convergence. In this alternative approach, differences in growth rates across countries would be mainly attributable to differences in *stocks* of human capital, as those condition countries′ ability to innovate or to adapt to new technologies and thereby catch up with the world technological frontier. Thus, in the basic model of section 1.2, the equilibrium R&D investment and therefore the steady-state growth rate were shown to be increasing in the aggregate supply of (skilled) labor L and in the productivity of research λ, both of which refer more to the *stock* and *efficiency* of human capital than to its rate of accumulation.

Now, whichever approach one takes, and the evidence so far supports the two approaches as being somewhat complementary, once again one may worry about growth models delivering too general a message, namely that more education is always growth enhancing. In this section, I try to go one step further and argue that the Acemoglu, Aghion, and Zilibotti specification (summarized by equation (8)), can be used to analyze the effects, not only of the total *amount* of education, but more importantly of the *organization* of education, on growth in countries at different stages of development.

This section, which is based on Aghion, Meghir, and Vandenbussche 2003, focuses on one particular aspect of the organization of education systems, namely the mix of primary, secondary, and higher education. We consider a variant of the Acemoglu, Aghion, and Zilibotti model outlined in section 1.4.2, in which innovation requires highly educated labor, whereas imitation can be performed by both highly educated and lower-skill workers.

A main prediction emerging from this model is that the closer a country gets to the world technology frontier, the more growth-enhancing it becomes to invest in higher education. The intuition follows directly from the Rybczynski theorem in international trade. Stated in the context of a two-sector-two-input economy, this theorem says that an increase in the supply of input in the sector that uses that input more intensively should increase "output" in that sector more than proportionally. To transpose this result to the context of our model, consider the effect of an increase in the supply of skilled labor, keeping the supply of unskilled labor fixed and for given a. Given that skilled workers contribute relatively more to productivity growth and profits if employed in innovation rather than in imitation, the demand for additional skilled labor will tend to be higher in innovation. But then the marginal productivity of unskilled labor should also increase more in innovation than in imitation, hence a net flow of unskilled workers should also move from imitation into innovation. This, in turn, will enhance further the marginal productivity of skilled labor in innovation, thereby inducing an ever greater fraction of skilled labor to move to innovation. Now the closer the country is to the technology frontier (i.e., the higher a), the stronger this Rybszynski effect as a higher a increases the efficiency of both skilled and unskilled labor in innovation relative to imitation. A second, reinforcing, reason is that an increase in the fraction of skilled labor reduces the amount of unskilled labor available in the economy, hence reducing the marginal productivity of skilled labor in imitation, all the more the closer the country is to the frontier.

Aghion, Meghir, and Vandenbussche (2003) then confront this prediction with cross-country evidence on higher education, distance to frontier, and productivity growth. The prediction that higher education has stronger growth-enhancing effects close to the technological frontier can be tested using cross-regional or cross-country data. Thus Aghion, Meghir, and Vandenbussche consider a panel dataset of nineteen OECD countries over the period 1960–2000. Output and investment data are drawn from Penn World Tables 6.1 (2002) and human capital data from Barro-Lee (2000). The Barro-Lee data indicate the fraction of a country's population that has reached a certain level of schooling at intervals of five years, so Aghion, Meghir, and Vandenbussche use the fraction that has received some higher education together with their measure of TFP (itself constructed assuming a constant labor share of 0.7 across countries) to perform the following regression:

$$g_{j,t} = \alpha_{0,j} + \alpha_1 dist_{j,t-1} + \alpha_2 \Lambda_{j,t-1} + \alpha_3 (dist_{j,t-1} * \Lambda_{j,t-1}) + u_{j,t},$$

where $g_{j,t}$ is country j's growth rate over a five-year period, $dist_{j,t-1}$ is country j's closeness to the technological frontier at $t-1$ (i.e., 5 years before), $\Lambda_{j,t-1}$ is the fraction of the working age population with some higher education in the previous period and $\alpha_{0,j}$ is a country dummy controlling for country fixed effects. The closeness variable is instrumented with its lagged value at $t-2$, the fraction variable is instrumented using expenditure on tertiary education per capita lagged by two periods, and the interaction term is instrumented using the interaction between the two instruments for closeness and for the fraction variables. Finally, the standard errors we report allow for serial correlation and heteroskedasticity.

Aghion, Meghir, and Vandenbussche (2003) find a positive and significant interaction between our education measure and closeness to the frontier, as predicted by the theory in the previous section 1.4.2. This result demonstrates that it is more important to expand years of higher education close to the technological frontier.

1.5 Academic Institutions

1.5.1 Introduction

In the preceding sections, I have emphasized the role of R&D undertaken by private firms in the growth process. There exists also a different type of organization that also undertakes research, namely, academic institutions. What distinguishes *academic research* from *private research*? A common view is that academic research is more "basic" than private-sector research. However, the notion of "basicness" is not clearly spelled out in general, and moreover, a recent National Science Foundation (NSF) survey finds that more than 22 percent of all basic research (as defined by the NSF) in the United States during the period 1993–1997 was performed by private enterprises. A second approach emphasizes appropriability problems that are supposedly more acute at earlier stages of a research project; this, in turn, would provide a rationale for early stages to be performed within academic institutions in order to reduce the scope for underinvestment. Yet private firms outsource research from universities, even on projects that involve a high degree of knowledge spillovers; more fundamentally, if it was just a problem of appropriability and spillovers, then the state could simply subsidize basic research, and there would be no need for new organizations other than private firms.

This question is closely related to the design of the right policies to promote innovation. In particular, is the Bayh-Dole Act (passed in the early 1980s in the United States to allow universities to patent the outcome of federally funded research) a good idea? While Bayh-Dole can be understood as a response to lack of appropriability, is there something as "too much patent protection for basic research"?

There exists already a whole literature on academic research and the so-called republic of science. For example, Nelson (1959) and Arrow (1962) emphasize the appropriability problem (certain kinds of ideas cannot be fully appropriated by those who develop them). More recently, Carmichael (1988) and Dasgupta and David (1994) have emphasized differences in objective functions and incentives systems between academia and the private sector; however, why not simply change the incentive system within private-sector firms to increase overall efficiency of the innovation process? Finally, recent empirical literature, for example, Henderson, Jaffe, and Trajtenberg 1998, Murray and Stern 2004, and Lach and Schankerman 2004, analyze the effects of Bayh-Dole on flow and the importance of university patents, and question the existence of an "anti-commons effect" of IPR protection.

Departing from these attempts, Aghion, Dewatripont, and Stein (2005), develop a new approach of the role of academia in the innovation process that uses quality ladders and also focuses on control rights. More specifically, they focus on academia as a *commitment to leave control rights* to researchers. Moreover, they model research as a multistage process (see also Hellmann and Perotti 2004) whereby an initial idea goes through successive improvements before becoming commercializable. The basic trade-off Aghion, Dewatripont, and Stein emphasize between academia and private-sector research can be summarized as follows: an academic researcher may pursue commercially useless research (this is the cost of delegating control rights to the scientist), but on the other hand the academic researcher accepts lower wages compared to private-sector researchers because of the private satisfaction it obtains from research itself. The key insight in their paper is that when research is multi-stage, academia has comparative advantage in earlier stages, whereas private research has a comparative advantage in later stages of the research process.

I now present the mechanics of the Aghion, Dewatripont, and Stein model.

1.5.2 *Basic Framework*

The modeling strategy is to keep at a minimum the differences between academia and private sector, profit maximizing firms. An economically viable product (e.g., a new drug) starts with an idea I_0 that can be built upon by researchers, leading to ideas $I_1, I_2, \ldots,$ until idea I_k that generates economic value V. For each of the k stages, one researcher can work on the idea. In academic institutions, the scientist is *free* to pursue his own strategy. A *practical* strategy yields probability p of being successful, namely, of moving to the next stage, whereas the *alternative* strategy yields probability 0 of moving to the next stage. With probability α, a scientist will have zero disutility for the practical strategy, but with probability $(1-\alpha)$ she has disutility z from the practical strategy. Ex ante, the scientist does not know what her preferences will be, that is, what type of research she would like to undertake.

In the private sector, the scientist's boss can *direct* her research, and hence will impose the practical strategy. This raises the probability of success from αp to p, but it also imposes an ex ante disutility of $(1-\alpha)z$ on the scientist. If researchers have outside option R, then academic wages will be

$$w_a = R,$$

whereas private-sector research wages will be

$$w_p = R + (1-\alpha)z.$$

Now let us start from the last stage k and solve the model by backward induction. If managed by the private sector, the expected value of the research line as of the beginning of stage k, is equal to

$$P_k = pV - w_p = pV - (R + (1-\alpha)z).$$

If managed by academia, the expected value of the research line as of the beginning of stage k, is equal to

$$A_k = \alpha pV - w_a = \alpha pV - R.$$

Therefore, private-sector research dominates academic research in stage k whenever $P_k > A_k$, that is, if $pV > z$. Let the maximum expected profitability at stage k be denoted by Π_k, so that

$$\Pi_k = \max\{P_k, A_k\}$$

Consider stage $(k-1)$. The value of the line as of the beginning of stage $(k-1)$ if that stage is managed by the private sector, is equal to $P_{k-1} = p\Pi_k - w_p$, whereas the value of the line as of the beginning of stage $(k-1)$ if that stage is managed in academia, is equal to $A_k = \alpha p\Pi_k - w_a$. Now, let us move back to stage $i < k-1$. At that stage, the private sector generates payoff

$$P_i = p\Pi_{i+1} - w_p,$$

whereas academia generates payoff

$$A_i = \alpha p\Pi_{i+1} - w_a,$$

where

$$\Pi_{i+1} = \max\{P_{i+1}, A_{i+1}\}.$$

Therefore private-sector management dominates academia whenever

$$P_i > A_i \Leftrightarrow p\Pi_{i+1} > z.$$

This simple setup has interesting implications. First, since Π_{i+1} increases in i, it immediately follows that academic institutions tend to dominate private-sector research in earlier stages, that is for lower values of i. Second, for k sufficiently large, the line is not viable if entirely managed by the private sector. Indeed the value of a line entirely managed by the private sector as of stage 1, is equal to

$$\Pi_1 = p^k V - (1 + p + \cdots + p^{k-1}) w_p,$$

which becomes obviously negative when k goes to infinity and $z > 0$. In fact, academia is viable at an earlier stage than private sector if

$$p\Pi_i - w_p < \alpha p \Pi_i - w_a = 0$$

for some i, which in turn requires that

$$w_a = R < \alpha z.$$

More generally, it can be shown that there exists a unique cutoff point i^* such that it is socially optimal that research be done in academic institutions if $i < i^*$, and by the private sector if $i > i^*$. Moreover, the cutoff point i^* is decreasing in V, and increasing in α and z.

This model can be extended in several interesting directions. For example, it can be used to analyze the role for hybrid organizations, such as private firms that grant some academic freedom to their researchers, or academic institutions that introduce high-powered incentives, for example, in the form of patent rights. But most importantly it highlights that an economy wishing to innovate requires an institutional setup that allows academic institutions, as well as private firms, to undertake research.

1.6 Conclusion

In this chapter I argued that the endogenous growth model with quality-improving innovations provides a framework for taking a closer look at the relationship among institutions, institutional change, and economic growth.

Far from closing the field, the chapter suggests many avenues for future research. For example, on growth and convergence, more research remains to be done to identify the main determinants of cross-country convergence and divergence.[12] Also important is the need to analyze the role of international intellectual property right protections and foreign direct investment in preventing or favoring convergence. On growth and industrial organization, I have restricted attention to product market competition among existing firms. But what can I say about entry and its impact on incumbents' innovation activities?[13] On institutions, I have just touched upon the question of how technical change interacts with organizational change. Do countries or firms/sectors actually get stuck in institutional traps of the kind described in section 1.4? What enables such traps to disappear over time? How do political economy considerations interact with this process?

If I had to select just three topics for further thinking on the role of institutions and policy in the growth process using the Aghion and Howitt framework, I would suggest the following. First, consider the role of basic science in generating (very) long-term growth. Do fundamental innovations (or the so-called general-purpose technologies) require the same incentive system and the same rewards as industrial innovations? How can one design incentive systems in universities so that university research would best complement private research? A second aspect is the interplay between growth and volatility. Is R&D and innovation procyclical or countercyclical, and is macroeconomic volatility always detrimental to innovation and growth? Answering this question in turn opens up a whole new research topic on the macropolicy of growth.[14] A third aspect is the extent to which our growth paradigm can be applied to less developed economies. In particular, can we use the new growth approach developed in this chapter to revisit the important issue of poverty reduction?[15] All these exciting questions are left for future research.

Notes

This chapter draws unrestrainedly from a joint chapter with Peter Howitt in the forthcoming *Handbook of Economic Growth* (Aghion and Durlauf 2005).

1. The chapter is close to Aghion and Howitt 2005, which examines the implications of the Schumpeterian growth model. Here I emphasize the institutional aspect and introduce an additional mechanisms through which institutions may impact innovation and growth.

2. The richest group was Western Europe in 1820 and the "Euopean Offshoots" (Australia, Canada, New Zealand, and the United States) in 1998. The poorest group was Africa in both years.

3. Baumol (1986), Durlauf and Johnson (1995), Quah (1993, 1997), and Mayer-Foulkes (2002, 2003).

4. For simplicity, I suppose that everyone has linear intertemporal preferences with a constant discount factor β.

5. The "credit multipler" assumed here is much like that of Bernanke and Gertler (1989), as modified by Aghion, Banerjee, and Piketty (1999).

6. See LaPorta et al. 1998 for a detailed explanation of legal origins and its relevance as an instrument for financial development.

7. I refer the reader to section 1.3.2 where I confront theory and empirics on the relationship between competition/entry and innovation/productivity growth.

8. We thus depart slightly from our formulation in the previous sections: here we take the probability of innovation, not the R&D effort, as the optimization variable. However the two formulations are equivalent: that the innovation probability $f(n) = \mu$ is a concave function of the effort n, is equivalent to saying that the effort is a convex function of the probability.

9. Imitation does not destroy the rents of non-innovating firms. I assume nevertheless that the firm ignores any continuation value in its R&D decision.

10. Although perhaps only the second part of the inverse U will be observable.

11. For example, see Acemoglu 1996, Acemoglu 2002, Aghion, Howitt, and Violante 2002, and Aghion 2002.

12. In Aghion, Howitt, and Mayer-Foulkes 2005, we emphasize the role of credit constraints in R&D as a distinguishing factor between the countries that converge in growth rates and in levels toward the frontier, those that converge only in growth rates, and those that follow a divergent path toward a lower rate of long-run growth. Whether credit constraints or other factors such as health, education, and property rights protection, are key to this threefold classification remains an open question.

13. See Aghion et al. 2004 and Aghion et al. 2005 for preliminary work on entry and growth.

14. See Aghion, Angeletos, Banerjee, and Manova 2004.

15. See Aghion and Armendáriz de Aghion 2004 for some preliminary thoughts on this aspect.

References

Acemoglu, Daron. 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *Quarterly Journal of Economics* 111: 779–804.

Acemoglu, Daron. 2002. "Directed Technical Change." *Review of Economic Studies* 69: 781–810.

Acemoglu, D., P. Aghion, R. Griffith, and F. Zilibotti. 2003. "Technology, Hold-up, and Vertical Integration: What Do We Learn From Micro Data?" Mimeo., IFS, London.

Acemoglu, D., P. Aghion, and F. Zilibotti. 2002. "Distance to Frontier, Selection, and Economic Growth." NBER working paper 9066.

Aghion, P. 2002. "Schumpeterian Growth Theory and the Dynamics of Income Inequality." *Econometrica* 70, no. 3: 855–882.

Aghion, P., A. V. Banerjee, and T. Piketty. 1999. "Dualism and Macroeconomic Volatility." *Quarterly Journal of Economics* 114: 1359–1397.

Aghion, P., B. Bloom, R. Blundell, R. Griffith, and P. Howitt. 2002. "Competition and Innovation: An Inverted-U Relationship." NBER working paper 9269.

Aghion, P., R. Blundell, R. Griffith, P. Howitt, and S. Prantl. 2004. "Entry and Productivity Growth: Evidence from Micro-level Panel Data." *Journal of the European Economic Association* 2, nos. 2–3: 265–276.

Aghion, P., R. Burgess, S. Redding, and F. Zilibotti. 2005. "Entry Liberalization and Inequality in Industrial Performance" (joint with Philippe Aghion, Robin Burgess and Fabrizio Zilibotti). *Journal of the European Economic Association* 3, nos. 2–3: 291–302.

Aghion, P., M. Dewatripont, and J. Stein. 2005. "Academic Freedom, Private-sector Focus, and the Process of Innovation." Mimeo., Harvard University.

Aghion, P., and S. Durlauf. 2005. *Handbook of Economic Growth*. Amsterdam: North-Holland.

Aghion, P., C. Harris, P. Howitt, and J. Vickers. 2001. "Competition, Imitation, and Growth with Step-by-step Innovation." *Review of Economic Studies* 68: 467–492.

Aghion, P., and P. Howitt. 2005. "Growth with Quality-improving Innovations: An Integrated Framework." Forthcoming in *Handbook of Economic Growth*, ed. P. Aghion and S. Durlauf. Amsterdam: North-Holland.

Aghion, P., P. Howitt, and D. Mayer-Foulkes. 2005. "The Effect of Financial Development on Convergence: Theory and Evidence." *Quarterly Journal of Economics* 120, no. 1: 173–222.

Aghion, P., P. Howitt, and G. L. Violante. 2002. "General Purpose Technology and Wage Inequality." *Journal of Economic Growth* 7: 315–345.

Aghion, P., C. Meghir, and J. Vandenbussche. 2003. "Productivity Growth and the Composition of Education Spending." Unpublished.

Arrow, Kenneth. 1962. "Economics of Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity*, ed. R. Nelson, 609–625. Princeton, NJ: Princeton University Press.

Banerjee, A., and A. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101: 274–298.

Barro, R. J., and X. Sala-i-Martin. 1992. "Convergence." *Journal of Political Economy* 100: 223–251.

Baumol, W. J. 1986. "Productivity Growth, Convergence, and Welfare." *American Economic Review* 76: 1072–1085.

Benhabib, J., and M. Spiegel. 2002. "Human Capital and Technology Diffusion." Unpublished, New York University.

Bernanke, B., and M. Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79: 14–31.

Blundell, R., R. Griffith, and J. Van Reenen. 1999. "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms." *Review of Economic Studies* 66: 529–554.

Carmichael, H. Lorne. 1988. "Incentives in Academics: Why Is There Tenure?" *Journal of Political Economy* 96: 453–472.

Corriveau, L. 1991. "Entrepreneurs, Growth and Cycles." Ph.D. dissertation, University of Western Ontario.

Dasgupta, Partha, and Paul David. 1994. "Towards a New Economics of Science." *Research Policy* 23: 487–521.

Durlauf, S., and P. Johnson. 1995. "Multiple Regimes and Cross-country Growth Behavior." *Journal of Applied Econometrics* 10: 365–384.

Evans, P. 1996. "Using Cross-country Variances to Evaluate Growth Theories." *Journal of Economic Dynamics and Control* 20: 1027–1049.

Geroski, P. 1995. *Market Structure, Corporate Performance and Innovative Activity*. New York: Oxford University Press.

Gerschenkron, Alexander. 1962. *Economic Backwardness in Historical Perspective: A Book of Essays*. Cambridge, MA: Belknap Press of Harvard University Press.

Hellmann, Thomas, and Enrico Perotti. 2004. "Circulation of Ideas: Firms versus Markets." Stanford University working paper.

Henderson, Rebecca, Adam Jaffe, and Manuel Trajtenberg. 1998. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988." *Review of Economics and Statistics* 80: 119–127.

Howitt, P., and D. Mayer-Foulkes. 2002. "R&D, Implementation and Stagnation: A Schumpeterian Theory of Convergence Clubs." NBER working paper 9104.

Lach, Saul, and Mark Schankerman. 2004. "Incentives and Invention in Universities." London School of Economics working paper.

LaPorta, R., F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106: 1113–1155.

Levine, R., N. Loayza, and T. Beck. 2000. "Financial Intermediation and Growth: Causality and Causes." *Journal of Monetary Economics* 46: 31–77.

Lucas, R. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22: 3–42.

Maddison, A. 2001. *The World Economy: A Millennial Perspective*. Development Centre Studies. Paris: OECD.

Mankiw, G., D. Romer, and D. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107: 407–437.

Mayer-Foulkes, D. 2002. "Global Divergence." Documento de Trabajo del CIDE, SDTE 250, División de Economía.

Mayer-Foulkes, D. 2003. "Convergence Clubs in Cross-country Life Expectancy Dynamics." In *Perspectives on Growth and Poverty*, ed. Rolph van der Hoeven and Anthony F. Shorrocks, 144–171. Tokyo: United Nations University Press.

Murray, Fiona, and Scott Stern. 2004. "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-commons Hypothesis." Northwestern University working paper.

Nelson, R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67: 297–306.

Nelson, R., and E. Phelps. 1966. "Investment in Humans, Technological Diffusion, and Economic Growth." *American Economic Review* 61: 69–75.

Nickell, S. 1996. "Competition and Corporate Performance." *Journal of Political Economy* 104: 724–746.

North, Douglass C., and Robert P. Thomas. 1973. *The Rise of the Western World: A New Economic History*. New York: Cambridge University Press.

Piketty, T., G. Postel-Vinay, and J.-L. Rosenthal. 2003. "Wealth Concentration in a Developing Economy: Paris and France, 1807–1994." Mimeo., EHESS, Paris.

Pritchett, L. 1997. "Divergence, Big-time." *Journal of Economic Perspectives* 11: 3–17.

Quah, D. 1993. "Empirical Cross-section Dynamics in Economic Growth." *European Economic Review* 37: 426–434.

Quah, D. 1996. "Convergence Empirics across Economies with (Some) Capital Mobility." *Journal of Economic Growth* 1: 95–124.

Quah, D. 1997. "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs." *Journal of Economic Growth* 2: 27–59.

Romer, P. 1990. "Endogenous Technical Change." *Journal of Political Economy* 98, no. 5: S71–S102.

Savvides, A., and M. Zachariadis. 2005. "International Technology Diffusion and the Growth of TFP in the Manufacturing Sector of Developing Economies." Forthcoming in *Review of Development Economics* 9, no. 4.

Scherer, F. 1967. "Market Structure and the Employment of Scientists and Engineers." *American Economic Review* 57: 524–531.

Segerstrom, P., T. Anant, and E. Dinopoulos. 1990. "A Schumpeterian Model of the Product Life Cycle." *American Economic Review* 80: 1077–1092.

Tirole, J. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

van de Klundert, T., and S. Smulders. 1997. "Growth, Competition, and Welfare." *Scandinavian Journal of Economics* 99: 99–118.

2 Financial Institutional Reform, Growth, and Equality

Costas Azariadis and
David de la Croix

2.1 Introduction

Trends toward less public regulation of financial markets for household debt are emerging in different parts of the world. Liberalization of financial markets in OECD countries since the 1980s is well documented. Examples of this are higher loan-to-value ratios, increased competition between mortgage institutions and banks, and higher borrowing limits on consumers' personal debt. In less developed countries, financial reform is more a question of creating lending institutions in order to promote investment in human and physical capital. A majority of developing countries is now undergoing significant structural transformations, one of the most controversial adjustment areas being that of financial markets and institutions. There is little consensus as to when to liberalize financial markets or how it should be done (see Fanelli and Medhora 1998). Finally, in Eastern Europe, a new financial intermediation system has been created allowing for credit to households in some segments of the market, but there is still some way to go.

Behind the slow implementation of reforms and/or the objections raised against liberalized financial markets, we find the idea that there are upfront costs that may be deterring. To understand the foundation of these criticisms, we study the medium- and long-term impact of credit reform on the growth and distribution of income in a life-cycle economy populated by agents who differ in their ability to acquire human capital.

In this economy, deregulation amounts to an anticipated lifting of all borrowing constraints on households; namely, it is equivalent to creating credit markets starting from a situation where such markets are absent. We describe the qualitative and quantitative outcomes of this financial "big bang" on incomes, inequality, and the welfare of particular social groups indexed by age and ability. Our starting point is that borrowing limits do not necessarily ration the poor, as it is assumed in much of the literature (see, e.g., Galor and Zeira 1993; Piketty 1997). They may ration instead the most efficient accumulators of human skills, that is, households with high potential income growth.

Important clues to the answer we are seeking are identified in papers by Jappelli and Pagano (1994), De Gregorio (1996), and De Gregorio and Kim (2000), which link market liberalization to economic growth and distribution.[1] We call these clues the *level effect* and the *growth effect* from credit market reforms.

The level effect of financial deregulation is strongest in the short to medium run. It reduces *net* household saving, slows down physical capital accumulation, and raises yields in societies without human capital. This mechanism was identified by Jappelli and Pagano (1994), who found some support for it in a panel of OECD countries. They conclude that financial deregulation in the eighties has contributed to the decline in national saving and growth rates in the OECD countries.[2]

Opposed to the level effect is the growth effect, identified by De Gregorio 1996. It refers to the rise in borrowing for investments in human skills, and the corresponding boost to long-run growth in small open societies that rely on human capital as their growth engine. Evidence for this channel appears to be mixed.

De Gregorio and Kim (2000) also find that financial reform is welfare improving but may raise the dispersion of earnings by permitting the more able to specialize in learning and the less able to specialize in working. As Becker (1964) had suggested, relaxing constraints on society's ablest households contributes to earnings inequality.

This chapter is based on the assumption that physical and human capital need to be studied *jointly* both because they oppose each other and because they interact in subtle ways. For example, as the level effect raises yields and lowers wage rates, it will undermine the growth effect and itself by inducing less schooling by unconstrained people and greater labor supply. Without a complete general equilibrium model, it seems very hard to guess how financial reform now will affect output in the medium run as well as the welfare of each currently living household.

Accordingly, section 2.2 sets up a simple economy with heterogeneous households, one consumption good, and two reproducible inputs—physical capital and human capital. In section 2.3, we characterize equilibria with a perfect loan market and with an extreme form of credit rationing, that is, a prohibition on all loans. We prove that the return on capital is always higher in the economy with perfect markets. The transitional and long-term response of output and inequality to financial reform depends critically on how common credit rationing was before credit market liberalization.

The remainder of the chapter conducts dynamic simulation experiments of financial deregulation in a model calibrated to fit the long-run economic performance of a panel of less developed countries in the 1960s. Specifically, we explore in section 2.4 the quantitative implications for per capita income growth and the Gini coefficients in these countries. We pay particular attention to the changes in welfare by cohort and ability group. We find that, even when credit constraints initially bind on relatively few people, the macroeconomic consequences of removing these constraints can be large, with upfront costs from a lower capital intensity and delayed benefits from long-term growth. Initial responses to financial deregulation are dictated by the adverse level effect: a decline in the growth of output, coupled with a rise in inequality and in real yields. The growth effect eventually takes over, boosting long-term growth by about one third of 1 percent per year. The impact of liberalization is adverse for all young households at the time of the reform and also for skilled older people.

The robustness of these results to changes in technology is investigated in section 2.5. In particular we show that, with CES technologies and low substitutability between capital and labor, financial reform shrinks the basin of attraction to the higher of the two balanced growth states. If the economy considered has a low initial capital-labor ratio, or if its total factor productivity is not high enough, then the lifting of borrowing constraints that comes from financial reform may redirect economic growth toward a poverty trap. Section 2.6 sums up the costs and benefits from financial reform and discusses policies that would make liberalization more agreeable to a majority of households.

2.2 The Problem of the Household

The model is an overlapping generations model in the spirit of Azariadis and Drazen (1990), extending their approach to heterogeneous households and imperfect credit markets. Time is discrete and goes from 0 to $+\infty$. Each generation consists in a continuum of households, with mass expanding at a constant rate $n > -1$.[3]

Each individual lives for two periods, youth and old age. The households of the same generation differ in their innate ability to work when young, ε^{Y}, and when old, ε^{O}. Ability ε^{Y} can be thought of as being related to physical *strength*, while ε^{O} incorporates elements related to the ability to learn, say *IQ*.[4] Their utility function is defined over consumption when adult c_t and consumption when old d_{t+1}:

$$\ln c_t + \beta \ln d_{t+1}, \qquad \beta \in \mathbb{R}_+. \tag{1}$$

A share of time λ_t is spent to build up human capital and $1 - \lambda_t$ to work. First-period income is allocated between consumption and savings s_t:

$$\varepsilon^{\mathrm{Y}}(1-\lambda_t)w_t\bar{h}_t = c_t + s_t. \tag{2}$$

The individual variables c_t, s_t, λ_t, and d_{t+1} will generally depend on ability. Economy-wide variables are w_t, the wage per unit of human capital, and $\bar{h}_t$, which denotes the average human capital of the old generation at time t. The endowment of efficient labor when young is $\varepsilon^{\mathrm{Y}}\bar{h}_t$. Following Azariadis and Drazen (1990), each young person benefits from the average human capital of the previous generation. Old-age human capital depends on the time spent on education when young; on the ability when old, ε^{O}; and on the average value of the previous generation's human capital:

$$h_{t+1} = \varepsilon^{\mathrm{O}}\psi(\lambda_t)\bar{h}_t. \tag{3}$$

We think of $\bar{h}_t$ as a measure of teacher quality. As we can see from equation (3), the individual characteristic ε^{O} reflects both the ability to work when old and the ability to learn (i.e., to accumulate human capital). The function ψ is assumed to be increasing and concave, and satisfies boundary conditions

$$\lim_{\lambda\to 0} \psi'(\lambda) = +\infty, \qquad \lim_{\lambda\to 1} \psi'(\lambda) = 0, \tag{4}$$

which ensure that it is always optimal to spend a strictly positive time span building human capital.

The ability type $(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}})$ is distributed over each generation according to a cumulative function G defined on $\mathbb{R}^2_+$. The economy-wide average human capital is

$$\bar{h}_t = \int_0^\infty \int_0^\infty h_t \, \mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}).$$

Old agents consume both labor earnings and capital income:

$$d_{t+1} = R_{t+1}s_t + w_{t+1}h_{t+1}. \tag{5}$$

R_{t+1} is the interest factor.

We denote the relative wage by

$$x_t \equiv \frac{w_{t+1}}{w_t R_{t+1}}.$$

From equations (2), (3), and (5), life-cycle income is proportional to the inherited human capital $\bar{h}_t$:

$$\Omega_t = w_t[\varepsilon^{Y}(1-\lambda_t) + x_t\varepsilon^{O}\psi(\lambda_t)]\bar{h}_t.$$

Since the duration of schooling λ_t does not enter the utility function, we can solve the household planning problem in two separate steps. When financial markets are perfect, there is no liquidity constraint on households, and the optimal length of schooling maximizes life-cycle income, satisfying the condition

$$\psi'(\lambda_t) = \frac{\varepsilon^{Y}}{\varepsilon^{O}x_t}. \tag{6}$$

This equation represents the trade-off between studying and working put forward by Ben-Porath (1967). This relationship implies that the length of schooling depends positively on discounted future wage (the benefit from education) and negatively on current wage (the opportunity cost). It also depends positively on the ratio of innate abilities $\varepsilon^{O}/\varepsilon^{Y}$. Inverting equation (6), we obtain

$$\lambda_t = \varphi(\varepsilon^{O}x_t/\varepsilon^{Y}), \qquad \varphi' > 0,\ \varphi(0) = 0.$$

Optimal savings are computed by maximizing utility subject to the budget constraints (2) and (5):

$$(1+\beta)s_t = \left(\beta\varepsilon^{Y}(1-\lambda_t)w_t - \frac{w_{t+1}}{R_{t+1}}\varepsilon^{O}\psi(\lambda_t)\right)\bar{h}_t. \tag{7}$$

We define the increasing function as follows:

$$\Phi(a) \equiv \varphi(a) + \frac{a}{\beta}\psi(\varphi(a)) \qquad \Phi' > 0. \tag{8}$$

This allows us to rewrite savings as

$$(1+\beta)s_t = \beta w_t\varepsilon^{Y}(1-\Phi(\varepsilon^{O}x_t/\varepsilon^{Y}))\bar{h}_t. \tag{9}$$

Note that there is a threshold $\tilde{\mu}$ bearing on relative ability $\varepsilon^{\mathrm{O}}/\varepsilon^{\mathrm{Y}}$ above which households borrow from financial markets. Indeed, we note from equations (6) and (7) that savings are positive if, and only if, $\beta(1-\lambda_t)\psi'(\lambda_t) > \psi(\lambda_t)$. As $\psi(.)$ is increasing in the interval $(0,1)$ and $\psi'(\lambda_t)(1-\lambda_t)$ is decreasing in λ_t, this inequality defines a critical value for schooling, $\tilde{\lambda}$, independent of time and such that

$$\lambda_t < \tilde{\lambda} \Leftrightarrow s_t > 0.$$

Since λ_t is a monotone function $\varphi(.)$ of ability, we can define the ability threshold as a function of the relative wage:

$$\tilde{\mu}_t = \frac{\varphi^{-1}(\tilde{\lambda})}{x_t} \equiv \frac{B}{x_t}. \tag{10}$$

This threshold again separates borrowers from lenders, that is,

$$\frac{\varepsilon^{\mathrm{O}}}{\varepsilon^{\mathrm{Y}}} < \tilde{\mu}_t \Leftrightarrow s_t > 0.$$

Hence, households in cohort t with relative ability above $\tilde{\mu}_t$ (or, equivalently, with steeply rising wage profiles) will borrow while other households will lend.

We define an imperfect credit market as an environment in which young households cannot credibly commit their future labor income as a collateral against current loans. As in Kehoe and Levine 1993, we assume that individuals are allowed to borrow up to the point where they are indifferent between repaying loans and suffering market exclusion. Since everyone dies at the end of the second period, default involves no penalty and is individually optimal. The borrowing constraint then takes a very simple form: $s_t \geq 0$.[5]

We saw earlier that the households with ability ratio $\varepsilon^{\mathrm{O}}/\varepsilon^{\mathrm{Y}}$ above the threshold $\tilde{\mu}_t = B/x_t$ borrow from financial markets. Those households will now be rationed. They will not participate to the credit market, maximizing instead an autarkic utility function obtained by replacing (2), (3), and (5) in (1):

$$\ln(1-\lambda_t) + \beta\ln(\psi(\lambda_t)) + \text{constants}.$$

The first-order condition is

$$\psi(\lambda_t) = \beta\psi'(\lambda_t)(1-\lambda_t).$$

Since $\psi(.)$ is increasing in the interval $(0,1)$ and $\psi'(\lambda_t)(1-\lambda_t)$ is decreasing in λ_t, this equation defines a unique solution $\tilde{\lambda}$, which does not depend on prices, or on ability type. It is the same as the threshold $\tilde{\lambda}$ defined in (10).

We can now summarize our results in the following proposition:

Proposition 1 Households whose ability profiles do not rise fast, i.e., $\varepsilon^{\mathrm{O}}/\varepsilon^{\mathrm{Y}} < \tilde{\mu}_t$, save a positive amount given by equation (9); their investment in education λ_t equals $\varphi(\varepsilon^{\mathrm{O}} x_t/\varepsilon^{\mathrm{Y}})$ and depends positively on $\varepsilon^{\mathrm{O}}/\varepsilon^{\mathrm{Y}}$. Households with fast-rising ability profiles, i.e., $\varepsilon^{\mathrm{O}}/\varepsilon^{\mathrm{Y}} > \tilde{\mu}_t$, are credit rationed, and invest the same amount in education, i.e., $\lambda_t = \tilde{\lambda} = \varphi(\tilde{\mu}_t x_t)$.

Households with a steep potential earnings profile would like to borrow in order to study longer, but credit rationing prevents them from doing so. All others have positive saving and study as long as they wish. Note that the threshold $\tilde{\mu}_t$ depends on prices through equation (10). For example, when yields are high, there will be fewer constrained households, other things being equal. Hence, although our borrowing constraint is very simple, *the proportion of rationed people depends on prices and hence varies over time.*

2.3 The Equilibrium

To characterize the equilibrium with perfect markets, we compute the average human capital of the next period as (using equation (3))

$$\bar{h}_{t+1} = \int_0^\infty \int_0^\infty h_{t+1}\,\mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}) = \int_0^\infty \int_0^\infty \varepsilon^{\mathrm{O}} \psi(\lambda_t) \bar{h}_t\,\mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}).$$

The growth rate of human capital, $g_p(x_t)$, is therefore

$$\frac{\bar{h}_{t+1}}{\bar{h}_t} = 1 + g_p(x_t) = \int_0^\infty \int_0^\infty \varepsilon^{\mathrm{O}} \psi(\varphi(\varepsilon^{\mathrm{O}} x_t/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}). \tag{11}$$

We also equate aggregate saving with the value of the capital stock. First we compute saving per young household from

$$\bar{s}_t = \int_0^\infty \int_0^\infty s_t\,\mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}) = \frac{\beta}{1+\beta} w_t \bar{h}_t \mathscr{S}_p(x_t), \tag{12}$$

where the function $\mathscr{S}_p(x_t)$ is defined as

$$\mathscr{S}_p(x_t) = \int_0^\infty \int_0^\infty \varepsilon^{\mathrm{Y}} (1 - \Phi(\varepsilon^{\mathrm{O}} x_t/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}).$$

We assume that firms operate a constant returns to scale technology $F(K_t, H_t)$ involving capital and labor inputs. Defining the capital-labor ratio as $k_t = K_t/H_t$, and an intensive production function $f(k_t)$, equilibrium factor prices are

$$w_t = f(k_t) - k_t f'(k_t) = \omega(k_t),$$

$$R_t = f'(k_t) = R(k_t).$$

This allows us to rewrite the relative wage x_t as a function of (k_t, k_{t+1}):

$$x_t = \frac{\omega(k_{t+1})}{\omega(k_t)R(k_{t+1})}. \tag{13}$$

The total labor supply per young person H_t is obtained by averaging over young and old workers, that is,

$$H_t = \mathscr{H}_p(x_t)\bar{h}_t, \tag{14}$$

where the function $\mathscr{H}_p(x_t)$ is defined as

$$\mathscr{H}_p(x_t) = \frac{1}{1+n} + \int_0^\infty \int_0^\infty \varepsilon^{\mathrm{Y}}(1 - \varphi(\varepsilon^{\mathrm{O}} x_t/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}}).$$

Equilibrium in the financial market requires

$$K_{t+1} = k_{t+1}H_{t+1} = \frac{\bar{s}_t}{1+n}.$$

After using equations (11), (12), and (14), we find the following:

$$\frac{(1+\beta)k_{t+1}}{\beta\omega(k_t)}\mathscr{H}_p(x_{t+1}) = \frac{\mathscr{S}_p(x_t)}{1+g_p(x_t)}\,\frac{1}{1+n} \tag{15}$$

Given initial conditions $(k_0, \bar{h}_0)$, a perfect foresight equilibrium can be characterized by a non-negative sequence $(x_t, k_{t+1}, \bar{h}_{t+1})_{t\geq 0}$, which solves equations (11), (13), and (15).

This dynamical system can be solved recursively when the production function is Cobb-Douglas, $f(k_t) = Ak_t^\alpha$, with complete depreciation of capital. Then we have

$$\frac{k_{t+1}}{\omega(k_t)} = \frac{\alpha}{1-\alpha}\,\frac{\omega(k_{t+1})}{\omega(k_t)R(k_{t+1})} = \frac{\alpha}{1-\alpha}x_t,$$

and equation (15) reduces to a first-order difference equation in x_t:

$$\frac{(1+\beta)\alpha}{(1-\alpha)\beta}\mathscr{H}_p(x_{t+1}) = \frac{1}{x_t}\frac{\mathscr{S}_p(x_t)}{1+g_p(x_t)}\frac{1}{1+n}.$$

In the presence of rationing, the average human capital grows at a rate $g_c(x_t) = \bar{h}_{t+1}/\bar{h}_t - 1$, which reflects the weight of constrained households; that is,

$$1+g_c(x_t) = \int_0^\infty \int_0^{\varepsilon^Y B/x_t} \varepsilon^O \psi(\varphi(\varepsilon^O x_t/\varepsilon^Y))\,\mathrm{d}G(\varepsilon^Y,\varepsilon^O) + \psi(\tilde{\lambda})\int_0^\infty \int_{\varepsilon^Y B/x_t}^\infty \varepsilon^O\,\mathrm{d}G(\varepsilon^Y,\varepsilon^O). \tag{16}$$

Average saving is

$$\bar{s}_t = \frac{\beta}{1+\beta} w_t \bar{h}_t \mathscr{S}_c(x_t),$$

where the function $\mathscr{S}_c(x_t)$ is defined as

$$\mathscr{S}_c(x_t) = \int_0^\infty \int_0^{\varepsilon^Y B/x_t} \varepsilon^Y (1-\Phi(\varepsilon^O x_t/\varepsilon^Y))\,\mathrm{d}G(\varepsilon^Y,\varepsilon^O),$$

instead of the expression in equation (12). Similarly, average labor supply no longer satisfies equation (14); it is given instead by

$$H_t = \mathscr{H}_c(x_t)\bar{h}_t,$$

where the function $\mathscr{H}_c(x_t)$ is defined as

$$\mathscr{H}_c(x_t) = \frac{1}{1+n} + \int_0^\infty \int_0^{\varepsilon^Y B/x_t} \varepsilon^Y (1-\varphi(\varepsilon^O x_t/\varepsilon^Y))\,\mathrm{d}G(\varepsilon^Y,\varepsilon^O) + (1-\tilde{\lambda})\int_0^\infty \int_{\varepsilon^Y B/x_t}^\infty \varepsilon^Y\,\mathrm{d}G(\varepsilon^Y,\varepsilon^O).$$

Labor supply is decreasing in x_t, that is, $\mathscr{H}_c'(.) < 0$, since better earnings prospects move households from work to school. Financial market equilibrium satisfies

$$\frac{(1+\beta)k_{t+1}}{\beta\omega(k_t)}\mathscr{H}_c(x_{t+1}) = \frac{\mathscr{S}_c(x_t)}{1+g_c(x_t)}\frac{1}{1+n}. \tag{17}$$

Given the initial conditions $(k_0, \bar{h}_0)$, a perfect foresight equilibrium with credit rationing is again a sequence $(x_t, k_{t+1}, \bar{h}_{t+1})_{t\geq 0}$, which solves equations (16), (13), and (17).

With the Cobb-Douglas production function, equilibria are solutions to the dynamical system:

$$\frac{(1+\beta)\alpha}{(1-\alpha)\beta}\mathscr{H}_c(x_{t+1}) = \frac{1}{x_t}\frac{\mathscr{S}_c(x_t)}{1+g_c(x_t)}\frac{1}{1+n}, \tag{18}$$

$$k_{t+1} = A\alpha x_t k_t^{\alpha}. \tag{19}$$

This system is recursive. Equation (18) can first be solved for the path of x_t. Equation (19) is obtained from the definition of x_t in equation (13); it describes the evolution of the capital-labor ratio. The growth rate of human capital is obtained from (16). The solution to (16), (18), and (19) is summed up in the following result:

Proposition 2 The system (16), (18), and (19) has a steady state (x_c, k_c, g_c) and equilibrium is unique in the neighborhood of that state.

Proof: See appendix. ▪

The same reasoning can be applied to the perfect market economy that also possesses a locally unique equilibrium in the neighborhood of the steady state (x_p, k_p, g_p).

As a general proposition, it is impossible to show that financial reform will spread inequality and promote long-term growth. For example, liberalization raises yields (see proposition 3) and improves the income of retirees. Since this effect is stronger for less able retirees with relatively high saving, it tends to reduce inequality. What happens to long-term growth depends on how young households weigh the mixed incentives they receive in free financial markets: less credit rationing permits them to invest more in schooling, while higher yields on physical capital shrink the present value of future earnings. We first state a key result according to which financial reform reduces aggregate saving and raises yields.

Proposition 3 Assuming a unique steady state, the economy with perfect markets has a lower long-run capital-labor ratio than the one with imperfect markets.

Proof: See appendix. ▪

To assess the effect of financial reform on the long-run growth rate of per capita output (which equals the long-run growth rate of average human capital), we should compare the perfect market growth rate, $g_p(x_p)$, with the credit-rationed growth rate, $g_c(x_c)$. Two opposite effects interact: for the same long-run yield $1/x$, $g_p(x) > g_c(x)$. Indeed, some agents are constrained in the imperfect market economy, they invest less than they want in education, and growth is slower. However, as the yield is higher in the perfect market economy ($x_p < x_c$), agents are discouraged from investing in education, and this may or may not outweigh the direct positive effect. The first effect will dominate if there are enough constrained agents in the economy with imperfect markets. To evaluate the effect on growth, we need to rely on numerical simulations.

What happens to the short-run growth rate of output depends on the interaction of several factors. First, the forward-looking relative wage x drops when the reform is announced, and investment in physical capital starts to fall immediately, which is bad for short-term growth (level effect). Second, the lifting of borrowing constraints permits more investment in education, which is good for growth (growth effect). Third, the supply of labor moves in the opposite direction from investment in education, which depresses short-run growth. Last, there are additional dynamic effects when the reform is anticipated. To assess the relative importance of these mechanisms, we must rely on simulations.

Two final comments on the specification of the model deserve attention.

In this model there are intergenerational externalities in human capital investment. The average level of human capital of the old generation raises the wage income of the young generation, and also raises the productivity of the human capital investment of the young generation. As a result, one would expect, even with perfect capital markets the market economy does not achieve the first best. When we compare the perfect market economy with the one with credit constraints, we are comparing two imperfect scenarios.[6] It will be of particular interest to study how the welfare of the different households is affected by financial reform.

At equilibrium there is also some inequality. The notion of inequality in this paper is very different from that in Galor and Zeira 1993 and Newman and Banerjee 1993. There, inherited wealth relaxes borrowing constraints. In these papers, there is nothing good out of inequality, and redistribution always improves efficiency. Here there is no inherited wealth, and inequality reflects differences in ability. Low inequality should therefore not be necessarily pursued on efficiency grounds.

2.4 Dynamic Simulations

In section 2.3, we established that financial reforms that relax the borrowing constraints on households will lower the capital-labor ratio and improve growth in the long run if the number of constrained households is sufficiently high. However, the transitional impact of these reforms is less clear-cut and hard to characterize analytically. In order to study the interplay of long-run and medium-run forces along the transition path, we rely on simulations of a calibrated version of the model. This also allows us to assess the quantitative importance of liberalization for growth and inequality.

2.4.1 Calibration

We first choose functional forms for the production function of human capital and the distribution of abilities. The production of human capital has to satisfy the two limit conditions (4) to guarantee an interior solution for all agents. We use

$$\psi(\lambda) = b\left(\frac{1}{\gamma}\lambda^{\gamma} - \lambda\right).$$

The abilities index $(\varepsilon^{\mathrm{Y}}, \varepsilon^{\mathrm{O}})$ is assumed to be distributed over the population according to a bivariate lognormal distribution; the mean[7] and variance-covariance matrix of the underlying normal distribution are respectively $(0, 0)$ and

$$\Sigma = \begin{pmatrix} \sigma_{\mathrm{Y}}^2 & \varrho\sigma_{\mathrm{Y}}\sigma_{\mathrm{O}} \\ \varrho\sigma_{\mathrm{Y}}\sigma_{\mathrm{O}} & \sigma_{\mathrm{O}}^2 \end{pmatrix}.$$

Since we have no direct information to calibrate the variance-covariance matrix, we carry out a sensitivity analysis of the correlation ϱ between the two ability variables and of their relative variance $\sigma_{\mathrm{Y}}^2/\sigma_{\mathrm{O}}^2$. The scope of the analysis is restricted by assuming a positive correlation, $\varrho > 0$. It also seems reasonable to assume that the ability to work when young is less widely dispersed than the ability to work when old. Indeed, ability in youth only reflects different endowments in efficient labor, while ability in old age also embodied the ability to accumulate human capital. We thus assume $\sigma_{\mathrm{Y}}^2/\sigma_{\mathrm{O}}^2 < 1$. Keeping this ratio constant, the absolute magnitude of the two variances are chosen to match an income inequality coefficient.

The productivity parameter of the Cobb-Douglas production function A plays no role given that the utility is logarithmic; it only scales the output and capital levels. The capital share parameter α is fixed to one third according to the consensus in the literature. The psychological discount factor of households is set to 1 percent per quarter. Assuming that one period of the model is twenty-five years, we have: $\beta = 0.99^{100} = 0.366$.

For fixed ϱ and σ_Y^2/σ_O^2, there are four remaining parameters to calibrate: the growth rate of population n is directly observable; the productivity parameter b governs the long-term growth rate of output per capita; given b, the parameter γ determines the time spent on education in the first period of life; and, finally, the variance parameter σ_O^2 influences the distribution of income. We chose these parameters so that the steady state of the equilibrium with credit rationing matches some moments of a typical economy with imperfect credit markets. This representative economy is obtained from averaging eight economies considered by Bandiera, Caprio, Honohan, and Schiantarelli (2000) to have had strongly imperfect credit markets in the sixties. These are Chile, Ghana, Indonesia, Korea, Malaysia, Mexico, Turkey, and Zimbabwe.

The average growth rate of population and output is computed over the period 1960–1970 using the GDP data of the Penn World Tables. For the share of time devoted to education, we assume that the first period of the model covers ages 12–37 and the second one corresponds to ages 37–62. Doing so supposes that secondary and higher education are an alternative to working, but elementary education is not. The percentage of time devoted to schooling is therefore computed by adding the variables “average years of secondary schooling in the total population” and “average years of higher schooling in the total population” from Barro-Lee and dividing them by twenty-five. Finally, we summarize the distribution of income by a Gini index from Deininger and Squire (1996).[8]

These computations lead to the following four moments: an annual growth rate of population of 2.73 percent, a long-term per capita growth rate of 2.903 percent per year, a Gini coefficient of 0.458, and a share of time devoted to education of 2.901 percent. The value of n matching the growth rate of population is $n = 0.962$. The value of the other three parameters depends on the assumptions on ϱ and σ_Y^2/σ_O^2.

Section 2A.4 gives the variance σ_{O}^2, which matches the Gini coefficient for different combinations of ϱ and $\sigma_{\mathrm{Y}}^2/\sigma_{\mathrm{O}}^2$. The parameters b and γ are picked to match output growth and schooling. Equilibrium outcomes are reported for the percentage of the young population rationed,

$$\int_0^{\infty}\int_{\varepsilon^{\mathrm{Y}}B/x_c}^{\infty} \varepsilon^{\mathrm{O}}\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}}),$$

the saving rate,

$$\frac{1-\alpha\mathscr{S}_c(x_c)}{\mathscr{H}_c(x_c)},$$

and the annual rate of return on capital,

$$\sqrt[25]{1/x_c}-1.$$

We draw three conclusions from this sensitivity analysis. First, the percentage of households subject to a borrowing constraint is never large and reaches at maximum 19 percent. Second, when the correlation between the two random ability indexes is large, few people are constrained: in that case, relative ability $\varepsilon^{\mathrm{O}}/\varepsilon^{\mathrm{Y}}$ displays little variation across households and few people want to borrow. Third, the saving rate lies between 8.8 percent and 9.8 percent[9] and the annual rate of return on capital is around 11.2 percent, whichever variance-covariance matrix we pick.

In order to choose a reasonable variance-covariance matrix Σ, we look at the characteristics of the distribution of income for different parameters values. Section 2A.5 reports income Gini indexes per cohort and the ratio of the mean to the median of the earnings distribution. We chose to use in the sequel $\varrho = 0.2$ and $\sigma_{\mathrm{Y}}^2/\sigma_{\mathrm{O}}^2 = 0.8$. A correlation of 0.2 seems reasonable, given a span of twenty-five years between the two ability shocks and the fact that ε^{O} incorporates the ability to learn while ε^{Y} does not. A relative variance of 0.8 reproduces a ratio of Gini indexes of $0.42/0.53 = 0.79$, which is close to U.S. data (see Diaz-Gimenez, Quadrini, and Rios Rull 1997). Figure 2.1 plots the corresponding density function of abilities. The vertical plane represents the threshold above which people are rationed. Constrained households lie on the left side of the picture and represent 15.5 percent of the population; they are those with a high income growth potential (either low ε^{Y} or high ε^{O}).

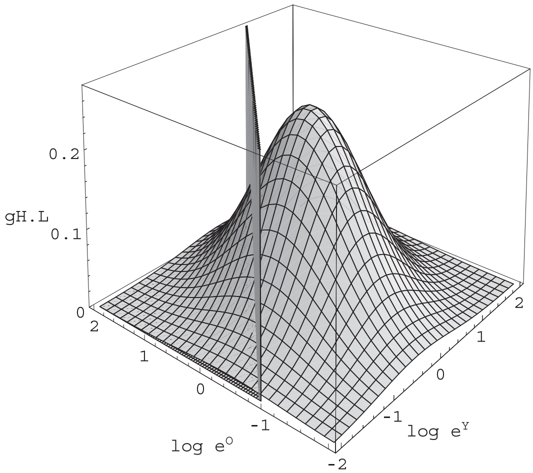


Figure 2.1
The distribution of abilities and rationed households.

2.4.2 *Response to Reform*

We now simulate the transition from a steady state with credit rationing to the one in the perfect market economy. The relaxation of the borrowing constraints takes place at time $t = 3$ and is anticipated one period in advance. Time $t = 1$ represents the initial steady state with credit constraints. Figure 2.2 represents the dynamic path of the three key variables, (x_t, k_t, and g_t^y), that is, relative wage, capital-labor ratio, and growth rate in per capita income. When liberalization is announced, the relative wage x_t looks forward; it jumps close to the steady-state level that will be reached at the time of the reform. This makes future wages less attractive and discourages investment in human capital at $t = 2$.

Because x_t is also the investment rate, the capital-labor ratio k_t starts declining at $t = 3$. The saving rate drops by half a percent. This decline in the stock of capital is key to explaining the drop in the annual growth rate at $t = 3$ from 2.9 percent to 2.7 percent over twenty-five years.

At $t = 3$ the ablest households are now allowed to borrow, increasing their investment in education and lengthening average schooling from 2.9 percent to 3.6 percent. This is not very large, but it is sufficient to drive growth above its initial level by about 0.15 percent.

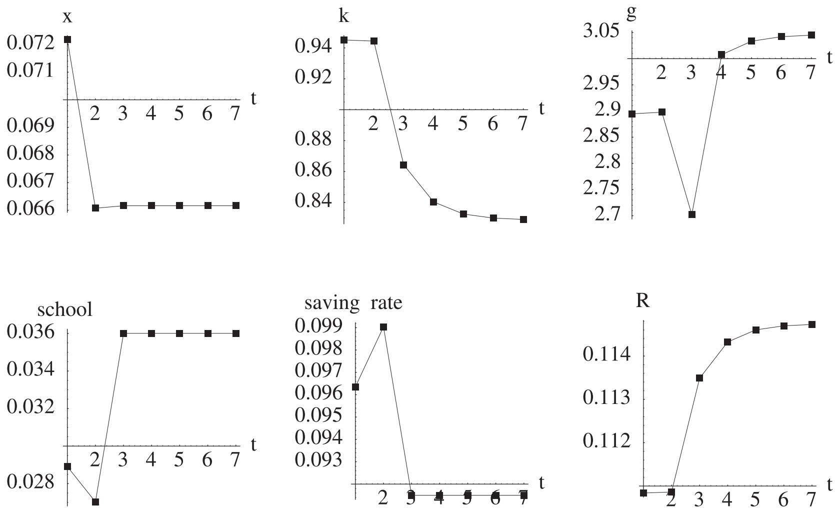


Figure 2.2
Dynamic responses to reform.

A sensitivity analysis of this magnitude to the chosen values of ρ and σ_Y^2/σ_O^2 is presented in section 2A.6: the gain is between 0 and 0.30 percent, and depends on the percentage of constrained households in the initial balanced growth path.

What might have happened if we had calibrated on the same set of economies for a different time period, or on an altogether different set of emerging economies? To see how our outcomes are sensitive to parameters, we summarize in table 2.1 the response of constraints, saving rates, and growth rates as the parameter structure changes relative to the baseline calibration. We conclude that the increase in long-term growth is largest in economies with high schooling and slow population growth, and smallest in economies with high capital share and low initial inequality. Changes typically show weak sensitivity to any single parameter and are almost completely insensitive to the pre-reform growth rate.

To better grasp the cost of this financial reform, figure 2.3 plots both the Gini coefficient and the difference between the GDP the economy would have enjoyed without reform and the one with the reform. Inequality peaks at $t = 3$ before stabilizing above its pre-reform level. The long-run effect is essentially explained by the fact that the ablest people can now fully exploit their advantage by going to school longer,

Table 2.1
Sensitivity analysis

	γ	B	σ_O^2	% constr.	Saving rate	Drop in saving rate	Gain in growth
Baseline case	0.234	0.739	0.74	15.5	9.6	−0.5	+0.15
More schooling (5% instead of 2.9%)	0.456	2.694	0.67	18.8	9.5	−0.6	+0.27
Slower population growth (1.6% instead of 2.9%)	0.210	0.619	0.74	18.0	8.5	−0.5	+0.17
Higher capital share ($\alpha = 1/2$)	0.370	1.982	0.67	8.2	8.1	−0.2	+0.08
Less inequality (Gini = 0.35)	0.232	0.877	0.41	10.5	8.8	−0.3	+0.07
Less output growth (1.5% instead of 2.9%)	0.234	0.523	0.75	15.6	9.6	−0.5	+0.15
More patience ($\beta = .995^{100}$)	0.152	0.359	0.77	12.1	14.0	−0.5	+0.08

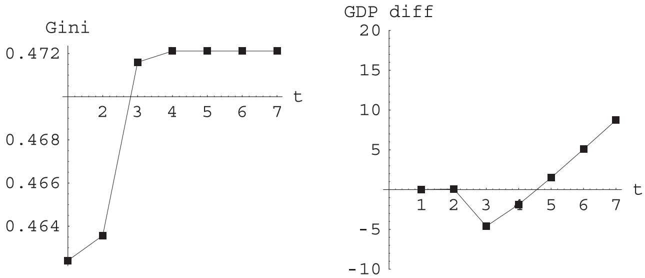


Figure 2.3
Costs of liberalization.

implying that old able persons are much richer in the perfect market economy than in the credit-constrained one.

The loss of output linked to the fall in physical capital also peaks at $t = 3$. It is around 5 percent at the time of the reform. It takes three periods to catch up and then overtake the level without reform.

Even though only 15.5 percent of the population was constrained in the initial state of the economy, financial reform leads to significant effects, both in the medium run and in the long run. We conclude that borrowing constraints may have a major impact on economic growth and inequality even if they affect a small fraction of households, provided that those include individuals with high income growth potential.

Gains from financial reform are displayed in figure 2.4, which describes the increment in life-cycle utility for members of different cohorts as a function of their abilities. Recall that the reform reduces the wage per unit of human capital from $t = 3$ onward and rises yields.

Looking first at the generations alive at the time of the liberalization, we can identify two gainers:

1. The cohort born at $t = 2$ (old at $t = 3$) with low relative ability $\varepsilon^O/\varepsilon^Y$ loses almost nothing in wages but do gain from the higher interest rate at $t = 3$; cf. the right side of panel (a).
2. The cohort born at $t = 3$ with high relative ability $\varepsilon^O/\varepsilon^Y$ gains from the lifting of the borrowing constraints; cf. the left side of panel (b).

On the contrary, a huge majority of young households born at $t = 3$ (cf. the right side of panel (b)) loses from liberalization, primarily because of lower wages per unit of human capital. Since in our model economy there is 1.962 young households for each old one, 32 percent of the total population living at $t = 3$ gains $([1.962 \times 11 + 74]/2.962 = 32)$.

Looking now at future generations, one out of two children of the generation born in $t = 4$ gain, essentially because they will benefit from the increase in GDP in their old days (see panel (c)). One hundred percent of the grandchildren gain (see panel (d)).

2.4.3 Compensating the Losers

Resistance to reform, which is at the root of slow financial liberalization, is directly related to a conflict going on between young and old, and between households with high- and low-income growth. Since this conflict will last for two generations and the reforms will increase total production, transfer schemes can be introduced to compensate the losers.

The timing of gains and losses suggests that public debt is one device that may allow all generations to share the gains from reform. In particular, suppose that the government pays subsidies to currently active households, by issuing public debt that will be repayed slowly by taxing future generations. Public debt will typically crowd out capital, amplifying the adverse level effect of the reform in the medium run

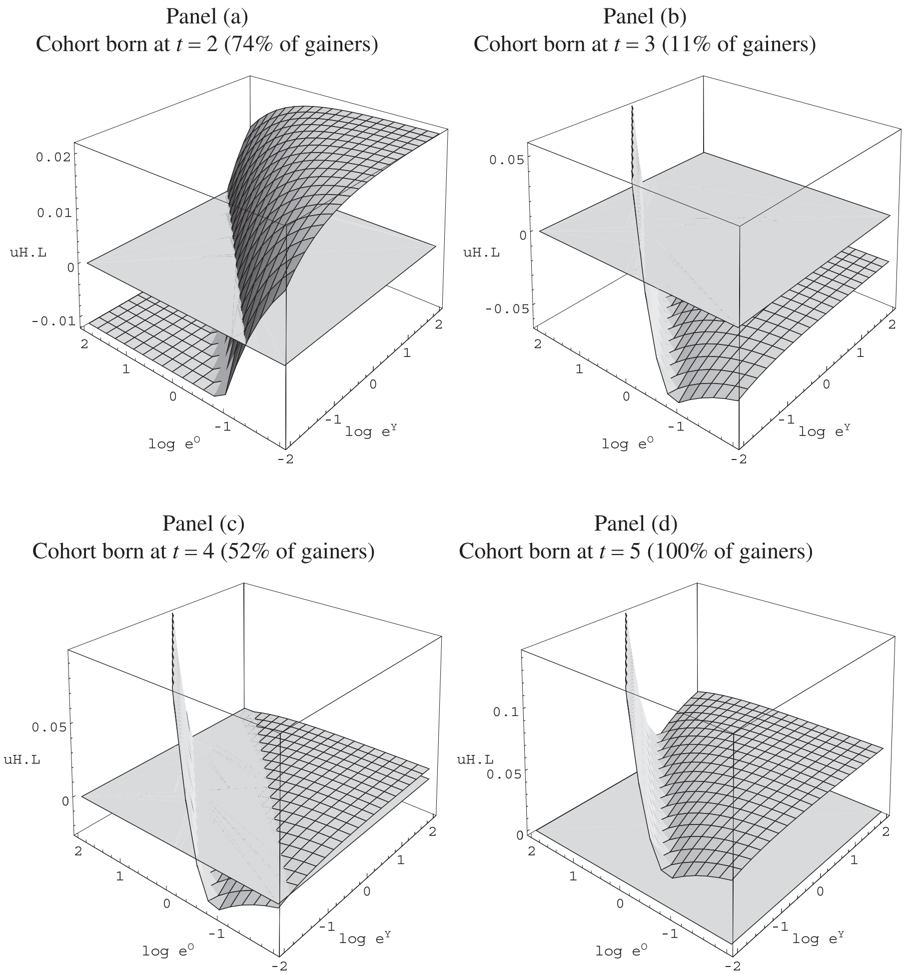


Figure 2.4
Gains in life-cycle utility by cohorts.

and undermining the favorable long-run growth effect. How to strike the right balance between medium-term redistribution and long-term incentives remains an open issue that deserves a careful analytical treatment.

Finally, as stressed by Fidrmuc and Noury (2003), ex ante acceptability of the reform can be ensured by a promise that the losers will be compensated ex post. However, such commitment may not be credible. Therefore, the reform program will likely receive greater political support ex ante when there are established frameworks for compensating losers such as an effective pension system compensating the old unskilled workers and/or when there are sufficient provisions for upholding the losers' interests (e.g., through a broadly representative system of government).

2.5 Reforms and Poverty Traps

Forty years ago, Arrow, Chenery, Minhas, and Solow (1961) taught us that economic analysis based on a unitary elasticity of substitution between labor and capital often leads to unduly restrictive conclusions. For example, estimates for developed countries consistently find that the elasticity of substitution is not different from unity, but much lower values have been found for LDCs.[10] This may reflect more limited technological options in emerging economies, namely, entrepreneurs choosing from the set of technologies in current or local use rather than from the broader set of all potential technologies.

In our specific context, we have two reasons to believe that lower substitution between production factors might affect the adjustment to financial reforms. First, it makes factor prices more sensitive to changes in the capital-labor ratio. Liberalization is thus expected to increase yields in a stronger way and to diminish the growth effect from human capital accumulation.

Second, CES technologies are consistent with poverty traps in the basic overlapping generations model (Azariadis 1996). If the initial capital-labor ratio is low enough, the economy will converge to the trivial steady state with zero capital instead of the one with high capital-labor ratio. In our setup, *financial reform tends to lower national saving and shrink the basin of attraction of the higher steady state*. As a result, more development paths will converge to the poverty trap. This is a powerful argument against reform: if the economy considered has an initial capital-labor ratio close to the region that leads to the poverty trap, the lifting of borrowing constraints that comes from financial reform may drive the economy *out* of the attraction basin of the high steady state.

Consider the class of CES production functions,

$$f(k) = A(\alpha k^{(\nu-1)/\nu} + 1 - \alpha)^{\nu/(\nu-1)},$$

with parameters $\nu, A > 0$ and $\alpha \in (0, 1)$. We set the elasticity of substitution ν equal to one-half, which we regard as a lower bound on the actual elasticity. To better assess the role of the low elasticity of substitution, the parameters b, σ^2, β, and γ keep the same value as in the Cobb-Douglas case. We adjust the parameters A and α in order to obtain a high steady state as close as possible to the previous case in terms of both growth rate and capital share in production. With $A = 53.5$ and $\alpha = 0.425$, we obtain a steady state with imperfect market displaying the same growth and capital share as previously. All the other variables are very close to their level in the Cobb-Douglas case, and 15.8 percent of young households face borrowing constraints.

Figures 2.5 and 2.6 display the response to financial reform that follows the same timing as in the Cobb-Douglas case, namely, the reform is announced at $t = 2$ and takes place at $t = 3$. Compared to figures 2.2 and 2.3, we find three differences. First, as expected, the effect on yields is stronger: the return on capital rises from 11 percent to 11.7 percent instead of going from 11 percent to 11.5 percent as it did in the Cobb-Douglas case. Second, the drop in output at $t = 3$ is almost of the same magnitude as previously, but the long-run gain is *lower*. Third, the gains from the reform take more time to materialize: GDP takes four periods instead of three to catch up. As a consequence of the weaker growth effect, the long-term gains are much more modest; after seven periods, GDP is 4 percent greater than it would be without reform, instead of 10 percent in the Cobb-Douglas case.

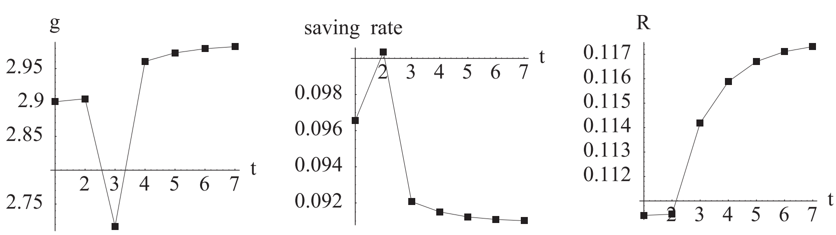


Figure 2.5
Dynamic responses to reform, CES case.

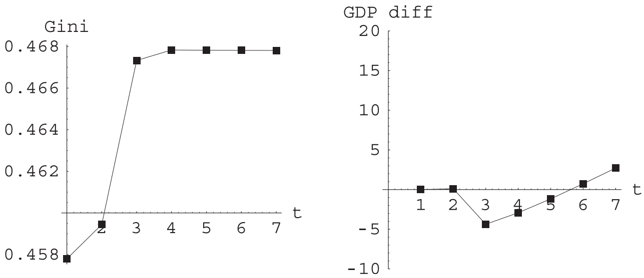


Figure 2.6
Costs of liberalization, CES case.

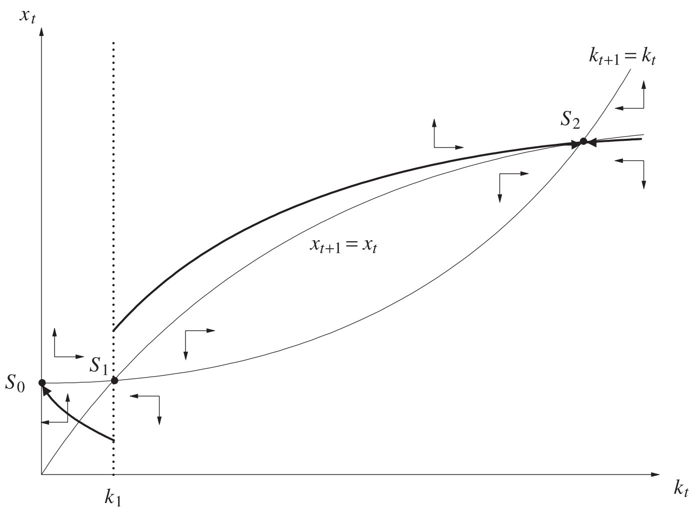


Figure 2.7
Phase diagram, CES case.

To evaluate more fully how financial reform alters the course of an emerging economy, we need to understand the global dynamics of an economy with credit rationing. This economy is described by equations (13) and (17), which lead to the phase diagram shown in figure 2.7. The phase lines $k_{t+1} = k_t$ and $x_{t+1} = x_t$ and the corresponding direction of motions are derived in section 2A.3. Depending on parameter values, the two phase lines may or may not intersect. Figure 2.7 represents the typical case where there are three steady states; point S_1 is a source and points S_0 and S_2 are saddles. If initial capital is below k_1, the equilibrium will converge to the trivial steady state S_0 in which there is no

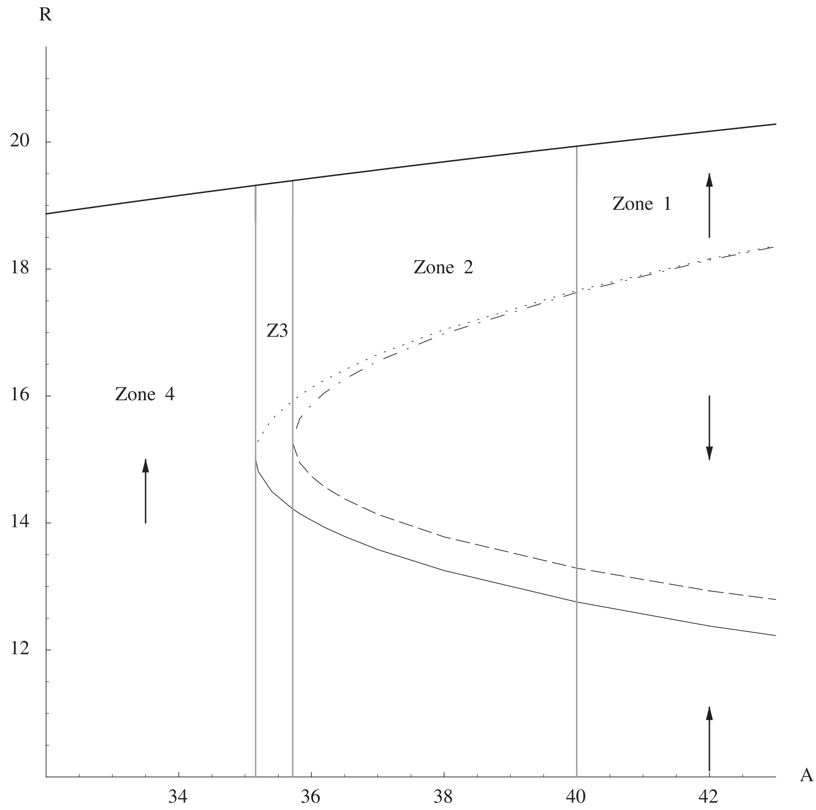


Figure 2.8
Bifurcation diagram for financial reforms.

production. If it is above, the equilibrium converges to S_2. Saddle paths are indicated by bold lines.

Credit market reform does not modify the position of the phase line $k_{t+1} = k_t$. Using the same arguments as in proposition 2, we can show that reforms moves the phaseline $x_{t+1} = x_t$ downward. Two situations may arise, depending on whether there is a positive steady state under a perfect credit market. This will depend crucially on values of the total factor productivity A and of the rate of time preference.

The bifurcation diagram in figure 2.8 shows how the existence of steady states, and their stability characteristics are sensitive to the value of the total factor productivity A. The annualized capital yield $R(k)$ is on the vertical axis, and total factor productivity on the horizontal one. All other parameters are set at their calibrated values of section

2.4. Reading the chart from bottom to top, the solid line indicates the saddlepoint-stable steady state of the economy with rationing. The dashed line above gives the corresponding saddlepoint-stable steady state of the economy with perfect credit market. The vertical distance between the two lines measures the increase in the long-run return on capital caused by the financial reform at each value of total factor productivity.

Dotted lines represent the unstable steady state of the economy with rationing (top) and without rationing (bottom), respectively. These lines also define the attraction basin of the stable steady state: if the economy starts with an initial return $R(k_0)$ outside that basin, then equilibrium will converge to the poverty steady state, and $R(k_t)$ converges to the solid line $f'(0)$. The vertical distance between the two dotted lines measures how much the attraction basin shrinks after the liberalization.

This diagram sums up the economy's response to financial reform in four different regions:

Zone 1: For $A > 40$ (corresponding to a no-liberalization annual growth rate $g^y > 2.62$) liberalization affects the unstable steady state and the attraction basin very little. This is because yields are high at the unstable steady state, and very few agents (less than 1 percent) are credit-rationed there.

Zone 2: For $40 > A > 35.718$ ($2.62 > g^y > 2.34$), liberalization shrinks the basin of attraction a bit more. If reform occurs when the economy is close to the low steady state, then liberalization will drive the equilibrium into the poverty trap.

Zone 3: For $35.718 > A > 35.1579$ ($2.34 > g^y > 2.17$), there is *no* steady state for the economy with complete markets. In this case, liberalization will lead the economy into the poverty trap for *any* initial value of the capital-labor ratio.

Zone 4: For $A < 35.1579$, there is no positive steady state. The economy will converge to the poverty trap with or without reform.

The policy implications of this picture depend very much on whether the economy is situated in Zone 2 or in Zone 3. In Zone 2, a natural policy implication is the traditional policy of foreign aid in terms of international debt that was a core policy implication of the development models of the sixties. The inefficiency in the model comes from reduced capital formation in the short run, which is more than compensated for by large human capital in the long run. Hence, in the line with the original "Washington consensus" view, financial liberalization coupled with foreign investment or foreign debt allows an economy to overcome the poverty trap. On the contrary, Zone 3 describes a "premature" liberalization. An economy with a total factor productivity (TFP) in this range should first build up its TFP by promoting structural microeconomic reforms before attempting financial reform.[11] Any foreign aid in this case would be wasteful.

2.6 Conclusions and Policy Implications

Financial reform in this chapter amounts to abolishing credit constraints on the most efficient human capital accumulators of an emerging economy. Calibrating the model to match the long-run operating characteristics (schooling, growth rate, income distribution) of a panel of eight economies in the sixties, we find that reform

1. eases constraints on individuals with rising lifetime ability profiles (15 percent of the population), accelerating long-term growth by about 0.15 percent per year;

2. reduces the household saving rate permanently and lowers the GDP growth rate temporarily by 0.3 percent per year, relative to the no-reform path. Post-reform output does not recover fully until several periods later, when the impact of higher skills overcomes the weakness of aggregate savings;

3. raises income inequality by a permanent margin;

4. lowers the life-cycle utility of nine out of ten people ages 12–37 at the time of reform as well as the ablest 25 percent among the older group ages 37–62. Without some type of compensation scheme, the losers from reform represent about two thirds of all economically active households;

5. improves the welfare of half the generation born at the time of the reform and of all members in all cohorts born later;

6. may permanently change for the worse the growth path of least developed economies, if it occurs prematurely, that is, before TFP becomes large enough. In particular, if the capital-labor elasticity of substitution is near one half, and physical capital and factor productivity are both low enough to drive the annualized net yield on capital up to 15–17 percent, then a financial reform of the type we consider here alters the course of economic growth permanently. Instead of converging to its pre-reform steady-state yield of 14–15 percent, the

post-reform economy is diverted to a poverty trap with an annualized capital yield of nearly 19 percent.

Even if we ignore the increased potential for a poverty trap, most rational households in the economy we describe would object to financial reform as we defined it. It comes as no surprise to us that opposition to less regulation and more competition in financial markets is so strong in actual economies; we are rather intrigued by the observation that majorities occasionally agree to reforms. Arguments in favor of reform are that altruism sways people to reckon the benefits that accrue to their descendants, and transfers from gainers persuade the losers to drop their objections.

Appendix

2A.1 Proof of Proposition 2

Proof: To prove this result, we show that there is a steady state in the dynamics of x_t given by equation (18), and that it is locally unstable. If this is true, the only possibility consistent with the existence of an equilibrium with perfect foresight is for the forward-looking variable x_t to be at steady state x for all $t \geq 0$. Given that $x_t = x_c\ \forall t$, the dynamics of k_t given by (19) converge monotonically to the steady state.

Equation (18) can be written as

$$J(x_{t+1}) = H(x_t).$$

Computing the limits of these functions on their interval of definition, we find

$$J(0) = \frac{(1+\beta)\alpha}{(1-\alpha)\beta}\frac{2+n}{1+n} > J(\infty) = \frac{(1+\beta)\alpha}{(1-\alpha)\beta}\left(\frac{2+n}{1+n} - \tilde{\lambda}\right)$$

$$H(0) = +\infty > H(\infty) = 0.$$

Given that $H(0) > J(0)$ and $H(\infty) < J(\infty)$, there is a steady state x such that $J(x) = H(x)$. The local instability of x is guaranteed by $-H'/J' > 1$. ∎

2A.2 Proof of Proposition 3

Proof: To compare the steady states in the two economies, we define the functions as follows:

$$T(x,i) = (1+n)\frac{(1+\beta)\alpha}{(1-\alpha)\beta}\left(\int_0^{\infty}\int_0^{i\varepsilon^{\mathrm{Y}}} \varepsilon^{\mathrm{O}}\psi(\varphi(\varepsilon^{\mathrm{O}}x_t/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}}) + \psi(\tilde{\lambda})\int_0^{\infty}\int_{i\varepsilon^{\mathrm{Y}}}^{\infty} \varepsilon^{\mathrm{O}}\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}})\right)x$$

$$W(x,i) = \frac{\int_0^{\infty}\int_0^{i\varepsilon^{\mathrm{Y}}} \varepsilon^{\mathrm{Y}}(1-\Phi(\varepsilon^{\mathrm{O}}x_t/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}})}{\frac{1}{1+n} + \int_0^{\infty}\int_0^{i\varepsilon^{\mathrm{Y}}} \varepsilon^{\mathrm{Y}}(1-\varphi(\varepsilon^{\mathrm{O}}x_t/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}}) + (1-\tilde{\lambda})\int_0^{\infty}\int_{i\varepsilon^{\mathrm{Y}}}^{\infty} \varepsilon^{\mathrm{Y}}\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}})}.$$

The steady state x_p of the perfect market economy is characterized by $T(x_p, \infty) = W(x_p, \infty)$. The one of the economy with credit rationing x_c is given by $T(x_c, B/x_c) = W(x_c, B/x_c)$.

The function T is increasing in both of its arguments. To evaluate the sign of the derivatives of $W(.)$, we replace the function Φ by its value from (8), and we obtain after some manipulations

$$1 - W(x,i) = \frac{1 + \int_0^{\infty}\int_0^{i\varepsilon^{\mathrm{Y}}} \frac{\varepsilon^{\mathrm{O}}x}{\beta}\psi(\varphi(\varepsilon^{\mathrm{O}}x/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}}) + \frac{B}{\beta}\psi(\tilde{\lambda})\int_0^{\infty}\int_{i\varepsilon^{\mathrm{Y}}}^{\infty} \mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}})}{\frac{1}{1+n} + \int_0^{\infty}\int_0^{i\varepsilon^{\mathrm{Y}}} \varepsilon^{\mathrm{Y}}(1-\varphi(\varepsilon^{\mathrm{O}}x/\varepsilon^{\mathrm{Y}}))\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}}) + (1-\tilde{\lambda})\int_0^{\infty}\int_{i\varepsilon^{\mathrm{Y}}}^{\infty} \varepsilon^{\mathrm{Y}}\,\mathrm{d}G(\varepsilon^{\mathrm{Y}},\varepsilon^{\mathrm{O}})},$$

which is increasing in x for fixed i and increasing in i for fixed x. We deduce that the function W is decreasing in both of its arguments.

Hence, the condition $T(x,i) - W(x,i) = 0$ defines an implicit function

$$x = Q(i) \qquad \text{with } Q' < 0.$$

Since i is infinite in the perfect market case and finite in the imperfect case, we obtain that $x_p < x_c$. Using (19), which holds for both economies, we obtain that the capital-labor ratio is lower in the perfect market economy. ∎

2A.3 Phase Diagram

The first relationship, equation (13),

$$x_t = \frac{\omega(k_{t+1})}{\omega(k_t)R(k_{t+1})},$$

describes an implicit function

$$k_{t+1} = \Gamma(k_t, x_t), \qquad \Gamma_k > 0,\ \Gamma_x > 0,$$

which is increasing in each argument. Note also that $\Gamma(0, x) > 0$ for any $x > 0$, because, for any CES production function with $\nu < 1$, $R(k)$ is bounded from above.

The locus of points where $k_{t+1} = k_t$ is defined by $x_t = 1/R(k_t)$, which is increasing and has a positive intercept $1/f'(0)$ for an elasticity of substitution $\nu < 1$. Above this line, $k_{t+1} > k_t$ because Γ is increasing in x_t.

The second relationship $x_{t+1} = \Psi(k_t, x_t)$ is derived from equation (17) where k_{t+1} has been replaced by $\Gamma(k_t, x_t)$:

$$\frac{1+\beta}{\beta}\mathscr{H}_c(x_{t+1}) = \frac{\mathscr{S}_c(x_t)}{1+g_c(x_t)}\,\frac{\omega(k_t)}{(1+n)\Gamma(k_t, x_t)}. \tag{20}$$

The lefthand side of this relation is decreasing in x_{t+1}, while the righthand side is decreasing in x_t. Furthermore, for any elasticity of substitution $\nu < 1$, one can show that $\omega(k)/\Gamma(k, x)$ is increasing in k for each fixed x. It follows that the function $\Psi(k, x)$ is decreasing in k and increasing in x:

$$x_{t+1} = \Psi(k_t, x_t), \qquad \Psi_k < 0,\ \Psi_x > 0.$$

The locus of points where $x_{t+1} = x_t$ defined by $x_t = \Psi(k_t, x_t)$ has a zero intercept: for any $x > 0$, the fact that $\Gamma(0, x) > 0$ implies that $x = 0$ is the only solution to the equation $x = \Psi(0, x)$. Furthermore, repeating the arguments in the proof of Proposition 2, we can show that $\Psi_x(k, x) > 1$ for each fixed k. In addition, the equation $x = \lim_{k\to\infty} \Psi(k, x)$ has a bounded solution in x. Therefore, the phase line $x_{t+1} = x_t$ is upward-sloped, starting below the phaseline $k_{t+1} = k_t$ at $k_t = 0$, and ending below it as $k_t \to \infty$.

2A.4 Sensitivity Analysis with Respect to Σ

Calibrated value of σ_O^2

	σ_Y^2/σ_O^2				
ϱ	0.2	0.4	0.6	0.8	1.0
0.0	1.45	1.07	0.88	0.74	0.64
0.2	1.44	1.09	0.89	0.74	0.65
0.4	1.44	1.09	0.90	0.76	0.65
0.6	1.44	1.11	0.90	0.77	0.68
0.8	1.42	1.11	0.92	0.78	0.69

Rationed households (percent of population)

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	14.2	15.8	17.2	18.1	18.8
0.2	12.4	13.6	14.6	15.6	16.3
0.4	10.3	10.8	11.6	12.4	13.1
0.6	8.0	7.6	7.7	8.3	8.9
0.8	5.4	3.7	3.1	3.0	3.4

Saving rate

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	9.4	9.6	9.7	9.8	9.8
0.2	9.3	9.4	9.5	9.6	9.7
0.4	9.1	9.3	9.4	9.5	9.6
0.6	9.0	9.1	9.2	9.4	9.5
0.8	8.8	8.9	9.1	9.2	9.4

Annual return on capital

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	11.2	11.1	11.1	11.0	11.0
0.2	11.3	11.2	11.1	11.1	11.0
0.4	11.3	11.3	11.2	11.2	11.1
0.6	11.4	11.3	11.3	11.2	11.2
0.8	11.5	11.4	11.4	11.3	11.2

2A.5 Income Distribution as a Function of Σ

Mean to median ratio for all earnings

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	1.29	1.37	1.40	1.41	1.42
0.2	1.30	1.35	1.39	1.43	1.42
0.4	1.30	1.36	1.40	1.40	1.44
0.6	1.29	1.37	1.41	1.43	1.42
0.8	1.28	1.36	1.40	1.43	1.44

Earnings Gini—young generation

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	0.30	0.36	0.40	0.42	0.43
0.2	0.30	0.36	0.40	0.42	0.44
0.4	0.30	0.36	0.40	0.42	0.44
0.6	0.29	0.36	0.40	0.43	0.43
0.8	0.29	0.36	0.40	0.42	0.45

Earnings Gini—old generation

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	0.70	0.62	0.57	0.52	0.49
0.2	0.69	0.62	0.56	0.53	0.49
0.4	0.70	0.62	0.56	0.53	0.49
0.6	0.71	0.62	0.57	0.52	0.48
0.8	0.71	0.62	0.56	0.52	0.47

2A.6 Growth Effect as a Function of Σ

Output growth in the perfect market economy

ϱ	σ_Y^2/σ_O^2 0.2	0.4	0.6	0.8	1.0
0.0	3.30	3.19	3.15	3.11	3.09
0.2	3.25	3.12	3.09	3.05	3.02
0.4	3.19	3.08	3.04	3.02	3.00
0.6	3.13	3.01	2.98	2.96	2.95
0.8	3.06	2.94	2.92	2.91	2.91

Notes

We thank participants to seminars in Toulouse, Stockholm, Louvain-la-Neuve, Ghent, Marseilles, and Venice (CESIfo) as well as R. Anderson, M. Cervellati, F. Heylen, and two anonymous referees for their comments on an earlier draft.

1. See also Bencivenga and Smith (1991) and Ljungqvist (1993) who stress information and commitment issues in financial markets.

2. A similar result for LDCs is obtained by Bandiera, Caprio, Honohan, and Schiantarelli (2000) who stress that liberalization—and, in particular, those elements that relax liquid-

ity constraints—may be associated with a fall in saving. Norman, Schmidt-Hebbel, and Serven (2000) also find that the relaxation of credit constraints leads to a decrease in the private saving rate.

3. A model with constant population would be slightly simpler, but allowing for population growth will make the calibration more realistic since $n = 0.962 \neq 0$ in the data we use.

4. A unidimensional heterogeneity would in fact be enough to derive the main results of the chapter, but the bivariate distribution is more realistic for calibration. With a univariate distribution, all the results would be reinforced.

5. A related formulation, due to Jappelli and Pagano (1994), would be to permit borrowing up to a “natural” debt limit that amounts to a fixed, and typically small, fraction of the present value of future income.

6. And it may even be the case that not having a credit market might be welfare enhancing from the theory of second best.

7. The mean can be normalized without loss of generality.

8. Where possible, the Gini coefficients are from 1970; otherwise, we used the closest available year. The Gini in the model is computed over the incomes of both young and old people at steady state.

9. This lies below the average saving rate of 15.49 percent computed from the data of Bandiera, Caprio, Honohan, and Schiantarelli (2000) but seems still acceptable.

10. For example, Sosin and Fairchild (1984) find an average elasticity of one half using a sample of 221 Latin American firms in the seventies.

11. Note that this range does not correspond to totally unreasonable values of the endogenous variables. For example, with $A = 35.4$, the steady state with imperfect markets has a return rate on capital of 14.5 percent, a capital share in output of 60 percent, and a growth rate of 2.28 percent.

References

Arrow, Kenneth, Hollis Chenery, B. S. Minhas, and Robert Solow. 1961. “Capital-labor Substitution and Economic Efficiency.” *The Review of Economic and Statistics* 43, no. 3: 225–250.

Azariadis, Costas. 1996. “The Economics of Poverty Traps—Part One: Complete Markets.” *Journal of Economic Growth* 1, no. 4: 449–486.

Azariadis, Costas, and Allan Drazen. 1990. “Threshold Externalities in Economic Development.” *Quarterly Journal of Economics* 105, no. 2: 501–526.

Bandiera, Oriana, Gerard Caprio, Patrick Honohan, and Fabio Schiantarelli. 2000. “Does Financial Reform Raise or Reduce Saving?” *Review of Economics and Statistics* 82, no. 2: 239–263.

Becker, Gary. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Ben-Porath, Yoram. 1967. “The Production of Human Capital and the Life-cycle of Earnings.” *Journal of Political Economy* 75, no. 4: 352–365.

Bencivenga, Valerie, and Bruce Smith. 1991. "Financial Intermediation and Endogenous Growth." *Review of Economic Studies* 58: 195–209.

De Gregorio, José. 1996. "Borrowing Constraints, Human Capital Accumulation and Growth." *Journal of Monetary Economics* 37, no. 1: 49–71.

De Gregorio, José, and Se-Jik Kim. 2000. "Credit Markets with Differences in Abilities: Education, Distribution and Growth." *International Economic Review* 41, no. 3: 579–607.

Deininger, Klaus, and Lyn Squire. 1996. "Measuring Income Inequality: A New Database." Development discussion paper no. 537, Harvard Institute for International Development.

Diaz-Gimenez, Javier, Vincenzo Quadrini, and Jose-Victor Rios Rull. 1997. "Dimensions of Inequality: Facts on the U.S. Distributions of Earnings, Income and Wealth." *Federal Reserve Bank of Minneapolis Quarterly Review* 21, no. 2: 3–21.

Fanelli, Jose, and Rohinton Medhora. 1998. *Financial Reform in Developing Countries*. London: Palgrave Macmillan.

Fidrmuc, Jan, and Abdul Noury. 2003. "Interest Groups, Stakeholders, and the Distribution of Benefits and Costs of Reform." ECARES, Brussels.

Galor, Oded, and Joseph Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Economic Studies* 60, no. 1: 35–52.

Jappelli, Tullio, and Marco Pagano. 1994. "Saving, Growth and Liquidity Constraints." *Quarterly Journal of Economics* 109, no. 1: 83–109.

Kehoe, Timothy, and David Levine. 1993. "Debt-constrained Asset Markets." *Review of Economic Studies* 60, no. 4: 865–888.

Ljungqvist, Lars. 1993. "Economic Underdevelopment: The Case of Missing Market for Human Capital." *Journal of Development Economics* 40, no. 2: 219–239.

Newman, Andrew, and Abhijit Banerjee. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101, no. 2: 274–298.

Norman, Loayza, Klaus Schmidt-Hebbel, and Luis Serven. 2000. "What Drives Private Saving across the World?" *Review of Economics and Statistics* 82, no. 2: 165–181.

Piketty, Thomas. 1997. "The Dynamics of Wealth Distribution and the Interest Rate with Credit Rationing." *Review of Economic Studies* 64: 173–189.

Sosin, Kim, and Loretta Fairchild. 1984. "Nonhomotheticity and Technological Bias in Production." *Review of Economics and Statistics* 66, no. 1: 44–50.

3 How Do Institutions Lead Some Countries to Produce So Much More Output per Worker than Others?

Theo S. Eicher, Cecilia García-Peñalosa, and Utku Teksoz

3.1 Introduction

Development accounting exercises have established that the observed per capita income differences across countries are only partially explained by variations in production inputs.[1] Of these large (up to thirty-six-fold) differences, about half are commonly attributed to a regression residual that Abramowitz (1956) termed "the economists' measure of ignorance." To capture the determinants of the sizable differences in residuals in turn, a voluminous empirical literature has emphasized the role of institutions. Cross-country regressions have shown that institutions are highly correlated with income per capita, and that institutions can explain up to thirtyfold per capita income differences between developed and developing countries.[2]

Previous empirical approaches to estimating the power of institutions in explaining per capita income rely on reduced forms, regressing output on institutions only. This method highlights the effect of institutions in a dramatic fashion, but sheds little light on the exact mechanism by which institutions actually affect output. Given the parsimonious setup of the regressions, this approach may also substantially overestimate the effect of institutions on output. The purpose of this chapter is to add detail to the popular reduced-form estimations and examine different hypotheses regarding the channels through which institutions affect income per capita.

Institutions alone do not produce output. Hence, their effect must be indirect, operating through their impact on either factor accumulation or productivity. Hall and Jones (1999) suggest that just under half of the impact of institutions on output is through their effect on factor accumulation, while the remainder is due to the impact of institutions on productivity. These results contrast with the results of Mankiw, Romer, and Weil (1992), which emphasize the importance of investment in human and physical capital.

In this chapter we combine the approach of Hall and Jones (1999) with that of Mankiw, Romer, and Weil (1992) to explain cross-country per capita income levels. Specifically, we examine whether specifications in which institutions are the sole determinant of output levels (as in Hall and Jones 1999, henceforth HJ) can be improved upon by allowing for the effect of factor inputs. Our hypothesis is that the main contribution of institutional quality to development is through its impact on the accumulation of human and physical capital.

To explore our hypothesis, we regress output on both institutions and factor inputs, allowing the data to determine the elasticity of output with respect to physical and human capital (as in Mankiw, Romer, and Weil 1992, henceforth MRW). We find that the inclusion of a measure of institutions into the MRW specification does yield a significant coefficient on institutions and reduces the residual significantly. The estimates on human capital and physical capital do not change significantly.

Augmenting Hall and Jones's specification with physical factors of production reduces the effect of institutions on output by a whole order of magnitude. Institutions retain only about 15 percent of their explanatory power to account for cross-country income levels as compared to the HJ results. This indicates that at least some part of the contribution of institutions to output is through institution-induced increases in physical factors of production.

Next we ask why is it that institutions affect factor accumulation. Both Hall and Jones (1999) and Mankiw, Romer, and Weil (1992) assume that the elasticities of output with respect to inputs are constant across countries. We propose that the quality of institutions affects the marginal productivities of factors, and hence output shares. A test of this hypothesis shows that once we allow for the factor elasticities to vary across countries, the direct effect of institutions on output vanishes entirely and only the moderating effect of institutions prevails.

Institutions thus truly moderate the effect of human and physical capital on output. Interestingly, while better institutions increase the contribution of capital to output, the result is reversed for the case of human capital. Our results imply that while human capital and institutions by themselves contribute positively to output, institutions matter more for development in low human capital countries. Conversely, the better institutions are, the less human capital matters in explaining differences in per capita income. Our results indicate that while physical capital and institutional quality are complements, human capital and institutions are substitutes in the development process.

The chapter is organized as follows. Section 3.2 reviews and compares the two competing explanations for differences in income levels across countries. Section 3.3 presents our empirical results. It starts with the estimations of the combined model, then moves on to allow for the interaction between institutional quality and factor inputs. Section 3.4 concludes.

3.2 Institutions and Output Levels

3.2.1 Development Accounting in the Absence of Institutions

Let us suppose that output in country *i* is produced according to

$$Y_i = A_i K_i^{\alpha} H_i^{1-\alpha}, \quad (1)$$

where K denotes the stock of physical capital, H is the stock of efficiency units of labor, and A is a measure of labor-augmenting productivity. Defining all magnitudes in per capita terms, $y = Y/L$, $k = K/L$, and $h = H/L$, we can rewrite output per worker as

$$\log y_i = \log A_i + \alpha \log k_i + (1 - \alpha) \log h_i, \quad (2)$$

which highlights that per capita output depends on factor inputs and on the level of productivity.

Hall and Jones (1999) analyze the power of factor inputs extensively to examine if additional factors, such as institutions, are required in order to understand any remaining, unexplained, cross-country income differences. In line with most previous work, their accounting exercise assumes the elasticity of output with respect to each input to be the same for all countries, and takes it to be equal to the value of the capital share in the United States, that is, $\alpha = 1/3$. Hall and Jones then replicate the well-known observation that differences in inputs explain only a small fraction of cross-country differences in output. The Solow residual, obtained when we rewrite (2) as

$$\log A_i = \log y_i - \alpha \log k_i - (1 - \alpha) \log h_i, \quad (3)$$

is in fact the main source of differences in per capita output across countries. Its correlation with per capita income is extremely high, as can be seen from figure 3.1, and differences in the residual explain almost 70 percent of income differences across countries.

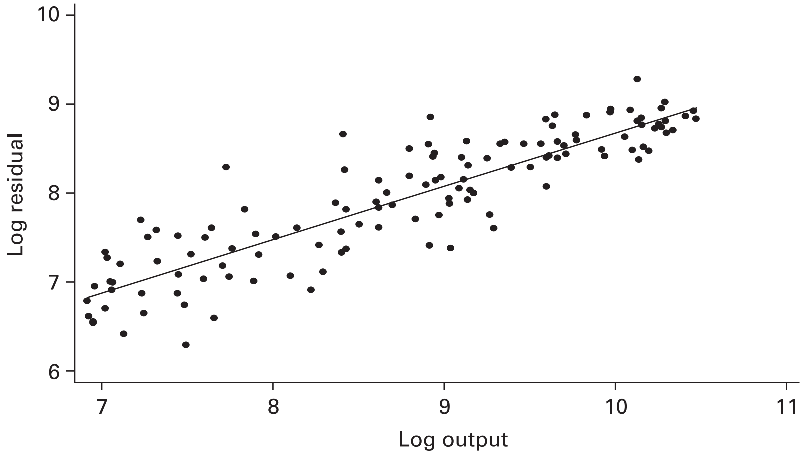


Figure 3.1
Output and the residual: growth accounting.

3.2.2 The Role of Institutions in Development Accounting

The high correlation between the residual and per capita income has led to the interpretation that A is a measure of the level of technology in a country. This implies that richer countries are richer because they use inputs more efficiently. Inspired by the work of North (1990), Hall and Jones (1999) maintain that a major determinant of aggregate productive efficiency in a country is the quality of its institutions. Institutions, they argue, are in fact the fundamental determinant of a country's long-run economic performance, as they determine both productivity and factor accumulation, and they proceed to test this hypothesis.

3.2.2.1 Measuring Institutions and Endogeneity

The first problem encountered when trying to assess the impact of institutions on output levels is that defining and measuring institutions is not straightforward, and the particular definition used may indeed influence the results. Knack and Keefer (1995, 1997) provided early empirical analyses on the growth effects of institutions. One of the novelties of these two papers was to introduce better measures of the institutional framework of countries.[3] They suggested using subjective data, variables constructed from surveys, and expert assessments such as International Country Risk Guide (ICRG) and Business Environment Risk Intelligence (BERI). These include variables such as contract enforceability, rule of law, or risk of expropriation, which can be used to construct indices of government anti-diversion policies.

Hall and Jones (1999) build on the Knack and Keefer indices and define a variable capturing the quality of institutions, which they call social infrastructure. It is the average of the government anti-diversion index and the Sachs-Warner index of trade openness, each of which in turn includes five different categories. The resulting index is measured on a scale of 0 to 1 and assigns a higher value to more desirable outcomes. An alternative approach has been proposed by Acemoglu, Johnson, and Robinson (2001), who measure institutions by the risk of expropriation.

The second concern when seeking to assess the effect of institutions on development is that the level of development itself impacts the quality of institutions. Major efforts have been undertaken to search for good instruments to control for endogeneity, and various correlates of Western European influence have been proposed as instruments for institutional variables. Hall and Jones (1999) use geographic or linguistic characteristics, while Acemoglu, Johnson, and Robinson (2001) employ settler mortality rates in countries colonized by the Europeans.[4]

3.2.2.2 Institutions and Output Levels

The growth accounting exercise performed by Hall and Jones (1999) implies that the correlation between the Solow residual obtained from equation (3) and institutional quality is 0.60. They suggest that the econometric specification that identifies the impact of institutions on income takes the form

$$\log y_i = \gamma_0 + \gamma_1 I_i + \varepsilon, \tag{4}$$

where I is a measure of the quality of institutions or social infrastructure, which differs across countries, and ε is an error term. Hall and Jones estimate equation (4) and find that institutions can account for over twenty-five-fold differences in per capita output. Similar results are obtained by Acemoglu, Johnson, and Robinson (2001) with their alternative measure of institutions and instrument.

The results in these papers have been confirmed by a number of subsequent studies,[5] and the overall evidence is that institutions play an overwhelming role in explaining differences in economic performance across countries. However, the insights from these parsimonious approaches are still limited. Thus far, the literature has treated institutions as black boxes. Nevertheless, it is imperative to understand how institutions work to make countries more (less) productive, and how they impact and interact with factor accumulation. We attempt to address this question in the next section.

3.3 The Effect of Institutions versus Factor Accumulation

3.3.1 Combined Models of Institutions and Factors

The approach of Hall and Jones (1999) and of Acemoglu, Johnson, and Robinson (2001) contrasts sharply with the traditional methods used to identify the determinants of cross-country per capita income, as in the paper by Mankiw, Romer, and Weil (1992), who regress output per capita on factor inputs. Rather than using the value of the capital share in the United States to account for the contributions of the various factors, Mankiw, Romer, and Weil estimate the elasticities of the production function econometrically. In particular, they assume that output in country *i* is produced according to

$$Y_i = AK_i^{\alpha} H_i^{\beta} L_i^{1-\alpha-\beta}, \qquad (5)$$

where L denotes the number of workers. Given our definition of output per worker above and taking logs, we can reexpress the preceding production function as

$$\log y_i = \log A + \alpha \log k_i + \beta \log h_i. \qquad (6)$$

The MRW approach is more general than the development accounting exercise in HJ, because it does not impose ex ante an elasticity of output, nor does it assume constant returns to accumulating factors. However, the former assumes that all countries share identical productivities,[6] an assumption that does not seem to be supported by the results in the latter.

The first question we want to address is whether large differences in the residual remain once we allow for the output elasticities to be determined by the data. Mankiw, Romer, and Weil (1992) and Hall and Jones (1999) use somewhat different data, with the former using per capita income for 1985 and secondary school enrollment rates as a measure of human capital, and the latter output per worker in 1988 and the stock of human capital. In order to render comparable results, we use the HJ output data in all specifications. Human capital data are either the original MRW or HJ, again to generate comparable results.

Table 3.1
Institutions in the augmented Solow model. Dependent variable: Log output per worker

	HJ	MRW	Combined model 1	Combined model 2
Institutions	5.142***		1.089***	.698**
	.343		.235	.249
Log *h* (enrollment rate)		.110	.099	
		.072	.069	
h (human capital stock)				.141
				.087
Log *k*		.603***	.525***	.562***
		.040	.048	.037
N	127	111	111	127
R-squared	0.58	0.91	0.92	0.91
Root MSE	0.70	0.33	0.31	0.33
Correl (A, Y/L)	0.89	0.30	0.27	0.31
Correl (A, Institutions)	0.60	0.25	0.01	0.00

Note: MRW specification without steady-state assumptions. Specifications in columns 2 to 4 are two-stage least squares regressions, where institutions are instrumented as in HJ. Subscripts ***/**/* denote 1 percent/5 percent/10 percent significance levels.

Our measure of institutions will be social infrastructure, as constructed by Hall and Jones. In all regressions, this variable is instrumented as in HJ. Aside from a reduced sample size, Acemoglu, Johnson, and Robinson (2001) instruments yield the same results.

Table 3.1 juxtaposes the basic empirical results. The first column reports the results of HJ, where institutions alone determine output levels. The second column presents a regression of output per capita on factor inputs, a general version of MRW. In their paper,[7] Mankiw, Romer, and Weil (1992) obtain a somewhat lower elasticity of output with respect to physical capital and a higher one for human capita, 0.48 and 0.23, respectively. However, the MRW estimates are within the 10 percent confidence interval implied by the estimates in column 2.

The bottom two lines of table 3.1 report the correlation of the *residual* with output per capita and institutions for the two approaches. In the HJ setup, this is the Solow residual obtained from equation (3); for the MRW specification, it is the residual resulting from the regression equation. The augmented Solow model provides a very good fit for the data. In particular, the correlation between the residual and output levels drops from 0.89 to 0.30, indicating that the estimates for the elasticities of output give a much better picture than imposing $\alpha = 1/3$.

Nevertheless, the resulting residual is still highly correlated with institutions (0.25).

The natural extension is to estimate a production function that includes both inputs and institutions. Suppose output is produced according to

$$Y_i = A_i K_i^{\alpha} H_i^{\beta} L_i^{1-\alpha-\beta}, \tag{7}$$

with the level of productivity, A_i, being a function of institutions. In particular, we stipulate that

$$A_i = A e^{\delta I_i}. \tag{8}$$

Output per capita is then a function of factor inputs, institutions, and a residual, taken to be the level of technology, and we can express it as

$$\log y_i = \log A + \alpha \log k_i + \beta \log h_i + \delta I_i + \varepsilon. \tag{9}$$

The third and fourth columns in table 3.1 report the results of the combined model in (9), using secondary school enrollment rates as in MRW, and the stock of human capital as calculated by Hall and Jones (1999). Following HJ, we introduce institutions into the regressions without taking logarithms.

The results from the regressions are surprisingly good. All factors have the expected sign, and the estimates are quite robust across specifications. In particular, the coefficient on institutions is positive and significant, suggesting that Hall and Jones (1999) could have also included factors of production, or that Mankiw, Romer, and Weil (1992) could have included institutional differences to derive more accurate estimates of contributions of physical inputs to explaining per capita income differences across countries.

Once capital and labor are included in the regression, the estimate for the effect of institutions on growth, although still positive and significant, drops by a whole order of magnitude. Institutions can now account for only between 15 and 20 percent of the variation in per capita incomes. At the same time, the inclusion of institutions shows that the elasticities of output with respect to human and physical capital barely change as compared to the basic MRW specification in column 1. These elasticities are somewhat lower in the specification with institutions.

Neither of the combined models represents a significant improvement over the MRW specification in terms of the R^2. In order to assess

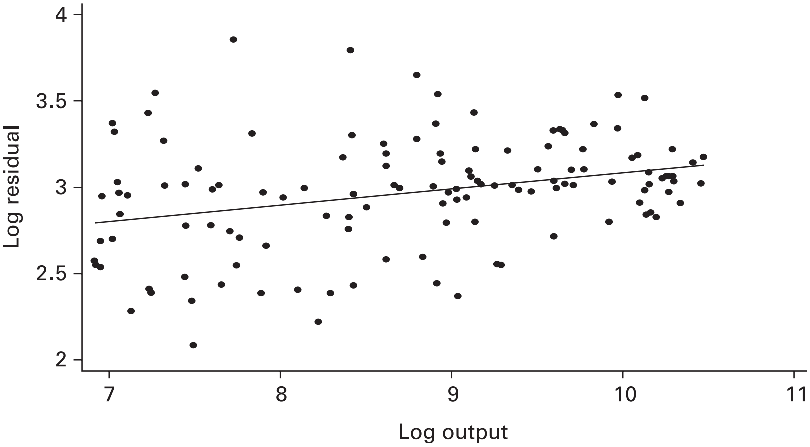


Figure 3.2
Output and the residual: the combined model.

the effectiveness of our specification, we examine how the combined models fare in terms of the Solow residual. The last two columns of table 3.1 show that the inclusion of institutions has an important effect: the correlation between the residual and output falls by 10 percent (column 3), while the correlation between the residual and institutions disappears entirely.

Figures 3.2 and 3.3 plot the residuals obtained from the combined model (model 1) and either per capita output or institutions. Although there is still some correlation with output, there is none with the measure of institutional quality. Our specification thus purges the residual of its institutional component, rendering it a true statistical residual due to measurement errors or violations of the structural assumptions in the Solow growth accounting framework (such as constant returns to scale).

3.3.2 Direct and Indirect Effects of Institutions

The regressions in table 3.1 imply that both institutions and factor accumulation matter for output levels. However, institutions by themselves do not produce output; their effect should actually be captured by the catalytic effect institutions have on the factors of production. In this section, we seek to understand how much of the variation in output is accounted for by the direct (and abstract) impact of institutions,

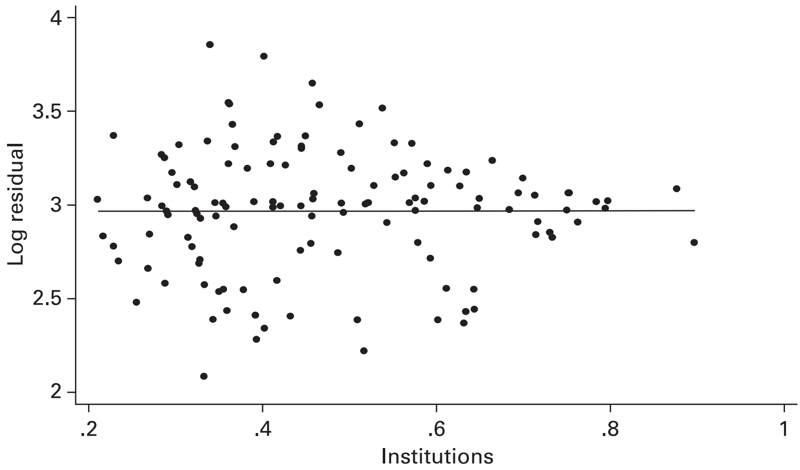


Figure 3.3
Institutions and the residual: the combined model.

Table 3.2
Direct and indirect contributions of institutions to per capita income

	Dependent variable				Combined contribution of H, K**
	$\alpha \log \frac{K}{L}$	$\beta \log \frac{H^*}{L}$	$\log A$	Institu-tions	
HJ	2.416	0.896	1.830		3.312
MRW	3.478	0.767	0.897		4.245
Combined model 1	3.745	0.325		1.072	4.070
Combined model 2	4.222	0.196		0.724	4.418

* H refers to MRW and HJ human capital variables, respectively, logged when necessary.
** Refers to the sum of columns 1 and 2. Coefficients in all intermediate regressions had significance levels of over 1 percent.

as opposed to the indirect effect of institutions that works through factor inputs.

Table 3.2 reports the direct and indirect effects of institutions by regressing inputs on institutions. The indirect effects were obtained by running the regression $x = \gamma + \gamma_1 Institutions + \varepsilon$, where x is either k, h, or A. The direct effect of institutions is the coefficient δ in equation (9), normalized so that the sum of coefficients is equal to 5.142, that is, to the coefficient on institutions obtained when these are the sole explanatory variable.

In row 1 we assess the contribution of inputs under the assumption that $\alpha = 1/3$ and $\beta = 1/3$, as in HJ. The contributions of inputs together with the residual, A, add up to 5.142, which is the total contribution of institutions as measured by the coefficient in table 3.1.

In the HJ specification in row 1, factors of production contribute about 64 percent to output, whereas the contribution of the Solow residual, A, accounts for the remaining 36 percent of the variation in output levels across countries. That is, factor accumulation plays a limited role, accounting for less than two thirds of output differences, and institutions seem to mainly affect aggregate productivity.

The rest of the table repeats this exercise for the Solow model augmented by Mankiw, Romer, and Weil (1992) and for our combined models. The second line uses the production elasticities obtained by Mankiw, Romer, and Weil, namely, $\alpha = 0.48$ and $\beta = 0.23$. With these elasticities, the role of factor accumulation becomes much more important: 82 percent of the effect of institutions occurs through human and physical capital accumulation. Similar results are obtained when we use the elasticities obtained from the combined model. Again, the main role of institutions is to encourage factor accumulation, with the direct effect accounting for between 14 and 21 percent of the overall impact.

The other major difference between the development accounting exercise and the results using estimated elasticities concerns the relative importance of physical and human capital accumulation. Imputing the value of α results in a contribution of institutions through human capital, which is almost a third of the total contribution of factors. The augmented Solow model (with and without institutions) features a much more important effect through physical capital, with only a small effect occurring through human capital accumulation (between 4 and 18 percent of the total contribution of factors).

3.3.3 *The Interaction between Institutions and Factors of Production*

Our preceding discussion implies that physical and human capital react rather differently to improvements in institutional quality. A reason for this could be that the elasticities of output with respect to factor endowments, and hence factor returns, depend on a country's institutional quality. That is, given the level of technology, the effect of a given stock of (physical or human) capital on output depends on the quality of a country's institutions.

While Mankiw, Romer, and Weil (1992) assume the level of technology to be common across countries and allow the output elasticities to be determined by the data, Hall and Jones (1999) impute the elasticities and allow technology to vary across countries. What both approaches share is the assumption that *factor shares are constant across countries*. Yet, the data cast doubt on this assumption. A number of recent studies document the extensive differences in factor shares across countries and over time (see Gollin 2002; Harrison 2002; Bentolila and Saint-Paul 2003). Such evidence raises the question of whether allowing the output elasticities to vary across countries can improve our understanding of income differences. If we assume that the elasticity of output with respect to the various inputs differs systematically across countries, we must propose a mechanism by which such differences arise. Here we stipulate that institutions crucially affect the productivity of factors and their shares in output.

In order to estimate the extent to which differences in output elasticities are driven by institutional differences, we further modify the production function used by Mankiw, Romer, and Weil (1992) and assume that output in country i is produced according to

$$Y_i = A_i K_i^{\alpha_i} H_i^{\beta_i} L_i^{1-\alpha_i-\beta_i}. \tag{10}$$

We propose that both the level of aggregate productivity and the elasticities of output with respect to the two inputs depend on the quality of institutions, I. As before, productivity is given by $A_i = Ae^{\delta I_i}$. Concerning the elasticities, we assume a simple linear formulation, whereby $\alpha_i = a + \alpha I_i$ and $\beta_i = b + \beta I_i$. We can then write output per capita as

$$\log y_i = \gamma_0 + \gamma_1 I_i + \gamma_2 \log k_i + \gamma_3 I_i \log k_i + \gamma_4 \log h_i + \gamma_5 I_i \log h_i, \tag{11}$$

with institutions affecting output through γ_1, γ_3, and γ_5, which capture both the direct effect of institutions on total factor productivity (TFP), which is constant across countries, as well as the effect of institutions on the input elasticities.

Table 3.3 reports the results of the estimation. Two surprising results emerge. First, the coefficient γ_1 is insignificant in both specifications. Institutions no longer affect TFP, which contrasts with the results in table 3.1. Rather, the effect of institutions on output is now entirely captured by its effect on the productivity of inputs. The alternative interpretation is that the HJ specification loses its validity once the effect

Table 3.3
Institutional effects on labor and capital productivities. Dependent variable: Output per worker (2SLS)

	Augmented model 1	Augmented model 2
Institutions	−.036	−1.376
	1.679	1.705
log *k*	.438***	.367***
	.095	.089
Institutions ∗ log *k*	.206	.471**
	.200	.210
log *h* (enrollment rate)	.300*	
	.161	
Institutions ∗ log *h*	−.514	
	.396	
h (human capital stock)		.776**
		.310
Institutions ∗ *h*		−1.297**
		.517
N	111	127
R-squared	0.93	0.91
Root MSE	.31	.33
Correl (A, Y/L)	0.27	0.30
Correl (A, Institutions)	0.00	0.00

Note: HJ and MRW specifications instrumented for institutions as in Hall and Jones 1999. Subscripts ***/**/* denote 1 percent/5 percent/10 percent significance levels.

of institutions on factor inputs and factor shares has been included. The second result is no less surprising: better institutions seem to increase the productivity of physical capital, but *reduce* that of human capital. Institutions increase the elasticity of output with respect to physical capital and labor, and reduce the elasticity with respect to human capital. Human capital and institutions by themselves have a positive impact; however, institutions matter more for growth in low human capital countries. The reverse way of thinking about this relationship is that the more human capital a country has, the less important institutions are.

Our results suggest that institutions and physical capital are complements. On the other hand, institutions and human capital are substitutes, in the sense that, given the stock of capital, a certain level of output can be produced either with good institutions and low human capital, or with poor institutions but a very educated labor force.

3.4 Conclusions

In this chapter, we provide a preliminary exploration of how institutions may directly affect per capita output. Our results indicate that the largest impact of institutions is through their effect on the factor productivity. While institutions have uniformly positive effects on the productivity of physical capital, our regressions indicate that institutions and human capital are substitutes. This can be interpreted as saying that institutions matter most for countries with low levels of human capital and least for those with high levels of educational attainment.

Two main implications emerge. First, the results provide evidence for an over-investment in human capital in some countries, raising the question of whether traditional justifications for public provision of education, based on a high social return to education, are still valid. Second, they indicate—in contrast to the Hall and Jones approach—improving institutions is not sufficient to generate increases in income levels. Since the main role of institutions is to increase the productivity of capital, improving institutions in countries with a very low level of investment will have only a small impact on output.

Last, our analysis has been static. Yet the results have important dynamic consequences. If better institutions increase the productivity of capital, they will create investment incentives, and hence foster future output. In fact this could be a possible explanation for the strong correlation between physical capital and institutions found in the data.

Notes

We thank seminar participants at Munich University, the Ifo Institute, and the Venice Conference on Institutions and Growth, as well as Philippe Aghion, Sascha Becker, and Stefan Klasen for helpful comments. Eicher thanks the German Science Foundation for financial support, and Munich University and the Ifo Institute for their hospitality. This research was conducted while García-Peñalosa was visiting CESifo.

1. See Caselli 2003 for a recent survey of development accounting.

2. See Knack and Keefer 1995, 1997; Hall and Jones 1999; Acemoglu, Johnson, and Robinson 2001, 2002; Easterly and Levine 2002.

3. Before Knack and Keefer 1995 and 1997, secure property rights / good institutions were proxied by the Gastil Index of political and civil liberties, and frequency of revolutions, coups, and political assassinations. However, results from such regressions were less satisfactory.

4. A number of other concerns have been raised regarding the simple cross-country regressions that we, and the original papers we discuss, undertake. These range from

issues of data quality to model uncertainty and parameter heterogeneity. See Durlauf, Johnson, and Temple 2004 for a discussion.

5. See, among others, Kaufman, Kray, and Zoido-Lobatón 1999; Easterly and Levine 2002; Grigorian and Martinez 2002; and Rodrik, Subramanian, and Trebbi 2002.

6. In their specification of the output levels regression equation, Mankiw, Romer, and Weil also assume that all countries are in their steady state, and they write the level of output as a function of investment shares, which in turn determine the steady state levels of human and physical capital. Our formulation is more general and simply uses factor endowments as the determinants of income levels.

7. The coefficients we report are implied by the growth regressions in MRW, which take into account that economies may not be at their steady states.

References

Abramowitz, Moses. 1956. "Resource and Output Trends in the United States since 1870." *American Economic Review* 46: 5–23.

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91, no. 5: 1369–1401.

Acemoglu, Daron, Simon Johnson, and James Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117: 1231–1291.

Bentolila, Samuel, and Gilles Saint-Paul. 2003. "Explaining Movements in the Labor Share." *Contributions to Macroeconomics* 3, no. 1: 1013–1065. Berkeley Electronic Press.

Caselli, Francesco. 2003. "The Missing Input: Accounting for Cross-country Income Differences." Forthcoming in *Handbook of Economic Growth*, ed. Philippe Aghion and Steven Durlauf. Amsterdam: North-Holland.

Durlauf, Steve N., Paul A. Johnson, and Jonathan R. W. Temple. 2004. "Growth Econometrics." University of Wisconsin working paper.

Easterly, William, and Ross Levine. 2002. "Tropics, Germs and Crops: How Endowments Influence Economic Development?" NBER working paper no. 9106.

Friedman, Eric, Simon Johnson, Daniel Kaufman, and Pablo Zoido-Lobaton. 1999. "Dodging the Grabbing Hand: Determinants of Unofficial Activity in 69 Countries." SITE working paper no. 148.

Gollin D. 2002. "Getting Income Shares Right." *Journal of Political Economy* 110, no. 2: 458–475.

Grigorian, David, and Albert Martinez. 2002. "Industrial Growth and the Quality of Institutions: What Do (Transition) Economies Have to Gain from the Rule of Law?" World Bank working paper no. 2475.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output than Others?" *Quarterly Journal of Economics* 114: 83–116.

Harrison, Ann E. 2002. "Has Globalization Eroded the Labor Share?" Mimeo., University of Berkeley.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999. "Governance Matters." World Bank policy research paper no. 2196.

Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-country Tests Using Alternative Measures." *Economics and Politics* 7: 207–227.

Knack, Stephen, and Philip Keefer. 1997. "Why Don't Poor Countries Catch Up? A Cross-national Test of an Institutional Explanation." *Economic Enquiry* 35: 590–602.

Lane, Jan-Erik, and Dominic Rohner. 2003. "Institution Building and Spillovers." Public Policy Programme working paper 40, National University of Singapore.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1999. "The Quality of Government." *Journal of Law, Economics and Organization* 15: 222–279.

Mankiw, Gregory N., David Romer, and David Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107: 407–437.

North, Douglas. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2002. "Institutions Rule: The Primacy of Institutions over Integration and Geography in Economic Development." IMF working paper no. 02/189.

Sachs, Jeffrey D. 2001. "Tropical Underdevelopment." NBER working paper no. 8119.

Sachs, Jeffrey D. 2003. "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income." NBER working paper no. 9490.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–118.

Summers, Robert, and Alan Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons 1950–1988." *Quarterly Journal of Economics* 106: 327–368.

4 Regulation and Economic Performance: Product Market Reforms and Productivity in the OECD

Giuseppe Nicoletti and
Stefano Scarpetta

4.1 Introduction

Over the past decade, a large number of theoretical and empirical studies have highlighted the channels through which institutional settings can affect crucial dimensions of economic performance. The large majority of this literature has focused on the effects of different institutional settings between industrial and developing countries on their economic performances. Yet even within the industrial economies the gaps in income per capita remain large. For example, in the early 2000s, per capita incomes were between 20 and 30 percent lower in Japan and the European Union than in the United States.[1] Moreover, growth episodes differed across industrial countries. For example, the acceleration in productivity growth observed in the United States since the mid-1990s has been shared only by a limited number of—generally small—OECD countries.[2]

Such evidence raises the question of why the process of economic convergence observed for some decades in the postwar period has been halted and even partially reversed. Is this a temporary phenomenon associated with the spread of the information and communication technology and its associated exuberance in investment, or is it rather a signal of more long-term forces that deserve a closer look by policymakers? Recent growth theory suggests that catch-up in real incomes is not mechanical. Rather it depends on factors such as market imperfections, agglomeration effects, or differences in endowments and institutional settings (see, e.g., Barro and Sala-i-Martin 1998, and chapter 1 in this volume). Indeed, the institutional setting in which firms operate is likely to affect both their level of employment and labor productivity, and through those aggregate output.

An essential component of the institutional framework is the extent of regulation in labor and product markets. A substantial literature has examined the impact of labor market policies on employment in OECD countries,[3] but the macroeconomic effects of regulations in the product market have received little attention. Such lack of attention is surprising for two reasons. First, the large cross-country variability in policy choices in this area and the magnitude of reforms aimed at promoting competition and productivity growth imply that such policies are good candidates for explaining cross-country differences in economic performance. Second, the effects of product market reforms on productivity and prices at the industry level have been well documented, but little is known about their macroeconomic implications. Yet as reforms spread to an increasing number of industries and increasingly include changes in general-purpose regulations, such as administrative procedures, their macroeconomic repercussions are likely to be significant.[4]

In this chapter, we assess the effect that the reforms of product market regulations undertaken by OECD countries over the past two decades have had on two determinants of growth, investment, and multifactor productivity. The aim of the chapter is twofold. First, we review the theoretical literature, taking a bottom-up approach to go from microeconomic arguments to macroeconomic outcomes. Second, we bring together the results obtained in some of our empirical work over the past few years with the aim of assessing the overall impact of regulatory reforms. Whenever possible, we compare our results with those found in other empirical studies.

We argue that product market regulations affect at least one of the two key components of the output gap across countries: the labor productivity gap. The latter can be further decomposed into the capital-labor ratio and the level of multifactor productivity. Existing theories on the microeconomics of regulation tell us that regulations can affect both of these factors through its impact on investment, managerial behavior, and the incentives to adopt new technologies and innovate. The effects are however complex and, at times, contradictory. For example, the regulation of natural monopolies can hinder entry of new firms, but it may also prevent abusive use of market power. The signs of the overall effects on both capital accumulation and productivity are hence a priori ambiguous.

In order to assess empirically the strength of the various effects, we draw on three studies that examine the impact of regulations on investment (domestic and foreign) and multifactor productivity. The studies use a consistent set of quantitative indicators, recently developed at the OECD, that measure differences in product market reform across countries, at both the economy-wide and the industry level. We look especially at regulations in non-manufacturing, where restrictions to competition and reforms to alleviate them have been more extensive. Throughout the chapter, we concentrate on the effects of policies aimed at strengthening private governance (e.g., through privatization) and opening up access to markets where competition is economically viable, though the empirical analyses we discuss often also look at the interactions between product market reforms and policy and institutions in other markets—in particular, the labor market.

The reported empirical results are based on panel estimates of fairly general specifications of investment and productivity across countries, time, and (when possible) industries. To reduce the risk of parameter heterogeneity, we focus on a subset of relatively similar OECD countries, and we account for differences in the response of sectoral productivities to regulations. All our crucial policy variables have a wide variability in all the dimensions of the panel, including over time, reflecting the significant privatizations and liberalizations that were implemented by OECD countries over the past three decades. We also pay particular attention to data and specification issues, such as sensitivity to outliers, changes in the set of explanatory variables, and the presence of nonlinearities.

Empirical results are reported in a summary form, concentrating on the implied quantitative implications for macroeconomic outcomes. They suggest that pro-competitive reforms tend to increase both investment and productivity. Through both these channels, product market reforms can lead to higher growth in GDP per capita. Clearly, quantitative assessments heavily depend on modeling choices and sample coverage. In particular, despite the focus on a homogeneous set of economies, unaccounted differences in the way they may respond to changes in regulation may bias the estimated coefficients. Therefore, implications based on such coefficients should be interpreted with caution, especially in view of the fact that the variables we consider are closely interdependent, and the global effects of reform can hardly be expressed as the sum of the effects estimated for each of them separately. Nonetheless, we believe that the reported results are sufficiently robust to provide at least an indication of the direction of the effects and possibly the magnitudes involved.

Our chapter contributes to recent literature on the institutional determinants of economic performance. In contrast to other empirical studies, we try to open the black box that is the link between institutions and growth by focusing on a very specific type of institutions. This allows us to assess some of the channels through which these institutions affect the determinants of growth. Two such channels, foreign direct investment and catch-up of industry-level multifactor productivities to OECD best practice, introduce an open economy element to the picture that is often missing in analyses of institutions and growth.

The plan of the chapter is as follows. In section 4.2 we provide a brief review of the literature on the economics of regulation and examine the main channels through which product market policies can affect investment and productivity, and consequently output. Section 4.3 describes the indicators of product market regulation used in the empirical analysis, as well as the trends in regulation observed in the OECD over the past two decades. We then discuss the econometric approaches—and their drawbacks—used in the empirical analyses that is summarized in section 4.4. We provide the main concluding remarks in section 4.5, where we also indicate possible avenues for further research.

4.2 Regulation and the Determinants of Output Growth and Income per Capita

4.2.1 Output Growth and Income per Capita

A country's level of output per capita can be mechanically decomposed into the average number of hours worked by the workforce multiplied by the level of labour productivity. Differences in per capita output levels across countries can then be explained by the extent of utilization of labor and by differences in labor productivity. Figure 4.1 presents a decomposition of real income levels relative to the United States in some OECD countries, and indicates that discrepancies in both hours worked and output per hour worked remain large even within this group of countries.

Differences in the institutional setting in which firms operate, and in particular differences in the extent of regulation in labor and product markets, are likely to affect both employment and labor productivity, and hence are candidate explanations for the divergent economic per-

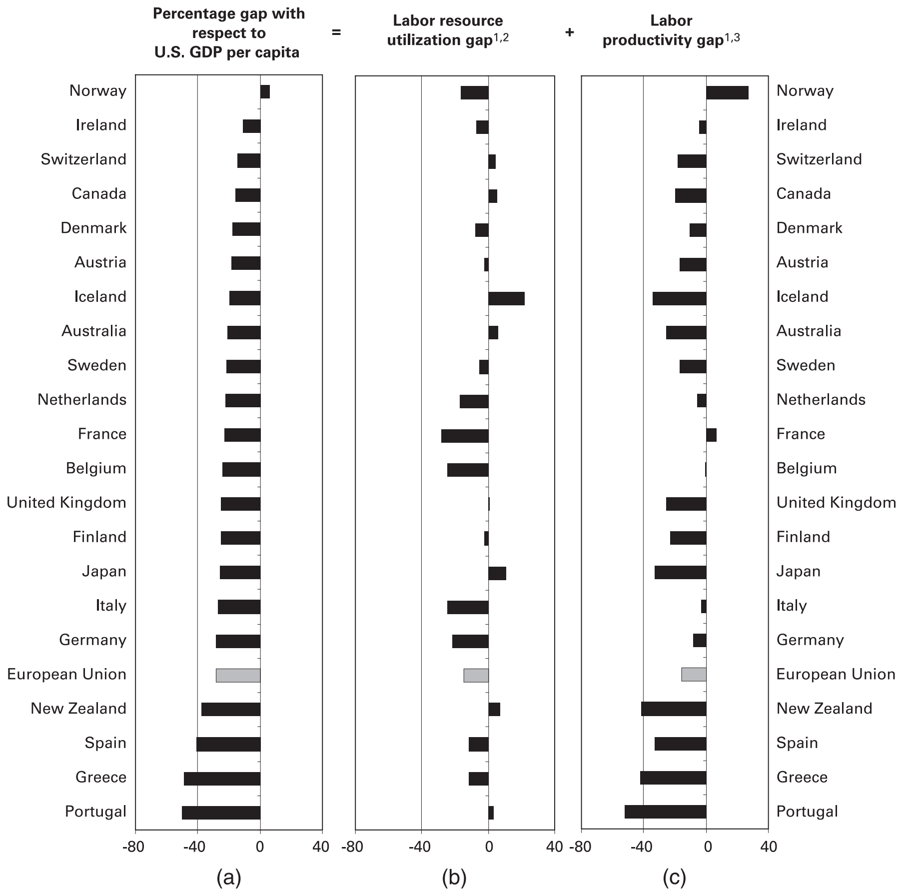


Figure 4.1
The sources of real income differences, 2002 (2000 Purchasing Power Parities).
Source: OECD Productivity Database.
1. Percentage gap with respect to the United States level.
2. Labor resource utilization is measured as trend total number of hours worked divided by population.
3. Labor productivity is measured as trend GDP per hour worked.
4. European Union except Luxembourg.

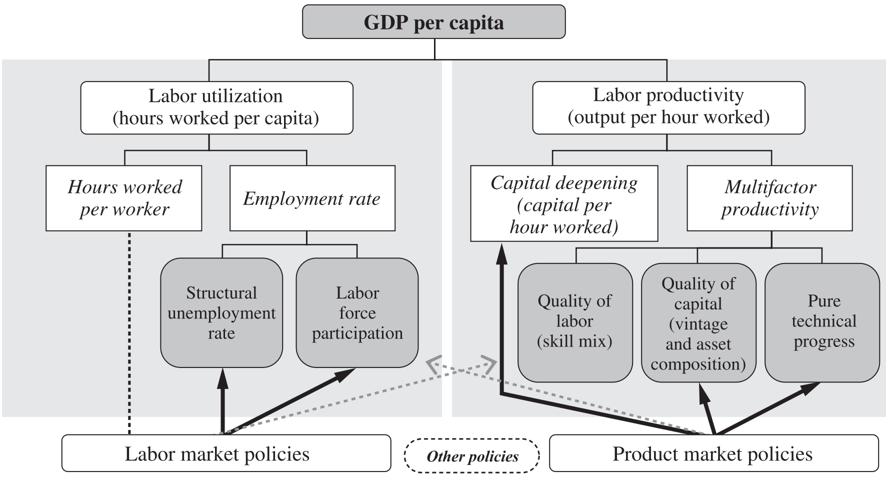


Figure 4.2
The proximate determinants of GDP per capita.

formances. Figure 4.2 illustrates the way in which such regulation can impact output levels. Labor market policies are a major determinant of both the employment rate and the number of hours worked, and the macroeconomic effects of labor market reforms have been extensively studied. However, little attention has been devoted to the impact of product market reforms despite the fact that recent policy changes in European product markets have sometimes been deeper than reforms in labor markets, as noted by Koedijk and Kremers (1996).

Product market regulations are likely to have macroeconomic effects mainly through their impact on labor productivity growth.[5] Labor productivity growth can be decomposed into changes in the capital-labor ratio—a measure of capital deepening—and changes in multifactor productivity (MFP), whose "broad" definition measures the residual growth in output that cannot be attributed to changes in the quantity of labor and capital used in production.[6] Changes in broad MFP growth can be further decomposed into changes in the "quality" of factor inputs—so-called embodied technological progress—and a residual factor that accounts for organizational changes and disembodied technological progress. Available information suggests that both improvements in the human capital of workers and shifts to more productive technologies and forms of organization have contributed to changes in productivity growth.[7] In particular, the shift toward infor-

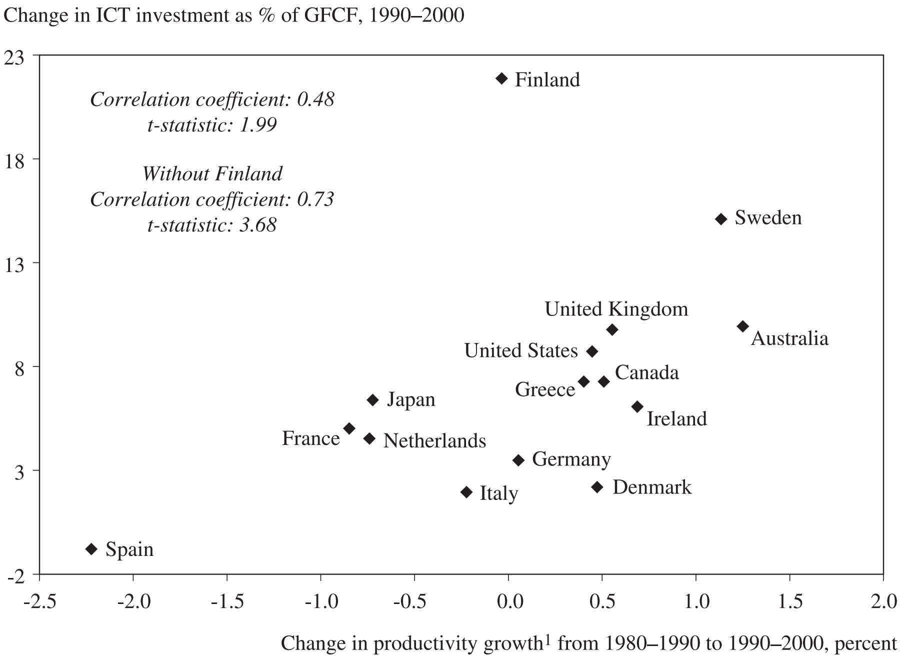


Figure 4.3
Pickup in hourly labor productivity growth and increase in information and communication technology investment.
Source: OECD Productivity Database.
1. Labor productivity is defined as output per hour worked.

mation and communication technology (ICT) equipment has been a strong factor behind the acceleration of productivity growth in some countries. At the beginning of the century, the share of ICT equipment in total investment was higher than in the previous decade in all countries for which data are available except Spain and Italy, and OECD countries' experience suggests a positive correlation between labor productivity acceleration and the increase in ICT investment over the 1990s (figure 4.3), reflecting both a surge in the level of investment (and in the capital-labor ratio) as well as an increase in the productivity of capital.[8]

Even narrowing down the definition of product market regulations to policies that affect competition among producers, it is easy to see that their effects are complex. They can curb competitive pressures and hinder (or prevent) entry of new firms in potentially competitive markets, but they can also favor competition in certain industries by ensuring that market power in natural monopoly segments is not used abusively and by providing the correct incentives to market participants. Sections 4.2.2 and 4.2.3 examine the main channels through which regulatory reforms in product markets impact on investment and multifactor productivity, and hence on productivity growth. In doing so, we will think of three broad categories of policies—barriers to domestic competition, barriers to foreign trade and investment, and cumbersome administrative procedures—all of which are likely to affect labor productivity.

4.2.2 Regulation, Competition, and Investment

4.2.2.1 Capital Spending

Changes in the regulatory environment may affect both domestic and foreign investment. Regulatory reforms of domestic product markets often take the form of a reduction of entry barriers in markets where competition is viable. As argued by Blanchard and Giavazzi (2003) and Alesina et al. (2003), a reduction in barriers to entry will increase the number of firms, leading to a decrease in the markup of prices over marginal costs and, therefore, of the shadow cost associated with capital and output expansion. This is likely to expand activity levels and stimulate capital formation. At the same time, less red tape and lighter regulatory burdens lower the costs of adjusting the capital stock, thereby boosting the capacity of firms to react to changes in fundamentals by expanding their productive capacity.

Regulation can affect investment through two additional channels, which can potentially offset the previously mentioned mechanisms. One consists of the Averch and Johnson (1962) effect. Certain regulations, such as ceilings on the rate of return on capital, encourage firms to over-accumulate capital in order to increase their overall remuneration. The removal of the ceilings will hence result in a downward adjustment of the capital stock. Last, ill-designed deregulation can fail to provide the right incentives to expand capacity. In particular, certain sectors such as network industries have been subject to a redesign of price regulation (e.g., from rate of return to price caps or inappropriate access pricing regimes) or to changes in industry structure (e.g., vertical separation of networks from service provision). Moreover, ill-designed deregulation may allow public enterprises to raise barrier to new entrants and curb competition.[9]

4.2.2.2 Foreign Direct Investment

The impact of regulation on foreign direct investment (FDI) is particularly relevant, because FDI not only raises the capital-labor ratio but also may be an important source of technological transfer. While the reasons to expect a negative effect of FDI restrictions on foreign investment are obvious, the role played by domestic regulations is a priori more ambiguous. Indeed, by raising production costs or entry barriers, domestic regulations can affect FDI in conflicting ways. On the one hand, regulations that increase costs in the host country can deter FDI by lowering its expected rate of return if the foreign subsidiary is used as a platform for reexporting final or intermediate goods back home or to other less regulated countries. On the other hand, if FDI aims at accessing the local market, cost-increasing regulations in the host country may encourage FDI because the foreign affiliate can take advantage of the production structure of the parent firm, which may be more efficient than in local firms if regulations in the investor country are more pro-competitive.[10]

Similar conflicting influences can be exerted by regulations that raise entry barriers in host countries. Such entry barriers clearly deter "greenfield" FDI aimed at establishing new firms or creating new production plants. However, by endowing local firms with market power, they can actually encourage inward FDI aimed at acquiring existing local firms, or merging foreign parents with these firms.

4.2.3 Regulation and Multifactor Productivity

There are basically three ways in which MFP improvements can be achieved: eliminating slack in the use of resources, adopting more efficient technologies, and increasing innovative effort. By affecting the incentives to innovate and improve efficiency, regulations that promote product market competition can have important effects on MFP performance.

4.2.3.1 Reducing Slack

There is an increasing consensus that product market policies that promote entrepreneurship and competition may contribute to shift the (country-specific) efficiency frontiers by raising the efficiency with which the inputs are used.[11] Regulatory policies may, in particular, help to eliminate X-inefficiencies through a number of different channels.[12] For example, competition creates greater opportunities for comparing performance, making it easier for the owners or the market to monitor managers. Competition is also likely to raise the risk of losing market shares at any given level of managerial effort, inducing managers to work harder so as to avoid this outcome.[13] Moreover, business-friendly regulations make it easier to implement efficiency improvements by reshuffling resources within and across firms.

It should be stressed that theoretical predictions of the effects of greater competition on managers' incentives are often "subtle and ambiguous" (Vickers 1995). Models using explicit incentives under information asymmetry do not lead to clear-cut implications (see, e.g., Holmström 1982), while intertemporal models using implicit (i.e., market-based) rewards suggest a positive link between competition and managerial effort if productivity shocks are more correlated across competitors than managerial abilities (Meyer and Vickers 1997). But competition could also lead to more slack if managers are highly responsive to monetary incentives, as the scope for performance-related pay is reduced (Scharfstein 1988).

4.2.3.2 Technology Diffusion and Adoption

The opening up of markets and increased competitive pressures provide both opportunities and strong incentives for firms to upgrade their capital stock and adopt new technologies to reach frontier production techniques. New entrants into a sector are likely to bring new vintages of technology, often embodied in capital goods, and this gives incumbents the opportunity to upgrade their capital through imitation. Moreover, the threat of losing market share vis-à-vis these more advanced competitors motivates existing firms to adopt new technologies and upgrade their machinery (Schmidt 1997; Aghion and Howitt 1998).

Technology diffusion may also be induced by spillovers from increased trade and foreign direct investment resulting from the reduction of border and nonborder barriers. Theoretical models highlight a number of mechanisms: the transfers of technology between parents and subsidiaries of multinationals, learning externalities for the host-country labor force, and spillovers related to the provision of high-technology intermediate inputs from the origin country. While the empirical evidence is mixed, recent cross-country and microeconomic studies suggest that these effects are significant, indicating that an increase in FDI is likely to be associated with higher levels of multifactor productivity.[14]

4.2.3.3 Innovation

There are three basic mechanisms that affect the incentives that firms have to innovate.[15] First, the attempt to acquire a competitive edge on rival firms often results in a stronger innovation effort. This can be due to the fact that in highly competitive markets, small price differences have big effects on market shares; competition thus raises the "bang for the buck" from cost-reducing productivity enhancements providing greater incentives to innovate. Alternatively, Klette and Griliches (2000) note that incumbents might be pushed to innovate in order to preempt rivals. Second, easier entry into (and exit from) innovative markets fosters market testing of new ideas and the process of "creative destruction."

Last, Aghion et al. (2001) model the pro-innovation impact of competition by noting that stronger competition may force managers to speed up the adoption of new technologies in order to avoid loss of control rights due to the risk of bankruptcy. Even when firms have similar cost structures (the case of "neck-and-neck" competition) and technological progress is more gradual, stronger competition may induce firms to increase R&D investment (conditional on the level of protection of intellectual property rights) in order to acquire a lead over their rivals. This channel, however, may be "bell-shaped" in the sense that the pressure for more innovation may be the highest at intermediate levels of competition (see Boone 2000a, 2000b; and Aghion et al. 2002), since "too much" competition may dissipate innovation rents curbing incentives to innovate.

4.2.4 Summing Up

The effect of product market reform on investment, both domestic and foreign, is ambiguous because there is a trade-off between the positive incentives generated by lower entry barriers or easier administrative procedures and the negative ones due to the removal of regulations that had led to over-investment or market power for foreign firms. Concerning MFP, there are strong arguments suggesting that increased competitive pressures resulting from product market reform are likely to stimulate productivity. However, offsetting factors and uncertainties inherent in the channels described earlier—in particular, the ambiguous effect on managerial incentives—indicate that the strength, if not the direction, of the link between product market competition and productivity performance remains an empirical issue.

It is also important to note that the investment and productivity gains induced in certain industries may lead to productivity gains for the economy as a whole. The domestic business environment is likely to be particularly important for efficiency in utilities and service industries, where competition from abroad is weaker and a difficult balance has to be struck between regulations and market forces due to market imperfections (e.g., in network industries). Enhancing competition in these industries can provide a "double dividend" because it may both increase the direct contribution of non-manufacturing to overall productivity growth and contribute to overall productivity growth indirectly, via improvements in the productivity of industries that use non-manufacturing products as intermediate inputs.[16]

4.3 Data and Econometric Specifications

4.3.1 Patterns of Product Market Reform

Most OECD countries have implemented sweeping regulatory reforms over the past two decades, with the aim of promoting entrepreneurship and competition. However, they did so at different speeds and to a different degree. The main elements of these reforms were (1) privatization, (2) entry and price liberalization in potentially competitive domestic markets, (3) pro-competitive regulation of natural monopoly markets (e.g., by regulating access to networks), and (4) further liberalization of international trade and foreign direct investment.[17]

To gauge the extent of these reforms, we use a set of cross-country quantitative indicators of regulatory reform. The indicators measure to what extent competition and firm choices are restricted where there are no a priori reasons for government interference, or where regulatory goals could plausibly be achieved by less coercive means. The indicators are constructed to measure regulation in either particular areas of the economy, specific industries, or the overall economy. Many of them focus on the non-manufacturing sector, which is the most regulated and sheltered part of the economy. Indeed, few explicit barriers to competition remain in markets for manufactured goods in the OECD countries.

All indicators take continuous values on a scale going from least to most restrictive of private governance and competition and report the situation at the end of the past decade, the latest period for which complete cross-country information was available at the time of this writing.[18] Box 4.1 provides a cursory view of the main indicators used in

Box 4.1
Indicators of Product Market Regulation

Barriers to Trade

These include tariff and nontariff barriers over the 1988–2001 period. The indicator of nontariff barriers is a frequency ratio: it corresponds to the proportion of (6-digit) tariff lines to which nontariff barriers apply. Nontariff barriers have been aggregated into indicators for two-digit industries using import weights corresponding to 1998 trade flows across OECD countries. A similar indicator for tariff barriers has been constructed. Details are provided in Nicoletti and Scarpetta 2003.

FDI Restrictions

Several types of restrictions are considered: limitations on foreign ownership, screening or notification procedures, and operational restrictions. These restrictions are reported for nine sectors, of which seven are services, and then aggregated into a single measure for the economy as a whole. The indicator covers the 1980–2000 period. Details can be found in Golub 2003.

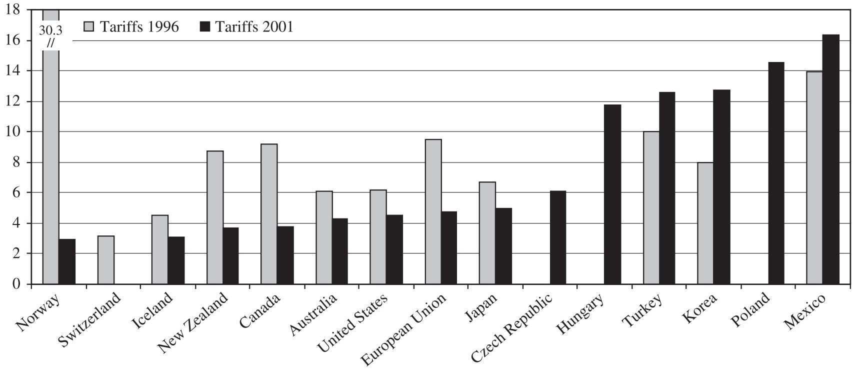
Non-manufacturing Regulation

The indicator covers regulations and market conditions in seven energy and service industries over the 1975–1998 period: gas, electricity, post, telecoms (mobile and fixed services), passenger air transport, railways (passenger and freight services), and road freight. The coverage of regulatory areas varies across industries. Barriers to entry are reported for all industries; public ownership is reported in all industries except road freight; vertical integration is documented for gas, electricity, and railways; market structure is documented for gas, telecommunications, and railways; and price controls are reported for road freight. The summary indicator is the simple average of the industry-level indicators. See Nicoletti and Scarpetta 2003 and Alesina et al. 2005 for details.

Economy-wide Indicators of Product Market Regulations

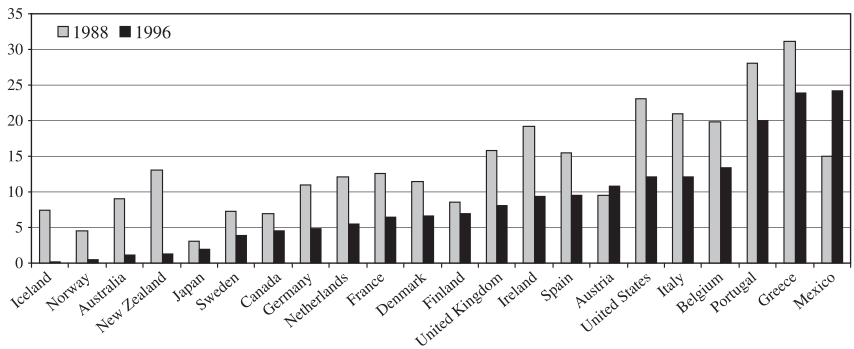
These indicators describe the 1998 policy environment in OECD product markets summarizing information on 139 general-purpose and industry-specific regulatory provisions restricting domestic market mechanisms (in potentially competitive environments) and international trade and investment. Details on data sources, scoring, and aggregation methods are provided in Nicoletti, Scarpetta, and Boylaud 1999.

Tariffs in 1996 and 2001[1]

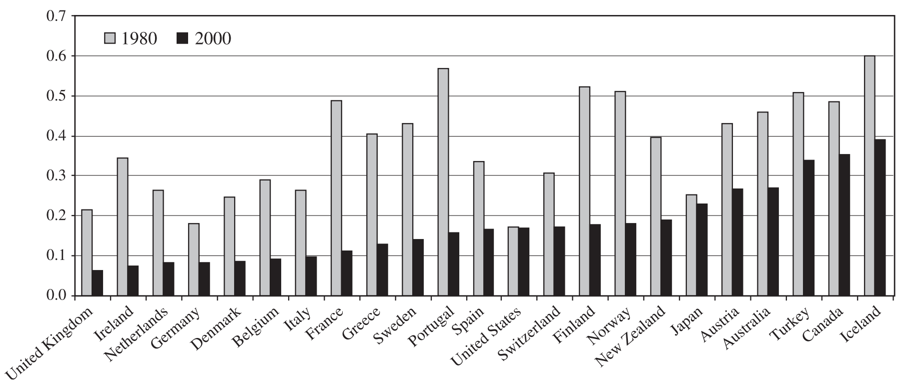


Import coverage of nontariff barriers[2]

Per cent



FDI restrictions over time, 1980–2000[3]



the empirical analyses surveyed here. The first three consists of indicators measuring the strength of barriers to trade, FDI restrictions, and the degree of regulation in non-manufacturing industries. These indicators have a full time-series dimension. We also use an economy-wide indicator covering product market regulations in both domestic markets and traded goods, which provides a summary view of the extent to which both economic and administrative regulations affected competition and private governance in each country. This indicator is only reported for 1998. It is important to note that these indicators are not intended to measure how effective regulations are in meeting their stated public policy goals; they merely quantify the "market unfriendliness" of regulations.

The variability in product market regulation both across countries and over time is substantial. Figure 4.4 focuses on border policies, showing the evolution of tariff and nontariff barriers and explicit restrictions to foreign direct investment. Trade restrictions have generally fallen in the OECD area. Over the 1996–2001 period, average tariffs have further declined in most countries, though the dispersion of bilateral tariff rates remains wide, reflecting tariff discrimination across trading partners. Available information on nontariff barriers up to 1996 suggests that these barriers have declined as well. However, they were still significant in many OECD countries, and more recent surveys suggest that such barriers have shifted from border restraints to domestic policies restricting access to markets. Foreign direct investment restrictions were also softened significantly over the past two decades, but cross-country differences remain large, with most European Union (EU) countries showing greater openness than the United States and Japan (largely due to complete liberalization of capital flows within the EU Single Market).[19]

Figure 4.5 highlights, by means of the summary indicator for non-manufacturing regulation, the general patterns of reform in non-manufacturing over the 1980–1998 period. The box plot shows the

Figure 4.4
Openness to trade and foreign direct investment.
Source: WTO Online Database; UNCTAD TRAINS Database; Golub 2003.
1. Simple average of applied MFN tariff rates.
2. OECD calculation based on UNCTAD data. Aggregation from two-digit level tariffs to national level using sectoral value-added weights.
3. The indicator ranges from 0 (least restrictive) to 1 (most restrictive). The most recent year for which data are available varies across countries between 1998 and 2000.

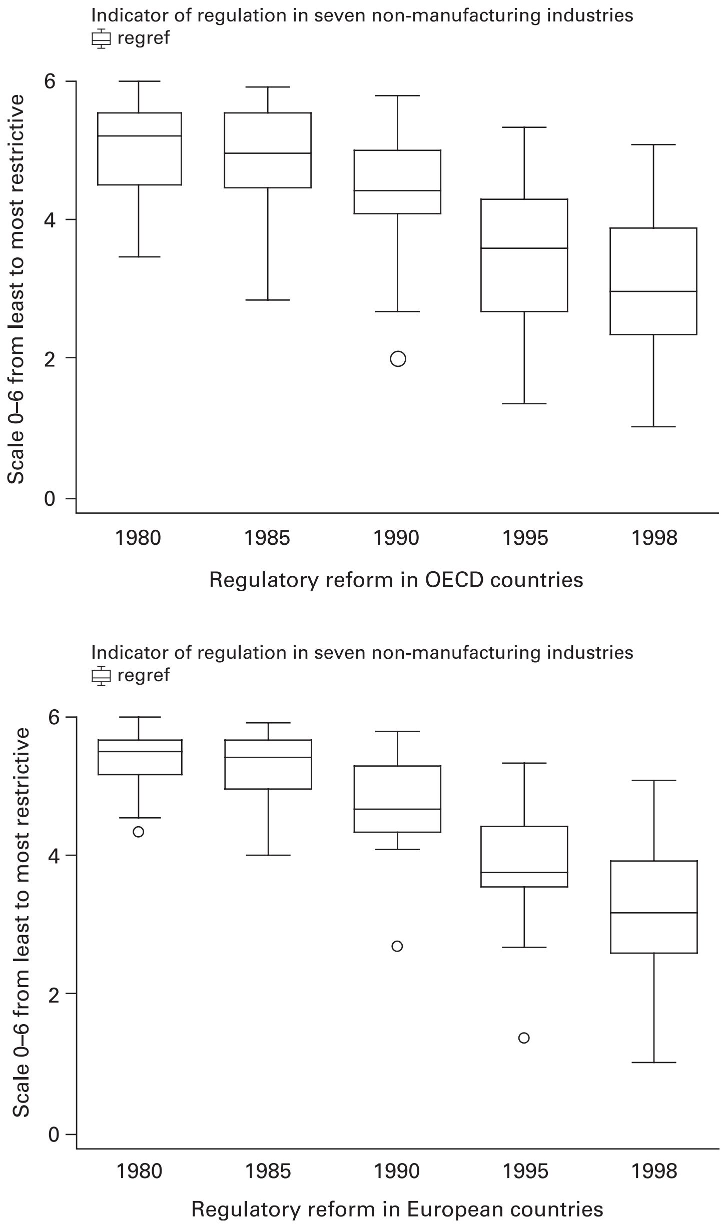


Figure 4.5
Regulatory reform in non-manufacturing.[1,2]
Source: Nicoletti and Scarpetta 2003.

median level of regulation and the dispersion of regulatory approaches across countries in each year. There was some policy convergence over the past two decades in absolute terms, with policies generally becoming friendlier to market mechanisms. However, mainly due to differences in initial conditions and in the pace of reform, regulatory policies diverged in relative terms, with a widening variance of approaches across countries in the most recent period. Paradoxically, the divergence in policies over the 1995–1998 period was widest within the European Union, despite efforts by the European Commission (EC) to harmonize the business environment in the EU Single Market.

Cross-country differences in regulation, and hence differences in the implementation of regulatory reform, can also be gauged looking at the economy-wide indicator of regulations (figure 4.6). In 1998, the United Kingdom had both the lighter administrative burdens (e.g., start-up costs) and the least restrictive market regulations (e.g., barriers to entry, public ownership). It was followed closely by the other common law countries, although economy-wide and/or industry-specific administrative regulations appeared to be somewhat heavier in some relatively liberal countries such as Ireland and New Zealand. At the opposite end, Italy was the most restrictive among the countries surveyed, with heavy administrative regulations also found in France, Belgium, and, to a lesser extent, Japan and Germany. Interestingly, countries with tight economic regulations also had burdensome administrative procedures on business enterprises.

4.3.2 Empirical Issues

The next section presents empirical results relating cross-country differences in the scope and pace of product market reform to investment and productivity outcomes. We look at domestic capital formation and multifactor productivity developments at the industry level and at

Figure 4.5 (continued)

1. The box plot shows, in each year, the median OECD value of the regulatory indicator (the horizontal line in the box), the third and second quartiles of the cross-country distribution (the edges of each box) and the extreme values (the two whiskers extending from the box). Dots identify outlier observations. For consistency, European countries include only EU members since 1980.

2. The indicator summarizes regulation (including barriers to entry, vertical separation, price control, and public ownership) in air, rail, and road transport; post and telecommunications; and electricity and gas.

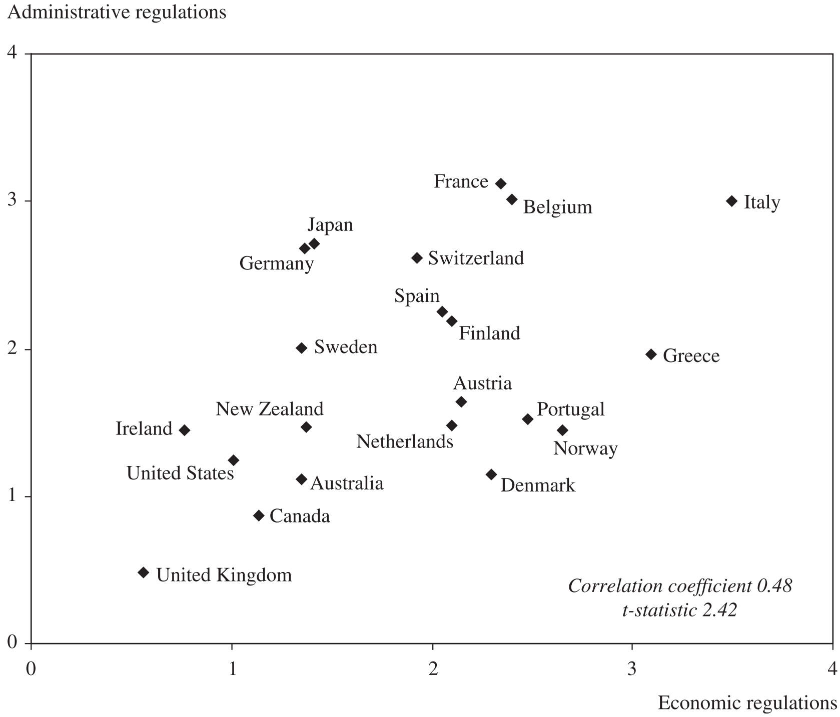


Figure 4.6
Overall regulatory approaches across countries.[1] Economic and administrative regulation.[2]
1. The scale of indicators is 0–6 from least to most restrictive.
2. Administrative regulation includes reporting, information and application procedures and burdens on business start-ups, implied by both economy-wide and sector-specific requirements; economic regulation includes all other domestic regulatory provisions affecting private governance and product-market competition (such as state control and legal barriers to entry in competitive markets).

bilateral and overall stocks of foreign direct investment. The phenomena addressed are different, and details on the empirical models estimated in each case are provided later. Nevertheless, the empirical approaches share some common features that we highlight here.

First, as mentioned in section 4.1, we try to move away from the estimation of aggregate growth relationships to look at some of the channels through which policies can affect GDP per capita performance. The models estimated are not "structural" in the sense of involving restrictions that allow direct testing of alternative economic theories of investment and productivity. However, their specifications are linear approximations based on widely accepted theories of the phenomena at hand. Thus, the capital formation specification explicitly derives from a standard model of forward-looking investment behavior with adjustment costs, the FDI specification is consistent with Markusen's "unified approach" to international investment, and the MFP specification is grounded in conditional convergence models that attribute a large role to international technological transfer. All these models attribute an important potential role to policies, in particular, those affecting private governance and competition.

Accordingly, we focus on the effects of liberalization and privatization policies, which have had a wide variability across countries, sectors, and time. However, we control either explicitly or implicitly for institutions and policies in other areas (e.g., the labor market), which often vary less in one or more of these dimensions. For instance, labor market policies are explicitly introduced in the models for domestic and foreign investment, while they are accounted for implicitly through fixed country and time dummies in the MFP model. Other potentially important policy areas, such as financial market regulation, are not covered due to lack of data. However, all the estimated models include a full set of fixed effects (for countries, sectors, and time) that attempt to capture unobserved influences on investment and productivity.

Most of the analysis is disaggregated at either the industry level (for domestic investment and MFP) or at the bilateral level (for FDI). This required a significant effort in data construction and verification (e.g., for obtaining sectoral MFP estimates or indicators of regulation). In this context, we took particular care to relate as much as possible the dependent variable to either industry-level or country-pair influences. For instance, sector-specific or country-pair-specific series for human capital and regulatory policies were used as regressors in the estimations. It should be noted, however, that while this disaggregation has the advantage of bringing the analysis closer to the source of the phenomenon being studied, it may also be more prone to problems of heterogeneity and measurement error. For instance, measured productivity and investment in service sectors are generally less reliable than in manufacturing or, for that matter, the economy as a whole. For other variables, however, such as regulations, measurement is probably more accurate at the industry level, and this may create a mismatch between the accuracy of measurement in the dependent and explanatory variables.

We tried to address the potential heterogeneity problems implicit in the panel estimation of average parameters in two ways. First, we restricted the analysis to a relatively homogenous subset of "core" OECD countries. Typically, these would include the EU members (before the recent enlargement), Switzerland, North American countries, Japan, Australia, and New Zealand. In this subset, country-specific responses to liberalization and privatization policies can be expected to be quite similar, although heterogeneity biases in coefficient estimates cannot be excluded. Second, we acknowledged potentially different reactions to such policies in different sectors by either restricting the analysis to a relatively homogeneous set of industries (such as network industries in the domestic investment estimates) or by testing differences in crucial coefficients across sectors (such as between the speed of catch-up and the reaction to regulation in manufacturing and services in the MFP estimates).

Finally, to verify the robustness of our estimates to potential measurement error and small sample problems, we systematically screened regressions for the presence of outliers. This was done in two ways. In all estimations, we identified outliers using a combination of methods based on leverage values and studentized residuals.[20] We then estimated equations both including and excluding outliers. In some estimations, notably those for MFP, we checked the robustness of the value and significance of key coefficients to recursively dropping country-sector observations one at a time.

4.4 Empirical Evidence on the Effect of Product Market Regulations on Economic Performance

There is sufficient variability in approaches to market regulation across countries and over time to expect that some of the channels highlighted in previous sections could help explain differences in growth performances in the OECD area. To check this, we take a bottom-up approach and look at the evidence concerning the effects of regulation on the main determinants of growth: capital formation and multifactor productivity. Since the purpose of this chapter is to summarize this evidence, we report only the main features and results of the analyses, as well as (where possible) some quantitative inferences on the likely impact of regulation on macroeconomic aggregates. In interpreting these results, the reader should bear in mind the illustrative nature of any policy simulation based on aggregate regressions. For details on the empirical approaches characterizing each study, the reader may refer to the appendix or the original papers.

4.4.1 Regulation and Investment

There are very few studies looking at the effects of product market policies on aggregate investment. Yet casual observation suggests that various measures of investment have recently been related to product market developments. For instance, countries with a more restrictive regulatory environment tend to invest less in ICT (figure 4.7). Moreover, investment in crucial non-manufacturing industries appears to have been increasing in the "business-friendly" United States and United Kingdom at a faster rate than in "restrictive" large continental EU countries. The patterns of investment vary significantly between the two groups of countries: while in the United States and the United Kingdom investment as a share of the capital stock increased from 3.7 percent in 1975 to 8.2 percent in 1998, in the large continental European

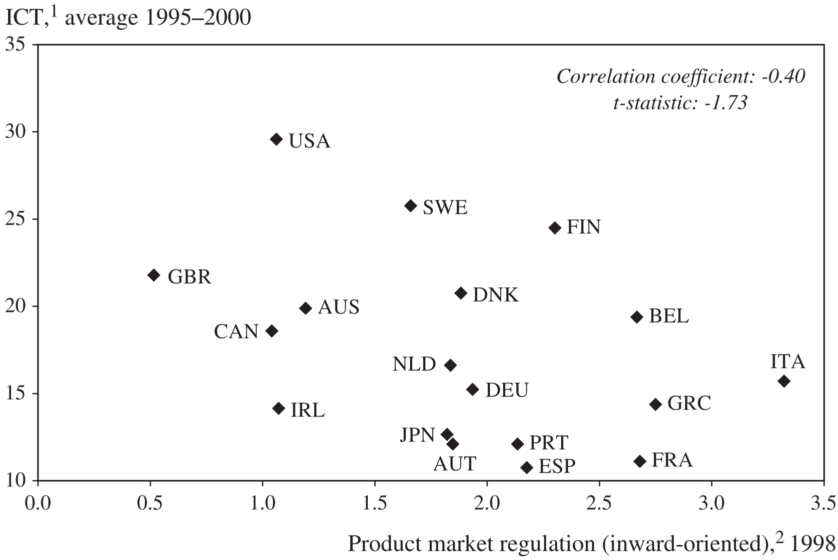


Figure 4.7
Regulation and investment in information and communication technology.
Source: OECD Productivity Database and Nicoletti, Scarpetta, and Boylaud 1999.
1. ICT investment as a percentage of non-residential gross fixed capital formation.
2. Indicator of economy-wide regulation excluding barriers to international trade and investment. The indicator ranges from 0 to 6, from least to most restrictive.

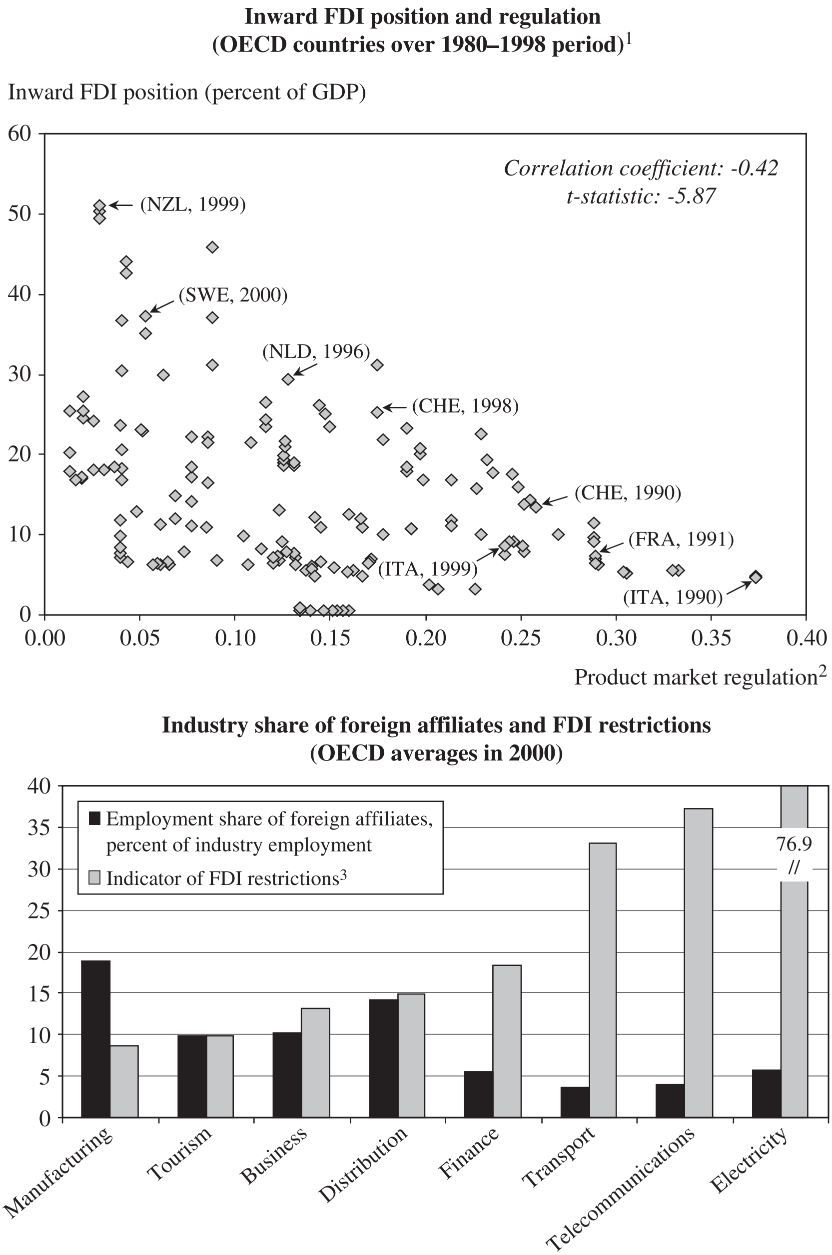


Figure 4.8
Regulation and foreign direct investment.
Source: Nicoletti et al. 2003.

countries the investment rate decreased by 5 percentage points from 9.4 percent to 4.4 percent. Interestingly, the United States and the United Kingdom strongly liberalized product markets before the nineties, while reforms were almost absent before then in Italy, Germany, and France. Moreover, changes in investment trends in each non-manufacturing industry are statistically associated with the timing of reforms.

Here, we summarize two recent studies that look at investment in the non-manufacturing sector (Alesina et al. 2003) and aggregate foreign direct investment (Nicoletti et al. 2003), respectively. The focus on non-manufacturing industries is justified by the fact that they are major users of ICT and there are large differences across countries in the extent to which these industries contribute to aggregate productivity growth. Moreover, these differences appear to be closely related to the extent of regulation, with most restrictive countries failing to obtain productivity gains. Looking at FDI is relevant because there is a strong presumption that FDI could significantly contribute to productivity growth. Casual observation suggests that countries and industries where regulations restrict access and make business difficult tend to receive less inward FDI (figure 4.8).

4.4.1.2 Investment in Non-manufacturing

Alesina et al. (2005) look at the effects of regulation on investment in the transport (airlines, road freight, and railways), communication (telecommunications and postal services), and utilities (electricity and gas) sectors. They estimate a simple dynamic panel model of investment and regulation, controlling for sector/country fixed effects and common or sector-specific year effects (see the appendix for more details). They measure regulation with the time-varying indicator described earlier, which captures entry barriers and the extent of public ownership, among other things. They find that regulatory reforms have had a significant positive impact on capital accumulation in

Figure 4.8 (continued)

1. Each point shows the combination of regulation and FDI in a given country and period. Some of these country/period contributions are shown for illustrative purposes.
2. Product of the indicator of economy-wide regulation in 1998 and the indicator of barriers to entry in seven non-manufacturing industries over the 1980–1998 period. 0–1 scale from least to most restrictive of competition.
3. The indicator ranges from 0 (least restrictive) to 100 (most restrictive).

Summary indicator of restrictions on inward foreign direct investment in nine sectors.

these industries. In particular, liberalization of entry in potentially competitive markets has a sizeable effect on long-run investment rates.

Their findings are consistent with the idea that a reduction in barriers to entry is likely to stimulate investment because it leads to a decrease of the markup and, therefore, of the shadow cost associated with capital and output expansion. This effect appears to outbalance potential downward pressures on investment that might have originated from changes in incentives due to the redesign of price regulation (e.g., from rate of return to price caps or inappropriate access pricing regimes) or changes in industry structure (e.g., vertical separation of networks from service provision). While in principle agency problems and political mandates affecting the behavior of public managers may lead to over-accumulation of capital, Alesina et al. (2005) also find that privatization spurs investment. This suggests that the reduction of barriers to entry for private firms associated with the elimination of state control on business enterprises more than compensates for the reduced importance of potential overinvestment problems due to managerial incentives. Interestingly, evidence also exists that the marginal effect of deregulation on investment is greater when the policy reform is large and when changes occur starting from relatively low levels of regulation. In other words, small changes in a heavily regulated environment are not likely to produce much of an effect. The implications of the analysis are clear: regulatory reforms that substantially lower entry barriers are likely to spur investment. The authors show that the empirical results are robust to several sensitivity checks and extensions.

Using these results, it is possible to derive some (highly tentative) quantitative estimates of the potential effects of product market reforms on investment. For instance, a decrease of the indicator of regulation from its third-quartile value to its first-quartile value would generate, according to the estimated model, an increase in the investment rate of approximately two percentage points in the long run, which is quite large. Considering that the sectors analyzed are highly capital intensive, the increase of investment as a percentage of gross output or value added would be even larger. Another way to gather a sense of the magnitude of the changes is to make some experiments with actual values of the indices in different time periods in one country or across countries. For instance, the estimated model would predict that the regulatory reforms implemented in the United Kingdom in the transport and communications sector over the period 1984–1998 would raise the investment rate in the long run by 2.5 percentage points over the same period. The actual increase was 3.0 percentage points. The estimated model would also predict that if Germany and France were to align regulation in non-manufacturing industries with U.S. standards their investment rate would increase by 2.3 percentage points in the long run. Taking the 1994–1998 period average as a benchmark, this would raise Germany's rate from 5.6 percent to 7.9 percent and France's rate from 5.9 percent to 8.2 percent, both much closer to the U.S. average level of 9.0 percent. The same thought experiment would raise Italy's investment rate by 3.3 percentage points, from 6.8 percent to 10.1 percent.

4.4.1.3 Foreign Direct Investment

There is a large literature looking at the structural determinants of FDI flows from the point of view of both investor and host countries (see, e.g., Markusen and Maskus 2001a). Surprisingly, however, less attention has been devoted to the role of policies. While the effects of corporate taxation have been extensively studied (see the survey in de Mooij and Ederveen 2003), very few empirical studies have looked at the influence of other border and, especially, nonborder policies. Interest in this issue has increased recently as a "unified approach" to the analysis of trade and the multinational enterprise has gained ground (Markusen and Maskus 2001b).[21] For instance, Markusen (2002) used this approach to assess the effects of both trade and investment liberalization on bilateral FDI flows.

Nicoletti et al. 2003 is, to our knowledge, the only study that looks explicitly at the effects of a broad range of product market policies on FDI, including both investment barriers and domestic regulations.[22] As argued in section 4.2, this is one of the areas in which the impact of regulation is a priori ambiguous.

Using some of the policy indicators described in section 4.2, Nicoletti et al. (2003) estimate two distinct FDI models: an empirical specification of the determinants of (aggregate) bilateral outward FDI stocks based on Markusen's "unified approach"; and a reduced-form specification of the determinants of total inward FDI stocks (see the appendix for more details). In both specifications, FDI stocks are determined by a set of structural factors—including gravitational ones (such as distance, transaction costs, total market size, etc.) and others reflecting comparative advantage and scale effects (relative factor endowments, relative market size)—and policies in the investor and host countries (participation in free trade areas, tariff and nontariff barriers, FDI restrictions, labor market arrangements, infrastructure investment, and product market regulations). These specifications are estimated on a panel of OECD countries over the past two decades, controlling for a range of unexplained effects (including host-specific, investor-specific, country-pair-specific, and time effects).

FDI restrictions by the host country are estimated to have a significant negative impact on its bilateral FDI stock with specific partners and, consistently, are also found to significantly depress its total inward FDI position (relative to all its partner countries). Similarly, product market regulations that curb competition at home are estimated to have a negative and significant effect on FDI, but what is relevant for bilateral FDI positions is the relative stringency of regulations in the host and investor countries. Put simply, the net effect of regulations that curb competition is to make the host country less attractive for international investors located in countries where regulations are less restrictive. This is confirmed by the significantly negative impact of anti-competitive regulations in the host country (relative to the OECD average) on its total FDI inward position.

With the usual caveat, empirical estimates can be used to quantify the long-run impact on inward FDI positions of changes in policies that affect FDI restrictions and product market regulation.[23] The estimates suggest that bringing down FDI restrictions in all OECD countries to the level of restrictions in the United Kingdom, the least restrictive country according to the indicator used in the empirical analysis, would have a sizeable impact on global integration. The effects of such reforms on FDI inward positions depend on how restrictive each country was before the policy move. Relatively restrictive countries could increase their total FDI inward position (which is typically low in terms of GDP) by between 40 and 80 percent, but even in countries that are estimated to be already relatively liberal the gains could amount to around 20 percent of their initial inward position. Overall, such policy reforms could increase OECD-wide inward positions by almost 20 percent, significantly raising economic integration in the area.

Reducing anti-competitive product market regulations is also likely to significantly increase FDI inward positions. If all OECD countries were to reduce the level of their product market regulations to that of the United Kingdom (again the least restrictive country according to the indicator used in the analysis), OECD-wide inward positions would increase by over 10 percent relative to the initial inward position. Since bilateral FDI outward positions are estimated to depend on the relative stringency of regulation in the home and host countries, relatively restrictive host countries—such as Greece, Italy, and France—that host FDI from relatively liberal countries could increase their FDI inward positions by as much as 60 to 80 percent through regulatory reform. Conversely, countries that are relatively liberal would see the relative attractiveness of their product markets either unchanged (such as in the United States, New Zealand, and Sweden) or even reduced (such as in the United Kingdom and Australia).

4.4.2 Regulation and Multifactor Productivity

Improvements in MFP play a crucial role in the process of economic growth, and in OECD countries they accounted for between one third and one half of the average business sector GDP growth observed over the past two decades (OECD 2003). Cross-country evidence suggests that countries that have extensively reformed their product markets (as measured by our indicators of regulation) have also experienced an acceleration of MFP over the 1990s, while the productivity slowdown (or stagnation) has continued in other countries (figure 4.9).

In Nicoletti and Scarpetta (2003), we move from this aggregate bivariate evidence to a multivariate regression analysis of the driving forces of industry-level MFP growth. We estimate a multifactor productivity equation derived from a production function in which technological progress is a function of country/industry-specific factors, as well as a catch-up term that measures the distance from the technological frontier in each industry (see the appendix).[24] This framework allows testing for the direct effect of institutions and regulations on estimated productivity, as well as for the indirect influences of these factors via the process of technology transfer.[25]

We find that various measures of anti-competitive product market regulation (both economy-wide and industry-specific) significantly curb productivity performance at the industry level.[26] In particular, the long-run costs of anti-competitive regulation, in terms of foregone productivity improvements, are higher in countries that are further away from the technological frontier. This negative effect on productivity catch-up may result, for instance, from lower incentives for organizational and technological change in markets where competition is

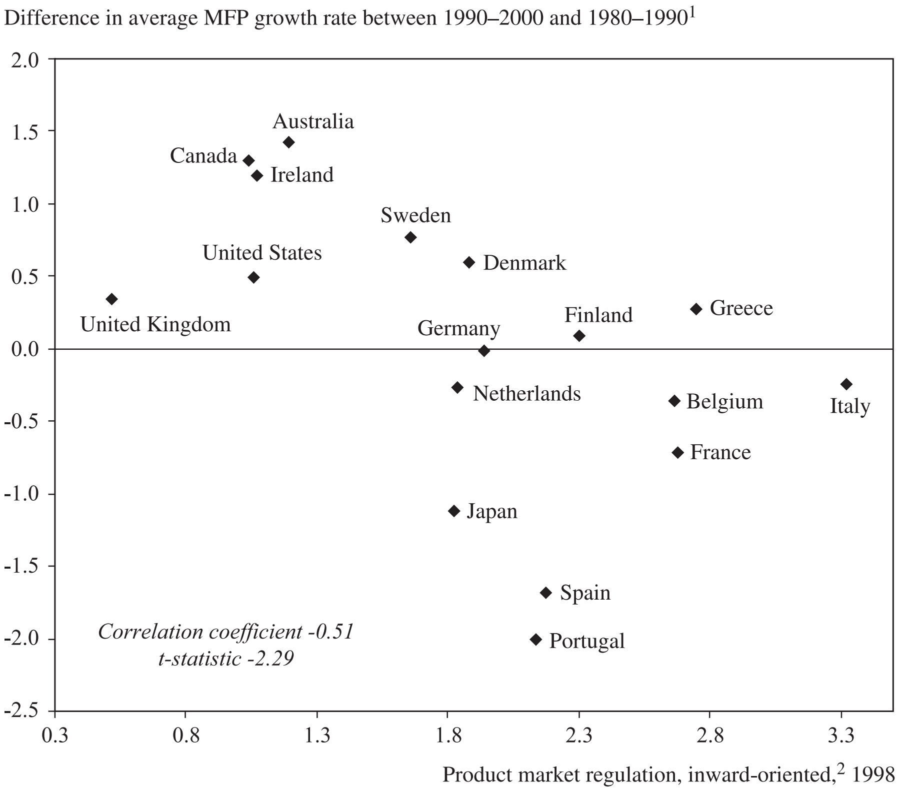


Figure 4.9
Multifactor productivity acceleration and product market regulation.
Source: OECD Productivity Database.
1. Adjusted for hours worked.
2. Indicator of economy-wide regulation excluding barriers to international trade and investment. The indicator ranges from 0 to 6, from least to most restrictive. See Nicoletti, Scarpetta, and Boylaud 1999.

weak due to state interference (e.g., entry barriers or price controls), and, in addition, a narrower scope for knowledge spillovers in markets where entry is restricted. The empirical results also suggest that, by increasing competitive pressures, regulatory reform will promote productivity in each individual industry, regardless of its position with respect to the technology frontier. Our findings are consistent with a growing empirical literature that has looked at the links between competition and productivity at the aggregate and especially at the industry and micro levels.[27]

Table 4.1
Change in the annual percentage growth of multifactor productivity in EU countries implied by the alignment of policies to OECD best practices over ten years

	Contributions of		
	Overall regulatory reform	Industry-specific reforms	Total effect
Austria	0.10	0.32	0.42
Belgium	0.15	0.45	0.60
Denmark	0.10	0.27	0.37
Finland	0.04	0.55	0.59
France	0.19	0.43	0.62
Germany	0.08	0.62	0.70
Greece	0.29	0.83	1.12
Italy	0.22	0.48	0.70
Netherlands	0.11	0.34	0.44
Norway	0.02	0.36	0.38
Portugal	0.23	0.42	0.65
Spain	0.12	0.28	0.41
Sweden	0.01	0.50	0.51
United Kingdom	0.00	0.11	0.11

Source: Nicoletti and Scarpetta 2003.
Note: The simulations are based on the results of a panel regression on twenty-three industries in eighteen OECD countries over the 1984–1998 period.

Our empirical results can be used to illustrate the potential productivity gains that would be induced by regulatory reforms. Bearing in mind the limits of such simulations, a product market reform that would align industry-specific regulations with those of the most liberal OECD country is estimated to reduce the MFP gap vis-à-vis the leading country by around 10 percent, in the long run, in high-gap countries such as Greece, and by around four to six percentage points in several other continental European countries and Japan. Put differently, aligning the overall regulatory stance with that of the most liberal OECD country could increase the annual rate of MFP growth in continental EU countries by between 0.4 and 1.1 percent over a period of ten years (table 4.1).

The effect of entry regulations is likely to be particularly important for productivity performance in industries in which technology is rapidly evolving, such as ICT-producing and ICT-using industries. In these industries, new entrants play an important role in

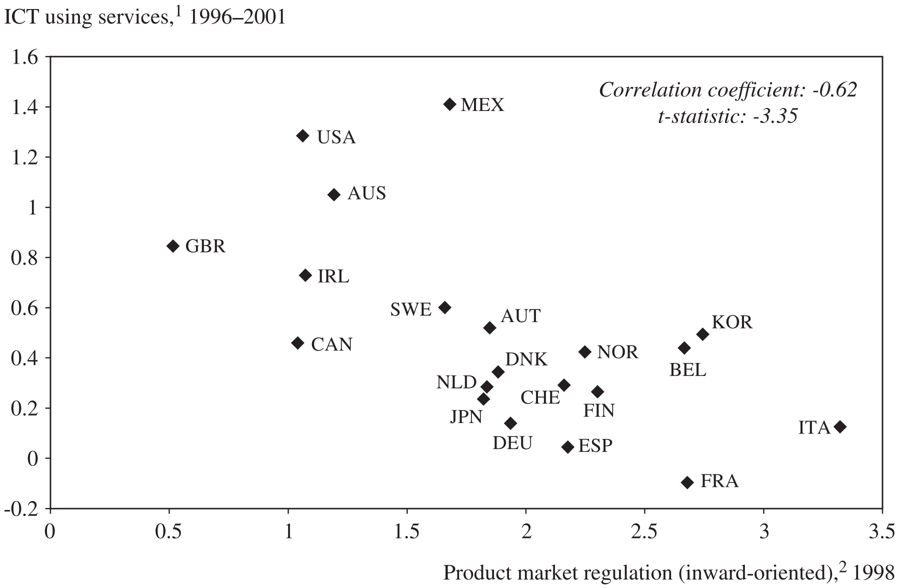


Figure 4.10
Regulation and the contribution of ICT-using services to aggregate productivity growth.
Source: OECD Productivity Database and Nicoletti, Scarpetta, and Boylaud 1999.
1. Contribution to aggregate labor productivity growth.
2. Indicator of economy-wide regulation excluding barriers to international trade and investment. The indicator ranges from 0 to 6, from least to most restrictive.

introducing new vintages of technology (see Scarpetta et al. 2002). Therefore, product market regulations that minimize the prospective costs faced by new entrants are likely to create favorable conditions for increasing the contribution of ICT to productivity growth. As shown by Gust and Marquez (2003) regulatory burdens are likely to have slowed down the adoption of ICT in restrictive countries. While the relationship between regulation and ICT adoption awaits further empirical research, figure 4.10 suggests that in countries that underwent extensive product market reforms, it was easier to translate such investment into productivity improvements in crucial ICT-using sectors, thereby increasing their contribution to aggregate productivity growth.

4.5 Conclusions

In this chapter, we looked at the possible links between product market regulation and growth in the OECD area over the past two decades. The chapter documents how differences in labor utilization, investment in new technologies, and MFP growth underpin the observed cross-country divergences in growth. These factors contributed to accelerate growth in the United States, the United Kingdom, Canada, and a few smaller OECD economies, while they held back growth in large continental EU countries and Japan. We cast the cross-country dispersion in growth patterns against product market reforms made over the past two decades showing that, despite efforts in virtually all OECD countries to make regulations more market-friendly, the cross-country dispersion in product market approaches has also increased due to differences in the pace and depth of reforms. In particular, European countries, despite market integration, EC competition policies, and the European Monetary Union, have been characterized by diverging regulatory reform patterns although new data covering the most recent years suggest that this divergent process of reform has been reversed in Europe.

Can these diverging patterns of reform contribute to explain the puzzling disparities in growth outcomes? The results described in this chapter suggest that there are several links between product market policies and growth performance. In particular, lower barriers to trade and competition in less regulated countries seem to have increased the level and rate of growth of productivity by stimulating business investment and promoting innovation and technological catch-up. These policies can explain part of the growth advantage experienced by the United States and other English-speaking and small EU countries over the past two decades. Thus, regulatory reforms in product markets seem to be an essential element of any "agenda for growth." In the future, this would appear to apply especially to large continental European countries and Japan, which experienced a persistent productivity slowdown and widening GDP per capita gaps vis-à-vis the United States over the past decade.

How reliable and general are these inferences? While overall the regulation-growth link seems robust to different model specifications and sample coverage within the OECD area, several caveats should be pointed out. First, more analysis needs to be done to extend these results to non-OECD countries, where the impact of competition-oriented policies on some of the determinants of growth could be different. Second, several missing links remain. For instance, we have produced only indirect evidence of the effects of product market reforms on innovation. More specific analysis of the link between regulation (including intellectual property rights) and aggregate innovative activity is needed. Moreover, there is still very little evidence on the relationship between economic growth and financial market policies, an important aspect of product market reform that has been left out of our discussion. Lastly, it should be recognized that the channels going from product market policies to performance identified in this chapter cannot be considered in isolation, because investment and productivity outcomes are closely related. Integrating these channels would perhaps provide a more nuanced view of the effects of reform on overall economic growth.

Appendix: Empirical Analyses of Regulation and Performance

4A.1 Regulation and Domestic Investment (Alesina et al. 2005)

The empirical investigation of the links between regulations and domestic investment was based on estimation of various versions of an unrestricted dynamic model of investment of the form

$$(I/K)_{ijt} = \sum_{s=1}^{2} \alpha_s (I/K)_{ijt-s} + \sum_{s=0}^{2} \beta_s REG_{ijt-s} + \gamma_{ij} + \zeta_t + (or\ \zeta_{jt}) + \varepsilon_{ijt},$$

where t represents years, i denotes countries and j sectors; I denotes investment, K the capital stock; REG is the product market regulation index; and the remaining terms capture country/sector-specific fixed effects, and common (or sector-specific) year dummies. The analysis controlled for endogeneity of regulation (by instrumenting it with lags and with some of its structural determinants according to recent political economy literature) and technology changes (through the sector-specific year dummies). The empirical analysis covered three broad non-manufacturing sectors in twelve OECD countries over the 1975–1998 period. The model was estimated with both dynamic fixed effects and generalized method of moments (Arellano and Bond 1991) techniques. Results are robust to changes in estimation approaches and to a number of other extensions, such as allowing for short-run heterogeneity in responses of investment to regulation, controlling for additional country-specific variables (notably, labor market regulation) and country/industry-specific variables (e.g., factor prices).

4A.2 *Product Market Policies and FDI* (**Nicoletti et al. 2003**)

The bilateral estimations covered twenty-eight OECD countries and partners over the 1980–2000 period and were based on equations relating FDI outward stocks to non-policy-related factors, and the relative costs of trading and investing implied by policies in the home or partner countries (Markusen and Maskus 2001b; Egger 2001). The building blocks are (1) standard variables expressing gravity forces, factor proportions, or other economic variables likely to affect FDI (e.g., R&D intensity and exchange rates); (2) indicators of openness (multilateral and bilateral tariffs, multilateral indicators of nontariff barriers, dummies for free trade agreements, and FDI restrictions), domestic product market regulations, and labor market flexibility; and (3) indicators of infrastructure supply. Thus, the basic bilateral model was

$$Y_{ijt} = \sum_{x} \beta_x X_{ijt} + \sum_{c} \beta_c C_{it} + \sum_{p} \beta_p P_{jt} + \alpha_i + \alpha_j + \alpha_t + \alpha_{it} + \alpha_{ij} + \alpha_{jt} + u_{ijt},$$

where Y_{ijt} stands for the logarithm of bilateral FDI outward stocks from country i to partner j at time t (with $i = 1, 2, \ldots, I$; $j = 1, 2, \ldots, J$; and $t = 1, 2, \ldots, T$); X_{ijt} are policy and non-policy-related variables that are specific to a given country-partner pair; C_{it} are country-specific variables; and P_{jt} are partner-specific variables. The α-type variables stand for specific effects that control for all combinations of bilateral, country-, or partner-specific and time-varying or time-invariant unobserved factors. Since estimating dummies for all these factors is not viable, due to an excessive loss of degrees of freedom and high potential multicollinearity, we transformed variables according to Erkel-Rousse and Mirza 2002, decomposing the estimation in two equations in which all variables are expressed as deviations from the mean investor or, alternatively, the mean host. This reduced the number of unobserved components to be estimated parametrically while at the same time preserving the desirable properties of the relevant coefficient estimates. All estimates controlled for outliers and heteroskedasticity.

The model for total inward FDI positions accounted for the possibility that the adjustment of actual to desired stocks of FDI is costly and takes time. Therefore, equations for total FDI inward position were of the dynamic partial adjustment kind, with the total FDI inward position in each period also depending on the realized inward position in the previous period (see Cheng and Kwan 2000). The estimated dynamic panel specification for total FDI inward positions was

$$\ln Y_{it} = \gamma \ln Y_{it-1} + \sum_{x} \delta_x X_{it} + \sum_{z} \lambda_z Z_{it} + v_i + \varepsilon_{it},$$

where the X_{it} are non-policy-related variables, the Z_{it} are policy variables, δ_x and λ_z are parameters to be estimated, v_i are unobserved country-specific time-invariant effects, and ε_{it} is a random disturbance. Estimation was carried out using the generalized method of moments (Arellano and Bond 1991).

4A.3 Product Market Regulations and Multifactor Productivity (Nicoletti and Scarpetta 2003)

The empirical analysis of MFP growth is centered on a catch-up specification of productivity, whereby, within each industry, the production possibility set is influenced by technological and organizational transfers from the technology frontier country (indexed L) to other countries. We further extend the conventional model by assuming that, in each period t, MFP growth in industry k and country i depends on country and industry characteristics (human capital, hc_{ikt}, regulation, pmr_{ikt}, and other unobserved effects) as well as the state of knowledge in the technology leader country (country with the highest level of MFP). In particular, an MFP advance in the frontier country is assumed to produce faster MFP growth in follower countries with the size of this impact increasing with each country's distance from the technological leader (see Scarpetta and Tressel 2002 for more details). Thus the MFP equation is

$$\Delta \ln MFP_{ikt} = \alpha_k \cdot \Delta \ln MFP_{Lkt} + \beta_k \cdot RMFP_{ikt-1} + \gamma_k \cdot pmr_{ikt} \cdot RMFP_{ikt} + \delta hc_{ikt} + \lambda_k pmr_{ikt} + f_i + g_k + d_t + \eta_{ikt},$$

where η is the usual error term, and the equation includes dummies that control for unexplained country-specific (f), industry-specific (g), and time-specific (d) factors. In the equation, *pmr* is the synthetic indicator of product market regulations, which varies over countries, time, and/or industries, depending on the specification, and *RMFP* is the ratio of MFP to the level found in the leader country. Note that α indicates the standard pace of technological transfer from the leader, β quantifies the importance of the technological transfer that depends on the size of the technology gap, and δ shows how the level of human capital affects the pace of technical progress in each country and industry. Moreover, λ shows the direct impact of regulation on productivity growth and γ gauges whether regulation hinders technology transfers from the technological leader. Most coefficients are sector-specific to account for potential heterogeneity bias. The empirical analysis covers twenty-three industries in manufacturing and business services in eighteen OECD countries over the period 1984–1998. The model was estimated using a standard panel data fixed-effects approach controlling for outliers and heteroskedasticity. In those specifications that use country-wide indicators of regulations instead of industry-specific indicators, we also adjusted standard errors and variance-covariance matrices of the estimators for cluster level effects on country-industry using the procedure suggested by Moulton (1996). Moreover, a detailed sensitivity analysis was performed to test the robustness of results: it showed that the results were robust to recursively dropping country/industry observations and to different measures of MFP and human capital. In particular, alternative measures of MFP were considered that control for quality changes in labor input and for the presence of price markups over marginal costs (see Scarpetta and Tressel 2002).

Notes

We thank Alain Desdoigts, Cecilia García-Peñalosa, participants in the CESIFO Venice Summer Institute 2004 on Institutions and Growth, and three anonymous referees for their useful comments on a previous version of the chapter. The opinions expressed are the authors' own and cannot be attributed to the organizations with which they are affiliated.

1. The EU income gap worsened even discounting the effect of unification in Germany.

2. Within Europe, there were starkly contrasting developments: productivity continued to grow fast in Ireland, some Nordic countries, and Portugal, while its growth remained disappointing in the Netherlands, Spain, and, especially, Switzerland.

3. For studies that have looked at the macroeconomic effects of labor market reforms, see, for example, Nickell 1997; Nickell and Layard 1998; Elmeskov, Martin, and Scarpetta 1998; Blanchard 2000.

4. In an early attempt to relate reforms to growth, Koedijk and Kremers (1996) noted that policy changes in European product markets have sometimes been deeper than reforms in labor markets.

5. While the primary (short- and long-run) effects of anti-competitive product market regulation are to be expected on investment and productivity, regulatory hindrances to

competition may also have consequences for employment, both at the firm level and in the aggregate. See Blanchard and Giavazzi 2003, Ebell and Haefke 2003, and Nicoletti et al. (2001).

6. MFP estimates involve a number of difficult measurement problems. For instance, it is hard to make adjustment for quality and compositional changes in the labor input and, especially, the capital stock. Other potential sources of measurement error are economies of scale and mark-up pricing (see Morrison 1999).

7. The workforce throughout the OECD has gradually become better educated, as entering young cohorts have higher educational attainment than the exiting older cohorts.

8. The odd position of Finland, which invested enormously in ICT goods relative to other countries, can be partly explained by its specialization in ICT-related industries.

9. See Sappington and Sidak 2003.

10. Cost-increasing regulations in the investor country may also stimulate outward FDI by favoring the delocalisation of production plants in countries with less costly regulations. On the other hand, the costs implied by these regulations may cripple the ability of investor-country firms to internationalise production to the desired level. For instance, there is evidence that certain product market regulations can hinder firm growth and curb R&D spending (Nicoletti et al. 2001; Bassanini and Ernst 2002). Both factors can constitute a handicap for internationalization.

11. For instance, the effects on growth trajectories of reforms that improve the efficiency in the use of inputs have been recently stressed by Bergoeing et al. (2002).

12. See Winston 1993 for a review.

13. These channels are highlighted by Lazear and Rosen (1981), Nalebuff and Stiglitz (1983), and Aghion and Howitt (1998).

14. This literature has been recently surveyed by Keller (2004). See also Görg and Greenaway (2002).

15. Cross-country evidence on the effects of regulation on R&D spending is provided by Bassanini and Ernst (2002).

16. The role of intersectoral input-output linkages in transmitting and amplifying the effects of product market reform has been recently stressed by Faini et al. (2004).

17. Another important element of product market reform was liberalisation of international capital flows. However, the empirical analysis surveyed in this paper does not cover the effects of liberalisation in financial markets.

18. The construction of such indicators is an intricate business that is explained in detail in Nicoletti, Scarpetta, and Boylaud 1999; Nicoletti and Scarpetta 2003; Alesina et al. 2005; and Golub 2003. The complete set of indicators is currently being updated to reflect developments over the 1998–2003 period.

19. As noted by Kox, Lejour, and Montizan (2004), this liberalization is partly formal, because barriers related to the heterogeneity of regulations and administrative requirements across countries remain significant, even within free trade areas such as the European Union.

20. In the different specifications, the outliers have been identified using the DFITS statistics and the COVRATIO statistics, which, in turn, are based on the studentized residuals and the leverage values. The outliers are those annual observations for a given country that significantly increase the standard error of the regression or affect the estimated coefficients.

21. This approach merges early analyses, largely replicating gravitational models of trade, with models of the multinational enterprise that stress the joint determination of trade and FDI, economy-wide and firm-level economies of scale and the policy influences upon them.

22. The study also considers the effects of trade barriers on FDI, related to the so-called tariff-jumping rationale for horizontal FDI, as well as the effects of infrastructure policies. Here, we focus only on results concerning FDI restrictions and domestic regulations.

23. It is important to notice that, given the specification of some of the policy variables (which entail a comparison between policies of the investor country and the host country), the quantitative effects highlighted in these thought experiments cannot take into account diversion effects (i.e., FDI redirected from one country to another). To the extent that these are important, the simulation results may overestimate the effects of policy changes on the variables of interest.

24. Griffith, Redding, and van Reenen (2000) have, among others, used a similar approach. However, their study does not include regulatory variables, nor does it consider industry differences in important covariates (e.g., human capital). A number of other studies have looked at productivity convergence using country/industry data.

25. For example, if the adoption of new technologies relies partly on new firms, high entry barriers may reduce the pace of adoption (see, e.g., Boone 2000b).

26. For additional evidence of the productivity effects of entry liberalization at the industry level, see the papers in OECD 2001.

27. For a review of the available studies using industry-level data, see Scarpetta and Tressel 2002. For cross-country studies that explore the role of competition on productivity using markups and concentration indexes, see Cheung and Garcia Pascual 2001. For studies using firms' market shares, see Nickell 1996, Nickell, Nicolitsas, and Dryden 1997, and Disney, Haskel, and Heden 2000.

References

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt. 2002. "Competition and Innovation: An Inverted U Relationship." NBER working paper no. 9269, October.

Aghion, P., C. Harris, P. Howitt, and J. Vickers. 2001. "Competition, Imitation and Growth with Step-by-Step Innovation." *Review of Economic Studies* 68 (July): 467–492.

Aghion, P., and P. Howitt. 1998. *Endogenous Growth Theory*. Cambridge, MA: MIT Press.

Alesina, A., S. Ardagna, G. Nicoletti, and F. Schiantarelli. 2005. "Regulation and Investment." *Journal of the European Economic Association* 3, no. 4: 791–825.

Arellano, M., and S. Bond. 1991. "Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58, no. 2: 277–297.

Averch, H., and L. L. Johnson. 1962. "Behavior of the Firm under Regulatory Constraint." *American Economic Review* 52 (December): 1053–1069.

Barro, R. J., and X. Sala-i-Martin. 1998. *Economic Growth*. Cambridge, MA: MIT Press.

Bassanini, A., and E. Ernst. 2002. "Labour Market Institutions, Product Market Regulation and Innovation: Cross-country Evidence." OECD Economics Department working paper no. 316, Paris.

Bergoeing, R., P. J. Kehoe, T. J. Kehoe, and R. Soto. 2002. "Policy-driven Productivity in Chile and Mexico in the 1980's and 1990's." *American Economic Review* 2, no. 2: 16–21.

Blanchard, O. 2000. "Rents, Product and Labor Market Regulation, and Unemployment." Lionel Robbins Lecture, LSE, October.

Blanchard, O., and F. Giavazzi. 2003. "Macroeconomic Effects of Regulation and Deregulation in Goods and Labor Markets." *Quarterly Journal of Economics* 18, no. 3: 879–907.

Boone, J. 2000a. "Competition." Center discussion paper no. 2000-104, Tilburg University, Netherlands, October.

Boone, J. 2000b. "Competitive Pressure: The Effects on Investments in Product and Process Innovation." *RAND Journal of Economics* 31, no. 3: 549–569.

Cheng, L. K., and Y. K. Kwan. 2000. "What Are the Determinants of the Location of Foreign Direct Investment? The Chinese Experience." *Journal of International Economics* 51, no. 1: 379–400.

Cheung, Y. W., and A. Garcia Pascual. 2001. "Market Structure, Technology Spillovers, and Persistence in Productivity Differentials." CESifo working paper series no. 517.

de Mooj, R. A., and S. Ederveen. 2001. "Taxation and Foreign Direct Investment." CPB (Netherlands Bureau for Economic Policy Analysis) discussion paper no. 3.

Disney, R., J. Haskel, and Y. Heden. 2000. "Restructuring and Productivity Growth in UK Manufacturing." CEPR discussion paper no. 2463, May.

Ebell, M., and C. Haefke. 2003. "Product Market Deregulation and Labour Market Outcomes." IZA discussion paper no. 957.

Elmeskov, J., J. P. Martin, and S. Scarpetta. 1998. "Key Lessons for Labour Market Reforms: Evidence from OECD Countries' Experience." *Swedish Economic Policy Review* 5, no. 2: 205–252.

Erkel-Rousse, H., and D. Mirza. 2002. "Import Price Elasticities: Reconsidering the Evidence." *Canadian Journal of Economics* 35, no. 2: 282–306.

Faini, R., J. Haskell, G. Barba Navaretti, C. Scarpa, and C. Wey. 2004. *Contrasting Europe's Decline: Do Product Market Reforms Help?* Report presented at the 6th European Conference of the Fondazione Rodolfo Debenedetti, Lecce, Italy, June.

Golub, S. 2003. "Measures of Restrictions on Inward Foreign Direct Investment for OECD Countries." OECD Economics Department working paper no. 357, Paris.

Görg, H., and D. Greenaway. 2002. "Foreign Direct Investment and Intra-industry Spillovers: A Review of the Literature." GEP research paper 01/37.

Griffith, R., S. Redding, and J. van Reenen. 2000. "Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries." IFS working paper W00/02.

Gust, C., and J. Marquez. 2003. "International Comparisons of Productivity Growth: The Role of Information Technology and Regulatory Practices." In *Labour Economics*, Special Issue on Productivity, ed. Gilles Saint-Paul.

Holmström, B. 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13: 324–340.

Keller, W. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42 (September): 752–782.

Klette, T. J., and Z. Griliches. 2000. "Empirical Patterns of Firm Growth and R&D Investment: A Quality Ladder Model Interpretation." *Economic Journal* 110, no. 127 (April): 363–387.

Koedijk, K., and J. Kremers. 1996. "Market Opening, Regulation and Growth in Europe." *Economic Policy* 23: 443–460.

Kox, H., A. Lejour, and R. Montizan. 2004. "The Free Movement of Services within the EU." CPB document no. 69, The Hague, October.

Lazear, E. P., and S. Rosen. 1981. "Rank-order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89: 841–864.

Markusen, J. R. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press.

Markusen, J. R., and K. E. Maskus. 2001a. "General-equilibrium Approaches to the Multinational Firm: A Review of Theory and Evidence." NBER working paper no. 8334.

Markusen, J. R., and K. E. Maskus. 2001b. "A Unified Approach to Intra-industry Trade and Direct Foreign Investment." NBER working paper no. 8335.

Meyer, M., and J. Vickers. 1997. "Performance Comparisons and Dynamic Incentives." *Journal of Political Economy* 105, no. 3: 547–581.

Morrison, C. J. 1999. *Cost Structure and the Measurement of Economic Performance*. Norwell, MA: Kluwer Academic Press.

Moulton, B. R. 1996. "Bias in the Consumer Price Index: What Is the Evidence?" *Journal of Economic Perspectives* 10, no. 4: 159–177.

Nalebuff, B., and J. Stiglitz. 1983. "Information, Competition and Markets." *American Economic Review, Papers and Proceedings* 73: 278–293.

Nickell, S. 1996. "Competition and Corporate Performance." *Journal of Political Economy* 104: 724–746.

Nickell, S. 1997. "Unemployment and Labor Market Rigidities: Europe versus North America." *Journal of Economic Perspectives* 11, no. 3 (Summer): 55–74.

Nickell, S. 1999. "Product Markets and Labour Markets." *Labor Economics* 6: 1–20.

Nickell, S., and R. Layard. 1998. "Labour Market Institutions and Economic Performance." Discussion paper no. 407, Centre for Economic Performance, September. Forthcoming in *The Handbook of Labor Economics*, vol. 3.

Nickell, S., D. Nicolitsas, and N. Dryden. 1997. "What Makes Firms Perform Well?" *European Economic Review* 41: 783–796.

Nicoletti, G., A. Bassanini, E. Ernst, S. Jean, P. Santiago, and P. Swaim. 2001a. "Product and Labour Market Interactions in OECD Countries." OECD Economics Department working paper no. 312, Paris.

Nicoletti, G., R. C. G. Haffner, S. Nickell, S. Scarpetta, and G. Zoega. 2001b. "European Integration, Liberalization and Labor Market Performance." In *Welfare and Employment in a United Europe*, ed. G. Bertola, T. Boeri, and G. Nicoletti, 147–236. Cambridge, MA: MIT Press.

Nicoletti, G., S. Golub, D. Hajkova, D. Mirza, and K.-Y. Yoo. 2003. "The Influences of Policies on Trade and Foreign Direct Investment." *OECD Economic Studies* no. 36. Paris: OECD.

Nicoletti, G., and S. Scarpetta. 2003. "Regulation, Productivity and Growth." *Economic Policy* 36: 11–72.

Nicoletti, G., S. Scarpetta, and O. Boylaud. 1999. "Summary Indicators of Product Market Regulation with an Extension to Employment Protection Legislation." OECD Economics Department working paper no. 226, Paris.

OECD. 2001. *OECD Economic Studies* no. 32. Special issue on regulatory reform. Paris: OECD.

OECD. 2003. *The Sources of Growth in OECD Countries*. Paris: OECD.

Pissarides, C. A. 2001. "Company Start-up Costs and Employment." Center for Economic Performance discussion paper no. 520, London School of Economics.

Sappington, D. E., and J. G. Sidak. 2003. "Incentives for Anticompetitive Behavior by Public Enterprises." *Review of Industrial Organization* 22: 183–206.

Scarpetta, S., P. Hemmings, T. Tressel, and J. Woo. 2002. "The Role of Policy and Institutions for Productivity and Firm Dynamics: Evidence from Micro and Industry Data." OECD Economics Department working paper no. 329, Paris.

Scarpetta, S., and T. Tressel. 2002. "Productivity and Convergence in a Panel of OECD industries: Do Regulations and Institutions Matter?" OECD Economics Department working paper no. 342, Paris.

Scharfstein, D. 1988. "Product-market Competition and Managerial Slack." *The RAND Journal of Economics* 19: 147–155.

Schmidt, M. K. 1997. "Managerial Incentives and Product Market Competition." *Review of Economic Studies* 64: 191–213.

Vickers, J. 1995. "Concepts of Competition." *Oxford Economic Papers* 47, no. 1 (January): 1–23.

Winston, C. 1993. "Economic Deregulation: Days of Reckoning for Microeconomists." *Journal of Economic Literature* 31, no. 3: 1263–1289.

II

The Impact of Institutions on Innovation and Entrepreneurship

5 Institutions and Technological Innovation during Early Economic Growth: Evidence from the Great Inventors of the United States, 1790–1930

B. Zorina Khan and
Kenneth L. Sokoloff

Intellectual property institutions have long played a central role in discussions of economic growth. In recent years, there has been a revival of concern over the impact of patent institutions on the rate and direction of inventive activity, and on technological change more generally. Much of the analysis has focused on the most direct effects of granting an exclusive property right in technological knowledge: the enhanced returns that inventors can extract by enjoying a state-mandated monopoly on their discoveries; and the higher costs that users of new technologies or consumers of final goods have to bear as a result of a society's recognizing property rights in technical information. For some decades, the dominant stream of thought held that the dynamic gains associated with more rapid rates of technological progress, induced by offering greater material incentives to investors in research and development, outweighed the static losses suffered because of slower diffusion. Scholars and policymakers have become increasingly skeptical about this calculus, however, with particular scrutiny of whether it is applicable to developing countries. Their chief concerns are that strengthened intellectual property rights in developing-country contexts might lead to a high volume of licensing or royalty fees from these societies to the highly industrialized countries, and that there might not be much of a return in terms of higher domestic rates of invention or productivity growth.

This chapter draws on economic history to highlight another feature of intellectual property whose significance has received little attention. We argue that defining and enforcing a tradable asset in new technological knowledge is extremely important for fostering a market in technology, and for extending and increasing incentives for investment in inventive activity to segments of the population that would otherwise find it difficult to directly extract returns from their technological creativity. The chapter explores this feature in the context of an assessment of the evolution of intellectual property systems, and its impact on the course of technological progress in one of the early industrializers, the United States. The historical perspective is valuable for appraising current issues, even though the effects and appropriateness of specific institutions such as patent systems vary with the frontiers of technology as well as other circumstances that certainly change dramatically over time. In our view, the repercussions of providing broad access to property rights in new technological knowledge may be much more relevant and favorable for developing economies today than is generally realized.

In order to demonstrate just how revolutionary the U.S. patent system was in providing broad access to a secure and well-defined asset in new technological knowledge, we discuss the development of patent institutions in Britain and France, relative to America, in section 5.1. Section 5.2 presents evidence regarding individuals who made contributions to the technological frontier in the nineteenth and early twentieth centuries. We describe the composition of a sample of "great inventors" and their patterns of patenting from 1790 through the 1930s, and we examine patterns of geographical location, mobility, nativity, and access to privileges such as schooling at institutions of higher learning. Moreover, the data allow us to determine whether specific classes of these so-called great inventors differed in their tendency to rely on the use of their patent as an asset that allowed them to extract returns from their technological creativity. Section 5.3 concludes with a brief discussion of how the U.S. innovation of a modern patent system diffused over the nineteenth century, and the relevance to developing countries today of that country's experience with the role of institutions in fostering democratic invention.

5.1 Early Patent Systems

A fundamental and enduring concern of organized society is the design of institutions that encourage private actors to undertake investments conducive to improvements in social welfare. It is therefore hardly surprising that there is a long history of public policies crafted to stimulate would-be inventors, innovators, and investors to contribute to the advance and diffusion of technological knowledge. Appreciation of the potential importance of such policies grew over the late eighteenth and nineteenth centuries, as it became clear that ongoing technological progress was feasible, capable of altering the fate of nations, and responsive to material incentives. A wide variety of schemes was introduced and debated at one time or another, but by the end of the nineteenth century patent systems, and especially the model provided by the United States, emerged as the dominant method by which national governments promoted the growth of new technological knowledge. An institution that had been rather obscure a century before had come to be regarded as a virtual necessity for any country with plans to industrialize or modernize.[1]

As emerging nations decided to establish or revise their patent institutions, they were able to draw from the examples and experiences of the three leading industrialized countries of the early nineteenth century, each of which was also something of a pioneer in formulating public policy toward technology. Britain—the first industrial nation—stands out for having established a patent system that has been in operation for a longer period than any other in the world.[2] Patents were granted "by grace of the Crown" and were subject to any restrictions that the government cared to impose, including the expropriation of the patent without compensation. To a large degree by design, patent institutions offered rather limited incentives to inventors with only modest resources or to creators of incremental inventions. Specific features of the British system made it difficult for an inventor who did not already command capital to obtain and use a patent as a well-defined asset to mobilize that capital from others, or to extract a return to his technological creativity by selling it off. The orientation of the British system reflected a widespread view among that country's elite that significant (in the sense of technologically important, not being easily discoverable by many people, and thus worthy of property protection) contributions in technical knowledge were unlikely to come from individuals who did not already have access to the means to absorb the cost of a patent or to exploit the invention directly through a commercial enterprise.

The Statute of Monopolies in 1624 offered a grant of a patent for fourteen years for "the sole making or working of any manner of new manufacture within this realm to the first and true inventor."[3] But in Britain the interpretation of the "first and true inventor" included importers of inventions that had been created abroad, and seems to have also intruded on the determination of whether employers were entitled to patents on the ideas of their workers.[4] Not only were fees set extremely high (five to ten times annual per capita income well into the nineteenth century), but potential patentees were well advised to obtain the help of a patent agent to aid in negotiating the numerous steps and offices that were required for a cumbersome process of application in London. Before 1852 patent specifications were open to public inspection on payment of a fee, but they were not printed, published, or indexed.[5] The complicated system also inhibited the diffusion of information and made it difficult, if not prohibitive, for inventors outside London to conduct patent searches.

The defects of the British system led to numerous investigations and calls for institutional reform, especially after 1829. But it was not until the Crystal Palace Exhibition in 1851, where American inventors shocked observers with their creativity and called attention to their innovative patent institution, that actual legislation was enacted to meet some of the long-standing criticisms. In 1852 the British patent laws were revised in the first major adjustment of the system in two centuries. The patent application process was rationalized into one single Patent Office, and the fee structure was adjusted. A renewal system was adopted, making it cheaper to initially obtain a patent, but one taken to full term remained just as costly as before. The 1852 reforms undoubtedly instituted improvements over the former opaque procedures, but the system remained one based on registration rather than examination, and procedures continued to discourage the technologically creative who did not already have substantial capital to draw on.

This absence of an examination system was, we argue, extremely important. Without examination, there was great uncertainty about what a patent was really worth. Before the 1852 reform, for example, the lack of access to information about the specifications of patents already granted made it difficult to identify whether a purported invention (even one that had been patented) was truly novel and would stand up to challenge. The legal system added to the prevailing uncertainty and was biased against "mere" improvements. Patents were valid only for inventions that were novel (which was, of course, difficult to determine before 1852) and useful, and courts did not hesitate to enforce both conditions. Utility under the patent law was regarded as unrelated to the commercial success of the patented invention. Since the legal system was unpredictable, patent rights could not be regarded as settled unless the patent had been contested in court with a favorable outcome.[6] Moreover, as the law did not offer any relief to the purchaser of a patent that ultimately proved invalid or worthless, potential purchasers were well advised to engage in extensive searches before enter-

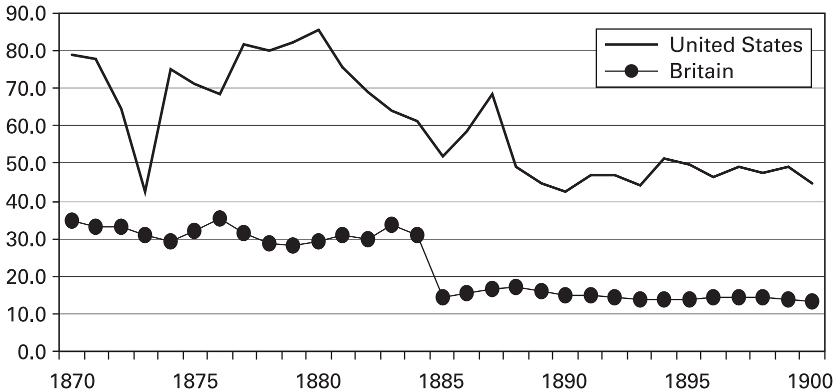


Figure 5.1
The ratio of all assignments to patents in the United States, as compared to the ratio of all assignments and licenses to patents in Britain, 1870 to 1900
Sources: U.S. Patent Office, *Annual Report of the Commissioner of Patents.* Washington, D.C.: G.P.O., various years; and Great Britain Patent Office. *Annual Report of the Commissioners of Patents* [after 1883: *Annual Report of the Comptroller-General of Patents, Designs and Trade Marks.*] London: H.M.S.O., various years.

ing into contracts.[7] When coupled with the lack of assurance inherent in a registration system, the purchase of a patent right involved a substantive amount of risk and high transactions costs—all indicative of a speculative instrument. It is therefore not surprising that the prevalence of assignments and licenses was significantly lower than in the United States (see figure 5.1).[8]

France provides another example of European institutions to promote invention and technical change. French policies consisted of an extensive array of rewards and incentives, and they illustrate the relative benefits and costs of alternative routes to statutory grants of patent rights.[9] Although offering some striking contrasts with the approach taken by Britain, overall they were alike in offering quite limited incentives for technologically creative individuals without either comfortable means or high status to invest in developing their ideas. The cost of obtaining patents was again very high relative to per capita income (much higher than in the United States, though lower than in Britain). Similarly, the reliance on a registration system, limited public disclosure of technical specifications, and the absence of a centralized way to track assignments and patent ownership, all made it difficult for patentees to use their patents as secure and well-defined assets to mobilize capital or to extract significant returns through selling or licensing off his or her rights. France was distinguished by its readiness to make extensive use of other means of rewarding inventors and innovators, but those sorts of returns were uncertain, especially since the processes involved in identifying those deserving of support tended to favor individuals who were already established economically, professionally, or socially.

Before the Revolution, the advance of technology was encouraged by the state granting inventors or introducers of inventions titles, pensions that sometimes extended to spouses and offspring, loans (some interest-free), lump-sum grants, bounties or subsidies for production, exemptions from taxes, or monopoly grants in the form of exclusive privileges. Alternatives to formal privileges illustrate the advantages and disadvantages of awards that were administered by the state on a case-by-case basis. These primarily nonmarket methods of allocation tended to be administratively costly and were imbued with the potential for corruption. It is, and was, evident that a system of grants and privileges could be arbitrary and based on noneconomic criteria. Eighteenth-century correspondence and records provide numerous examples of awards that were made based on court connections. Members of the scientific community who examined applications were not necessarily qualified to assess the potential value of many of the inventions. Moreover, the administrative and opportunity costs of such a system were nontrivial on the part of both supplicants and the state bureaucracy. Inventors were also aware that promises extended to them as inducements were not necessarily enforceable once the inventor had made fixed investments or made his discovery.[10] The technologically creative who were lacking in wealth, status, or connections to individuals with such privileges were at a pronounced disadvantage.

It was this complex network of state policies toward inventors and their inventions that was replaced after the outbreak of the French Revolution. The modern French patent system was established according to the laws of 1791 (amended in 1800) and 1844. The Revolutionary Assembly intended to avoid the excesses involved in previous grants of privileges and proclaimed that it had drafted the outlines of a system that constituted a distinct break with the past. But in effect, as Alexis de Tocqueville pointed out, many features of the institutions of the ancien régime survived the revolution, and this was no less evident in the workings of the patent system.[11] The 1791 statute stipulated patent fees that were costly, ranging from 300 livres through 1,500 livres, and the high price of protection led to difficulties for inventors from ordinary backgrounds. Patentees filed through a simple registration system without any need to specify what was new about their claim, and could persist in obtaining the grant even if warned that the patent was likely to be invalid. Indeed, on each patent document the following caveat was printed: "The government, in granting a patent without prior examination, does not in any manner guarantee either the priority, merit or success of an invention."[12]

The French patent statutes included a statement regarding the right of the public to view patent specifications, which echoed the "bargain" theory of patents that underlay American and British grants. In return for the limited monopoly right, the patentee was expected to describe the invention in such terms that a workman skilled in the arts could replicate the invention, and this information was expected to be "rendue publique." However, since no provision was made for the publication or diffusion of these descriptions, in effect the statutory clause was a dead letter. At least until the law of April 7, 1902, specifications were only available in manuscript form in the office in which they had originally been lodged, and printed information was limited to brief titles in patent indexes.[13] Moreover, the state remained involved in the discretionary promotion of invention and innovation through policies beyond the granting of patents such as cash awards and purchase of patent rights. As a result, inventors had an incentive to direct their attention to rent seeking activities as well as to productive efforts to commercialize their discoveries. Patent assignments were filed in the office of the prefect for the district, but since there was no central source of information it was difficult to trace the records for specific inventions. Like patents themselves, assignments and licenses were issued with a caveat emptor clause. This was partially due to the nature of patent property under a registration system, and partially to the uncertainties of legal jurisprudence in this area. In short, according to an informed nineteenth-century observer, patent rights evinced a "remarkably hazardous and uncertain nature." The basic structure and principles of the French patent system set forth in these early French statutes endured until after World War II.

The framers of the U.S. Constitution and its early laws were bold, ambitious, and optimistic, and they helped set the new nation on an institutional trajectory that was radically different from any in the Old World. One of the areas in which they made this dramatic break was in the patent system they constructed. The so-called Founding Fathers took the design of intellectual property institutions very seriously: for the first time in the world, an intellectual property clause was introduced in a national constitution; the intellectual property clause was approved unanimously, and the law establishing the patent system was one of the first passed by Congress; and prominent figures such as James Madison and George Washington played key roles in spelling out the provisions. The framers quite self-consciously made major changes to the structures employed in Europe, and nearly all of their alterations can be viewed as strengthening and extending incentives and opportunities for inventive activity to classes of the population that would not have enjoyed them under traditional intellectual property institutions.

The framers of the U.S. Constitution and of its early laws were familiar with European precedents, and so it might be reasonably inferred that their innovations in design were self-conscious and deliberate. The intellectual property clause providing for the patent and copyright statutes appears in the very first Article of the U.S. Constitution, whereby Congress was instructed to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." From what record of their thinking survives, the framers were intent on crafting a new type of system that would promote learning, technology, and commercial development, as well as create a repository of information on prior art. Their chosen approach to accomplishing these objectives was based on providing broad access to property rights in technology, which was achieved through low fees and an application process that was impersonal and relied on routine administrative procedures. Congress debated the question of appropriate fees, and the first patent law in 1790 set the rate at the minimal sum of $3.70 plus copy costs. In 1793 the fees were increased to $30 (less than 5 percent of those prevailing in Britain) and were maintained at this level until 1861. In that year, they were raised to $35, and the term was changed from fourteen years (with the possibility of an extension) to seventeen years (with no extensions).

Incentives for generating new technological knowledge were also fine-tuned by requiring that the patentee be "the first and true inventor" anywhere in the world. The law employed the language of the British statute in granting patents to "the first and true inventor," but unlike in Britain, the phrase was used literally, to grant patents for inventions that were original in the world, not simply within U.S. borders.

Moreover, a condition of the patent award was that the specifications of the invention be available to the public immediately on the issuance of the patent.[14] This latter condition not only sped up the diffusion of technological knowledge, but also—when coupled with strict enforcement of patent rights—aided in the commercialization of the technology. That strict enforcement was indeed soon forthcoming. Within a few decades, the federal judiciary evolved rules and procedures to enforce the rights of patentees and their assignees, and clearly considered the protection of the property right in new technological knowledge to be of vital importance.[15] Unlike in Britain, patent property was secured by the "Supreme Law of the land" (the Constitution) and, once granted, could not be overturned except for cases of outright fraud.

Another distinctive feature of the U.S. system was the requirement that all applications be subject to an examination for novelty. For the first few years after the Patent Act of 1790 was passed, a committee composed of the secretaries of state and war, and the attorney general examined the patent applications. This provision proved unwieldy and was replaced by a registration system in 1793, whereby disputes about the validity of a patent were to be resolved by the judiciary. The Patent Act of 1836 reintroduced the examination system, in a structure that remains in use today. Each application was to be scrutinized by technically trained examiners to ensure that the invention conformed to the law and constituted an original advance in technology.[16] Approval from technical experts reduced uncertainty about the validity of the patent and meant that the inventor could more easily use the grant to either mobilize capital to commercially develop the patented technology, or to sell or license off the rights to an individual or firm better positioned to directly exploit it. Private parties could always, as they did under the registration systems prevailing in Europe, expend the resources needed to make the same determination as the examiners, but there was a distributional impact, as well as scale economies and positive externalities, associated with the government's absorbing the cost of certifying a patent grant as legitimate and making the information public.[17] Trade in patented technologies was, as a result, much more extensive—even on a per patent basis—in the United States than elsewhere. Technologically creative people without the capital to go into business and directly exploit the fruits of their ingenuity were major beneficiaries.

It was not coincidental that the U.S. system was extraordinarily favorable to trade in patent rights. From the special provision made in the 1793 law for keeping a public registry of all assignments onward, it is clear that the framers of the system expected and desired an extensive market in patents to develop. It was well understood that the patent system enhanced potential private and social returns to invention all the more, by defining and extending broad access to tradable assets in technological knowledge to a wide spectrum of the population. A market orientation enabled patentees to extract income (or raise capital) from their ideas by selling them off to a party better positioned for commercial exploitation, and thereby encouraged a division of labor where creative individuals specialized in their comparative advantage. The U.S. system extended the protection of property rights to a much broader range of inventions than obtained in Britain or elsewhere in Europe (largely through the lower costs) and, when coupled with effective enforcement of the rights of the "first and true inventor," this meant that inventors could advantageously reveal information about their ideas to prospective buyers even before they received a patent grant. By the mid-1840s, trade in patents (and patenting) was booming, and growing legions of patent agents or lawyers had set up shop in major cities and other localities where rates of patenting were high. Although these agents focused initially on helping inventors obtain patents under the new system, it was not long before they assumed a major role in the marketing of inventions.[18]

Overall, there is no serious question that the United States patent system constituted a dramatic break from European antecedents with respect to policies to promote technological advance. The framers of the U.S. patent institutions held quite different expectations, relative to those of their counterparts in Europe, about the extent to which individuals from relatively ordinary or humble backgrounds could contribute to new technological knowledge, and about how responsive that group would be to expanded opportunities for realizing returns from inventive activity. The structure of the U.S. system was based on the conception that a wide range of individuals, whatever their social origins and standing, were capable of making significant contributions to the advance of technological knowledge, but that in order to realize that potential, broad access to property rights in their knowledge had to be provided. Such rights were especially critical for the technologically creative of limited means. Without clear and secure property rights to their inventions, how could they hope to mobilize the capital to exploit them directly? Certainly, without such property rights (dependent both on their ability to obtain a patent, and on a high likelihood—enhanced by the examination system—that it would stand up in court to a legal challenge), problems of asymmetric information and other high transactions costs would plague the attempts of inventors to attract investors, just as they would complicate the working out of arrangements between employers and employees to encourage the latter to develop their ideas for improvement and offer them up. Access to such rights was to be enhanced by lower fees, by an examination system that bolstered confidence that a patent would stand up to challenge, and by a judicial system that was effective at enforcement. Requirements of immediate full disclosure of technical specifications stimulated further progress, as the technologically creative could more easily learn about and build on what had been discovered. Such measures also encouraged a market in patented technologies that would both aid inventors in using their patents as collateral to attract investments and increase the returns to their inventions.

One would expect this system to have led to a more socially diverse composition of inventors and, in previous work, based on general samples of patentees, we showed how individuals from elite backgrounds accounted for a much smaller proportion of patentees in the United States than they did in Britain during the early nineteenth century.[19] Because many patents are of little or no value, however, this evidence may not conclusively demonstrate that providing broader and stronger incentives for inventive activity was of much technological significance.[20] Indeed, many observers, including those who were influential in maintaining the more socially restrictive patent systems that predominated in Europe until late in the nineteenth century, thought that little in the way of new technology that was novel or important could be expected from individuals who lacked sufficient capital to obtain patents and fund their commercial application: "even with the present expense there are so many trifling patents taken out. If the fee was much higher, parties that are now taking out patents for little speculative things . . . would not take them out. They are something like the dog in the manger; they prevent the public from benefiting by the invention or improvements on it for fourteen years, and yet do not benefit themselves."[21]

Section 5.2 therefore examines the backgrounds and careers of individuals who made contributions to the frontiers of technology, in order to determine the role of the patent system in providing broad access to opportunities for deriving income from investments in inventive activity. The focus on inventors who achieved renown for their contributions to the advance of technology seems fully appropriate for assessing whether the democratic orientation of the U.S. system supported important advances in technology.

5.2 Evidence from the "Great Inventors" in the United States

The idea that patent institutions might matter for the rate of invention, by either a group or the entire population, is based on the judgment that inventors (and those who invest resources to support their activity) are motivated in part by the prospects of realizing material returns. As we have shown in previous work, early nineteenth-century U.S. inventors were indeed highly entrepreneurial and quite sensitive to the potential gains that could be extracted from their discoveries. Some scholars might agree with this characterization, but still contend that the roster of patents was limited to trivial improvements that were only of marginal relevance to the sources of productivity growth and technological change. A common argument is that economic growth depended on discrete advances vested in such "great inventions" as the telegraph, the railroad, and the steam engine and that such ideas were generated through a different process. We drew on the *Dictionary of American Biography* to locate those so-called great inventors who were active in the United States at some point during their careers and born before 1886, and traced their patenting through 1930, in order to examine how different their patterns of behavior were from those of patentees in general. For each of the more than four hundred inventors (all men except for one woman), we collected biographical information as well as the records of a substantial proportion of the patents (roughly 4,500 out of 16,900) they were awarded over their careers.

Overall, we find that the so-called great inventors were quite similar to ordinary patentees. Indeed, to the extent they differed, it was because they were even more entrepreneurial or influenced by markets. Not only did their patenting activity vary procyclically, as did patenting overall, but they often shifted the direction of their inventive activity (as reflected in patents) when exogenous events such as the Civil War altered the relative returns that could be expected from different sorts of inventions.[22] It is also telling that more than 95 percent of the great inventors patented at least some of their inventions, and an even higher percentage took positive action to derive material benefit from them.[23]

Table 5.1
Regional shares of total patents, great inventor patents, and population, 1790–1930

Region	1790–1829 (%)	1830–1845 (%)	1846–1865 (%)	1866–1885 (%)	1886–1905 (%)	1906–1930 (%)
New England						
Patents	34.4%	30.1%	24.7%	19.7%	16.7%	11.4%
G.I. Patents	55.1	34.1	29.6	29.1	29.1	18.3
Population	21.0	13.2	10.1	9.1	7.6	7.2
Middle Atlantic						
Patents	54.5	52.3	48.3	40.6	37.6	30.8
G.I. Patents	35.5	57.7	55.7	51.5	41.1	62.0
Population	34.4	30.0	26.5	23.1	20.5	21.1
Midwest						
Patents	3.0	8.3	20.8	30.3	34.5	36.8
G.I. Patents	1.9	3.2	13.3	13.6	22.9	14.5
Population	3.3	17.3	29.2	34.0	36.0	32.6
South						
Patents	8.1	9.2	5.1	6.0	6.8	10.8
G.I. Patents	7.5	5.0	1.4	1.5	2.3	3.6
Population	41.3	39.7	32.9	31.9	31.5	31.7
West						
Patents	—	—	1.0	3.4	4.6	10.2
G.I. Patents	—	—	0.0	2.9	2.7	1.6
Population	—	—	1.4	1.9	4.5	7.5

Notes and sources: The population figures are from the decadal U.S. Census of Population. The regional distribution of total patents was computed from the Patent Office Annual Reports. The great inventor patents for the period before 1865 include all patents filed by great inventors to that date; after 1865, the distribution of great inventor patents refer to a sample of the patents obtained.

Table 5.1 provides another powerful indication that inventors, and great inventors especially, were concerned with material returns. This is reflected in the tendency for inventors to cluster disproportionately in geographic areas (such as New England and the Middle Atlantic) with better or easier access to low-cost transportation (such as navigable internal waterways) and to the institutional supports underlying the market for technology (such as patent agents and lawyers).[24] This geographic pattern is not explained by geographic differences in schooling levels or in the distribution of manufacturing workers. Rather, it seems consistent with the notion that inventors were more likely to focus on inventive activity if they were in (due either to birth or selective migration) locations where returns to inventive activity were higher.[25] Geographic differentials in patenting seem to have initially (during the early nineteenth century) been rooted in transportation-based disparities in access to broad markets, but persisted over time because institutional supports to carrying out and profiting from invention tended to cluster where patenting was higher. The self-reinforcing pattern was completed when technologically creative individuals chose to move to those places where the market for technology was concentrated, in order to better realize the returns to their specializing at their comparative advantage in inventive activity.[26] A final illustration of how entrepreneurial the great inventors were is the high rate of migration evident in all the birth cohorts (see table 5.2). The rates of interstate migration we estimate from their places of birth and

Table 5.2
Other descriptive statistics on great inventors, by birth cohort

	Pre–1820 birth cohorts		Post–1820 birth cohorts	
	Number	%	Number	%
Father's occupation				
	Inventors		*Inventors*	
Artisan	17	15.6%	36	20.3%
Farmer	43	39.4	37	20.9
Eng./Mach./Inventor	9	8.3	25	14.1
Professional/Merchant	28	25.7	61	34.7
Manufacturer	8	7.3	18	10.2
Other	4	3.7	—	—
Career patents received, by place of birth of inventor, and proportion of patents received in states other than place of birth				
	Patents		*Patents*	
Northern New England	92	87.0%	1697	97.6%
Southern New England	537	55.5	1785	64.7
New York	213	34.7	1872	58.3
Pennsylvania	45	64.4	459	67.6
Southern Middle Atlantic	118	91.5	266	60.3
South	48	64.4	260	87.0
Midwest/West	34	44.1	2670	89.6
Foreign Country	91	100.0	3483	100.0

Notes and sources: See text.

their residences during the years they received patents are much higher than those for the general population at similar ages.[27]

One of the key issues for our study is whether the framers of the U.S. patent institutions were correct in their assumption that individuals from modest or undistinguished backgrounds were capable of playing important roles in pushing out the technological frontier. Although the biographies contain some information about the parents of the great inventors (see table 5.2 for the occupations of their fathers), the relative paucity of data on the wealth holdings of the parents limits their usefulness for our purposes here. Instead, we take advantage of the abundant detail on the formal schooling and training the great inventors received early in their work or career histories. Historians of education agree, and the biographies confirm, that the age at which a young male left school and began to work was strongly associated with the economic resources of the parents during the nineteenth century.[28] This pattern stemmed both from most of the nation's secondary schools and universities being private and thus requiring significant tuition until late in the 1800s, and from the opportunity cost of an individual at school in lieu of at work. The evidence on the extent of formal schooling among the great inventors, therefore, bears not only on the necessity of a high level of formal schooling for making important contributions to technological progress, but also on what sort of material circumstances great inventors came from.

Figure 5.2 displays the distributions of patents across classes of great inventors distinguished in terms of the amount and type of formal schooling they received, and arrayed by birth cohort. It reveals that from the very earliest group (those born between 1739 and 1794) through the birth cohort of 1820–1845, roughly 75 to 80 percent of patents went to those with only primary or secondary schooling.[29] So modest were the educational backgrounds of these first generations of great American inventors, that 70 percent of those born during 1739–1794 had at best a primary education (at least as formally provided), with the proportion dropping to only just above 59 percent among those who entered the world between 1795 and 1819. Given that these birth cohorts were active and indeed dominant until the very last decades of the nineteenth century, these figures unambiguously indicate that people of rather humble backgrounds were capable of making important contributions to technological knowledge. Those who had received some schooling at institutions of higher learning are admittedly overrepresented, since the proportions of cohorts graduating

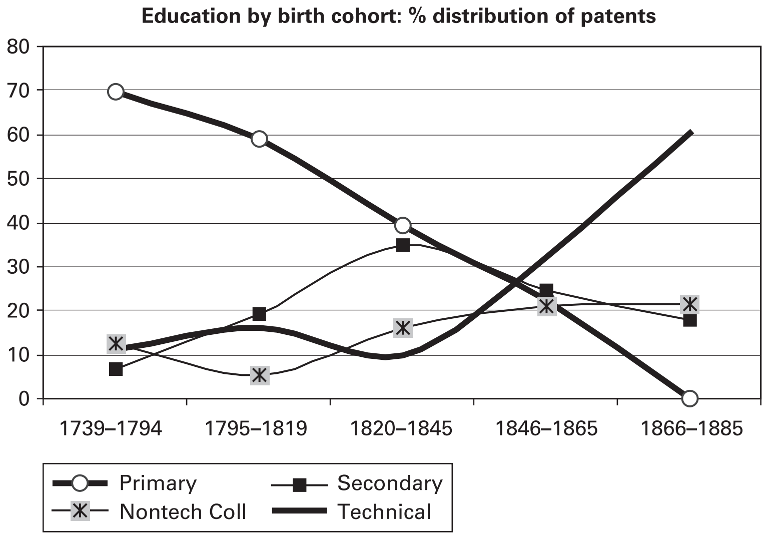


Figure 5.2
The distribution of great inventor patents by formal schooling and birth cohort.
Notes and sources: See the note to table 3 and text.

from secondary school or college were lower than 10 percent and 3 percent respectively as late as 1900.[30] But what is most striking is how individuals who had not enjoyed the advantages associated with a more advanced education accounted for such a large share of major inventions, and that those trained in engineering and/or the natural sciences (in college or beyond) did not play a major role until the birth cohort of 1846–1865. Moreover, in all of the birth cohorts, the great inventors who had only primary or secondary education received as many (and often more) patents over their careers as did their peers with more extensive formal schooling. The less-educated inventors also seem to have produced inventions that were as valuable or technically significant. Their patents were just as likely to be assigned, and just as likely to be cited in applications for patents from later inventors. Thus, the technologically creative seem to have been able to accumulate the skills and knowledge necessary to operate at the frontier largely on their own, or through their work experience as apprentices or younger employees, up until the Second Industrial Revolution.[31]

Some skeptics might suggest that the great inventors who had to make do with little or no formal schooling were not so disadvantaged. That is perhaps the point, at least in regard to the sources of technological creativity, but this should not be interpreted as meaning that this class of great inventors was as well off in material terms as those who went to college. Matthias Baldwin, James Eads, George Eastman, Thomas Edison, and Elias Howe are among the many great inventors who were compelled to go to work at an early age to support themselves or their families, and thus to forego much in the way of formal schooling. An even more fundamental question is whether this class of inventors was especially advantaged by the structure of the U.S. patent system, where the cost of obtaining a property right in the new technological knowledge one had discovered was low, where the state supported strict enforcement of those rights, and where (between 1790 and 1793, and from 1836 onward) the patent office invested substantial resources in determining the validity of patents before they were granted.

The biographies suggest that inventors with only primary or secondary schooling had more limited financial resources than those who were able to attend college. Given the financial institutions of that era, inventors lacking in wealth would surely have found it much more difficult to extract a return from their inventions, if they had to mobilize the capital to start or conduct a business on their own to exploit their idea directly without patent protection. The lower cost of obtaining a patent, and the certification that stemmed from having successfully passed an examination screening, should have made it much easier for inventors to market the new technology and either extract returns by selling off or licensing the rights to a firm better positioned for commercial exploitation, or to attract investment (by offering shares in a firm whose assets consisted largely of the patent rights to the new technology or commitments by the inventor) to support the continued efforts of the inventor.

Our evidence does indeed suggest that these features of the U.S. patent system were highly beneficial to inventors, and especially to those whose wealth would not have allowed them to directly exploit their inventions through manufacturing or other business activity. The ability to obtain patents provided a means for individuals whose chief asset was technological creativity, or accumulated human capital that was conducive to inventive activity, to extract a return from their talents by focusing on invention. Table 5.3 shows that a remarkably high proportion of the great inventors, generally near or above half, extracted much of the income from their inventions by selling or licensing off the rights to them. Moreover, it was just those groups that one would expect to be most concerned about trading their intellectual

Table 5.3
Distribution of great inventor patents by level of education and the major way in which the inventor extracted returns over his or her career: By birth cohorts, 1739–1885

	Level of education				
Birth cohort	Primary	Secondary	College	Engineering / Natural Sciences	Number or % of total
1739–1794 (row %)	69.5	6.8	12.5	11.3	*400*
avg. career patents	5.6	3.8	6.5	5.2	*75*
sell / license (col. %)	54.9	11.1	84.0	17.7	51.4%
prop. / direct (col. %)	36.5	74.1	2.0	44.7	35.6%
employee (col. %)	6.2	7.4	—	—	4.8%
1795–1819 (row %)	59.1	19.3	5.4	16.2	*709*
avg. career patents	20.0	14.4	17.3	12.1	*80*
sell / license (col. %)	58.2	81.0	42.1	60.4	62.1%
prop. / direct (col. %)	33.2	10.2	47.4	24.3	28.1%
employee (col. %)	8.4	8.8	—	13.5	8.8%
1820–1845 (row %)	39.2	34.7	16.3	9.7	*1221*
avg. career patents	41.8	44.0	29.4	23.7	*145*
sell / license (col. %)	50.7	31.8	37.4	72.8	44.0%
prop. / direct (col. %)	42.3	55.2	47.7	19.3	45.5%
employee (col. %)	7.7	13.0	14.9	7.0	10.2%
1846–1865 (row %)	22.2	24.5	20.9	32.4	*1438*
avg. career patents	158.3	73.6	78.6	55.3	*80*
sell / license (col. %)	94.5	68.5	46.2	57.1	66.0%
prop. / direct (col. %)	5.5	18.6	52.8	16.9	22.6%
employee (col. %)	—	12.9	—	23.6	10.4%
1866–1885 (row %)	0.2	17.9	21.4	60.5	*574*
avg. career patents	—	144.5	53.6	155.7	*26*
sell / license (col. %)	—	1.0	46.3	40.1	34.3%
prop. / direct (col. %)	100.0	98.1	49.6	18.7	39.7%
employee (col. %)	—	1.0	4.1	41.2	26.0%

Notes and sources: See the text. The table reports the distribution of great inventor patents across the schooling class of the patentee, by the birth cohort of the inventor; the average number of patents received by each inventor, by birth cohort and schooling class; and the distribution of patents across the principal method of the inventor extracting income, by birth cohort and schooling class. The numbers of patents and great inventors are reported in italics for each birth cohort. The classification of the way income was extracted was arrived at through a close reading of the biographies, and refers to the overall career of the inventor (all of his or her patents). The categories include inventors who frequently sold or licensed the rights to the technologies they patented, those who sought to directly extract the returns by being a principal in a firm that used the technology in production or produced a patented product, and those who were employees of such a firm. We

Table 5.3
(continued)

have omitted a category for those inventors who seem to have made no effort to extract income from their inventions. Our overall sample of Great Inventors was constructed in two waves. In the first wave (160 inventors), consisting primarily of those born before 1821, we collected the information for all of the patents they received through 1865, and retrieved the information on the number they received after 1865 for our estimates of the total career patents. In the second wave (249 inventors), we collected patents from every fifth year through 1930, and thus will be missing the patents received late in the careers of those of our inventors who were born in the 1870s and 1880s.

property that were indeed the most actively engaged in marketing their inventions. The great inventors with only a primary school education were most likely to realize the income from their inventions through sale or licensing, whereas those with a college education in a nontechnical field were generally among the least likely to follow that strategy.[32] With the exception of the birth cohort of 1739 to 1794 (where there are relatively few observations), the college-educated inventors were much more likely than others to extract the returns to their technological creativity by being a proprietor or principal in a firm that directly exploited the technology in production.[33] Inventors who chose to realize the fruits of their technological creativity in this way might not seem to have been so affected by the patent system, but in fact even this group benefited. They were obviously helped by holding a monopoly on the use of the respective technology, but many of them were also aided in mobilizing capital for their firms by being able to report patents (or contracts committing patents granted in the future) as assets. Patent portfolios were especially useful as a signal for those who wished to attract venture capital for exceptionally innovative projects that might otherwise have seemed overly risky.[34]

The estimates in table 5.3 of the relative prevalence of the approaches used by the great inventors to derive income from their inventions also indicate that the reliance on sales and licensing is quite high among the first birth cohort (51.4 percent on average), and remains high (62.1, 44.0, and 66.0 percent in the next three cohorts), until a marked decline among the last birth cohort, those born between 1866 and 1885. The proportion of great inventors who relied extensively on sales or licensing of patented technologies fell sharply from the levels of preceding cohorts, and there was a rise in the proportion that realized their returns through long-term associations (as either principals or employees) with a firm that directly exploited the technologies.

This finding parallels that of Lamoreaux and Sokoloff (1999a), whose analysis of different data indicated that there was a substantial increase in the likelihood of the most productive inventors forming long-term attachments with a particular assignee over the late nineteenth and early twentieth centuries.[35]

The patterns of variation over educational class and time in the relative prevalence of the different approaches employed by inventors in realizing the returns to their inventive activity, and in the relative productivity or prominence of different subgroups at invention, are both fascinating and complex. We have highlighted the role of a revolutionary, low-cost, examination-based patent system that, when coupled with favorable legal institutions and strong enforcement, not only encouraged a broad range of creative individuals and firms to invest more in inventive activity, but was especially crucial for those who began without much in the way of resources except for their technological creativity. A key feature of the story, however, is that much of the population possessed some familiarity with the basic elements of technology during this era. Moreover, apprenticeship or the widespread practice of leaving home during adolescence to pick up skills in a trade, a traditional social institution for the transmission and accumulation of more detailed technological knowledge, was both widely accessible and capable of adapting to many of the new developments and to the general quickening of the pace of advance over the nineteenth century. Technologically creative individuals without the resources to attend institutions of higher learning thus had avenues for acquiring the skills and knowledge necessary to be effective at invention, and they could later take advantage of the access to opportunities for inventive activity grounded in the patent system.

Good things generally come to an end, and in this case circumstances changed over time with the evolution of technology. Formal knowledge of science and engineering became increasingly important for making significant contributions at the technological frontier, particularly with the so-called Second Industrial Revolution, and the cost of carrying out inventive activity rose. Both of these developments served to narrow the range of the population that could generate important inventions, at least to the extent that technologically creative individuals from humble origins found it difficult to gain access to the programs in engineering or natural sciences that proliferated with the expansion of land-grant state universities during the late nineteenth century. Given the much higher costs of conducting inventive activity, those who were supplying the capital to fund such endeavors may have reasonably desired more in the way of credentials, as well as long-term commitments, from those they were supporting. This interpretation is obviously somewhat speculative but does seem to be consistent with the patterns in the data.

An alternative perspective is that many of the phenomena we have noted could be explained by changes in the sectoral composition of the economy. In this view, there were always some industries in which formal schooling in a technical field was nearly a prerequisite for significant invention, while in others inventors could make do with little or no formal schooling. The latter industries, such as agriculture or light manufacturing, may have featured prominently in the early industrial economy, and thus created opportunities for invention for the underschooled, but over time the more capital-intensive and science-based industries grew in importance. The sectoral shifts then led to the dominance among great inventors of those trained in engineering or the natural sciences, as well as to the rise of R&D laboratories in large integrated companies. Although some aspects of this account ring true, the estimates presented of variation in the educational backgrounds of the great inventors across sectors (and over time) in table 5.4 suggest that changes in the sectoral composition of the economy offer little explanatory power. Although inventors in the electrical/communications sector (i.e., electrical machinery and equipment, telegraph, telephone, radio, etc.) were always slightly more likely to have studied engineering or a natural science, in general the differences across sectors seem very small. Instead, the most striking pattern is that the educational backgrounds of inventors tended to move together over time, with each sector characterized by a marked increase in reliance on inventors educated in engineering or natural sciences during the last two birth cohorts. Although our classification of patents by sector is more aggregated than we would like, the data suggest that the change in the composition of inventors overall was driven more by developments extending across all sectors than by changes in the relative importance of different sectors.

Table 5.5 presents further evidence of a broad change over time in who was responsible for the most important technological advances, and the growing significance of supporting investments in human and other capital. As seen in the first panel of this table, there was a pronounced shift toward greater specialization at invention (and use of the patent system). The typical great inventor born after 1820 had a

Table 5.4
Distribution of great inventor patents across sector and education of inventor: By birth cohort, 1739–1885

	Agric.	Const./ Civ. Eng.	Elec./ Comm.	Manuf.	Transp.	Miscell.
1739–1794						
Number of patents	33	27	4	209	99	25
Sector share of patents	8.3%	6.8%	1.0%	52.6%	24.9%	6.3%
Primary (col. %)	60.6	59.3	—	69.4	73.7	84.0
College (col. %)	9.1	11.1	100.0	12.4	13.1	4.0
Eng./Nat. Sci. (col. %)	15.2	14.8	—	13.4	7.1	4.0
1795–1819						
Number of patents	61	37	6	316	218	67
Sector share of patents	8.7%	5.3%	0.9%	44.8%	30.9%	9.5%
Primary (col. %)	68.9	70.3	66.7	56.7	52.3	76.1
College (col. %)	21.3	5.4	—	5.4	—	9.0
Eng./Nat. Sci. (col. %)	4.9	24.3	33.3	14.6	21.6	11.9
1820–1845						
Number of patents	98	110	73	659	118	144
Sector share of patents	8.2%	9.2%	6.1%	54.8%	9.8%	12.8%
Primary (col. %)	24.5	41.8	11.0	44.8	49.2	27.1
College (col. %)	23.5	6.4	23.3	10.8	17.8	38.2
Eng./Nat. Sci. (col. %)	2.0	20.9	17.8	9.0	7.6	6.9
1846–1865						
Number of patents	40	154	413	430	261	128
Sector share of patents	2.8%	10.8%	29.0%	30.2%	18.3%	9.0%
Primary (col. %)	5.0	31.2	28.8	27.9	6.5	6.3
College (col. %)	7.5	19.5	7.8	13.3	35.3	66.4
Eng./Nat. Sci. (col. %)	42.5	35.1	37.5	23.0	33.3	6.3
1866–1885						
Number of patents	7	44	133	213	87	83
Sector share of patents	1.2%	7.8%	23.5%	37.6%	15.3%	14.6%
Primary (col. %)	—	—	—	—	—	—
College (col. %)	28.6	6.8	49.6	23.0	—	2.4
Eng./Nat. Sci. (col. %)	71.4	75.0	50.4	67.6	90.8	18.1

Notes and sources: See the text and the note to table 5.1. The distributions of patents across sectors of intended use are reported for each birth cohort of inventors. Within each sector and birth cohort, the table reports the distribution of patents across the educational level of the great inventor. The omitted schooling class is secondary schooling.

Table 5.5
Length of career at invention and age at first patent, by birth cohort

	Pre–1820 birth cohorts			Post–1820 birth cohorts		
Length of career at invention (yrs. between first and last patent)						
	Number of inventors	%	Avg. career pats	Number of inventors	%	Avg. career pats
0–5 years	37	23.1%	1.7 pats	22	8.5%	2.2 pats
6–10	8	5.0	3.5	8	3.1	6.0
11–20	21	13.1	6.1	36	13.8	25.9
21–30	35	21.9	9.3	45	17.3	30.6
>30 years	59	36.9	23.9	149	57.3	83.5

Age at first patent

Age at first patent	Number of inventors	%	Avg. career pats	Avg. citations	Number of inventors	%	Avg. career pats	Avg. citations
<20	8	5.0%	9.6	1.0	3	1.2%	117.7	8.7
20–24	22	13.8	19.0	0.7	35	13.7	130.6	9.9
25–29	22	13.8	20.0	0.4	72	27.7	64.8	4.2
30–34	35	21.9	17.8	0.5	66	25.4	46.1	4.0
35–39	28	17.5	7.3	0.3	33	12.7	39.0	2.5
40–44	14	8.8	5.9	0.1	18	6.9	28.7	2.8
45–55	21	13.1	4.6	0.0	18	6.9	19.9	1.4

Notes and sources: See text and footnotes to other tables.

much longer career at such an activity than those born before. Sixty percent of post-1820 inventors received patents for more than thirty years (as opposed to roughly 37 percent for the earlier cohort). Moreover, even controlling for the duration of career at invention, the later birth cohort generated many more patents. This tendency toward greater specialization was likely encouraged by the growth of the market for technology (which raised the returns to invention, and also helped inventors who demonstrated creativity in research and development mobilize support) and the increased significance of having accumulated specialized knowledge and human capital for productivity at invention.[36] Changes in the age of first invention may also reflect these developments. As indicated in the second panel of the table, over time great inventors began producing patented discoveries at somewhat younger ages, and on average those who got started earlier ultimately received many more patents over their careers (in total, and per year of career at invention). Accounted for partially by a sharp drop-off in the fraction of great inventors who filed their first patent at thirty-five or older, this distinct pattern seems to suggest that it became more important for the technologically creative to invest in specific human capital or establish a track record early if they were to be productive at invention in later years. It is quite interesting, however, that the great inventors who were older at first patent were proportionally (relative to the number of patents they received) just as likely to receive citations for their achievements; the implication is that the late starters may have generated fewer inventions over their careers, but that the average quality of their discoveries was roughly equivalent to those who began inventing and patenting earlier in life.

5.3 Institutions, Invention, and Economic Development

Invention was a remarkably democratic activity in the United States throughout the nineteenth century. Although individuals who had been able to study at institutions of higher learning were overrepresented among great inventors, those with little in the way of formal schooling were major contributors to the progress of technology. As we have argued before, this era of democratic invention owed much to the broad access to economic opportunities available in an environment where enterprises operated on a small scale, markets were rapidly expanding, and there were relatively modest barriers to entry. In this chapter, however, we called attention to a crucial feature of patent institutions whose role has not been fully appreciated. The U.S. patent system was revolutionary in its extension of property rights in technology to an extremely wide spectrum of the population. Moreover, it was exceptional in recognizing that it was in the public interest that patent rights, like other property rights, be clearly defined, well enforced, and easy to transact in. These were radical notions in a world accustomed to technology being a free good to all who had the capital to exploit it, except as limited by the authority of the government to arbitrarily grant a monopoly over it. It should be no surprise that they encountered fierce resistance in Old World Europe, at least until the exhibition at Crystal Palace intensified concern with the logic and design of intellectual property institutions.[37]

We have demonstrated the lack of support for the views of those nineteenth-century skeptics who contended that only an elite segment was capable of truly important invention, and therefore that an extension of property rights in technology to the general population would have no beneficial effect and might even retard the pace of technical progress. Although few of the celebrated inventors in Britain were of humble origins, such individuals were well represented among the great inventors of the United States. In the United States, this group was more likely to invest in inventive activity, not only because of the relatively lower cost of obtaining a patent, but also because the examination system facilitated the use of a patent as a general asset that could be sold, licensed, or offered as collateral for finance. This latter feature was of profound importance for technologically creative individuals who lacked the financial resources to exploit inventions directly. In short, the patent system was a key institution in the progress of technology, but it also stands out as a conduit for creativity and achievement among otherwise disadvantaged groups.

It might be natural to ask whether the U.S. system of strong and broadly extended property rights in new technological knowledge was so effective at harvesting the technological creativity of its population that it inspired other nations to adopt such a purportedly successful institutional innovation. As we have suggested, Britain and many other European countries did modify their patent institutions, especially after Crystal Palace, to make them more like the American system. On the whole, however, this institutional convergence might be considered somewhat slow, especially the convergence that was achieved without the influence of the international patent conventions convened over the last few decades of the nineteenth century (see table 5.6). Japan and Germany, for example, stand out as the only leading economies that seem to have enthusiastically (in their initial system designs) joined the United States in embracing the examination system (though Russia and the Scandinavian countries also went in that direction, having been much influenced by the German system, with Canada in turn presumably affected by the U.S. practice). How do we make sense of the pattern of institutional diffusion?

Tracing back at least as far as Simon Kuznets and Alexander Gerschenkron, scholars of long-term economic development have wondered whether follower countries naturally evolve, if not benefit from, a systematically different set of institutions than did the early industrializers. Gerschenkron's analysis focused on capital market institutions, but some of his insights might well apply to the question of what sort of patent institutions the follower countries of the nineteenth century should have adopted.[38] As with financial capital, follower countries

Table 5.6
Some descriptive characteristics of patent systems

	Examination system	Working req. or compul. lic.	Pats. for import or introd. of technology	Cost
Europe				
Austria				
1871	—	Y	Y	$$
1899	Y	Y	N	$$$$
Belgium				
1848	N	Y	Y	$$$
1871	N	Y	Y	$$$
1899	N	Y	N	$$
Denmark				
1871	—	Y	—	$$
1899	Y	Y	N	$$
France				
1848	N	Y	Y	$$$
1871	N	Y	—	$$$
1899	N	Y	N	$$$
Germany				
1891	Y	Y	N	$$$$
1899	Y	Y	N	$$$$
Great Britain				
1848	N	Y	Y	$$$$
1871	N	Y	Y	$$$$
1899	N	Y	N	$$$$
Italy				
1871	N	Y	Y	$$
1891	N	Y	Y	$$$
1899	N	Y	Y	$$$
Norway				
1871	Y	Y	N	$$
1899	Y	Y	N	$$
Portugal				
1848	N	Y	Y	$
1899	N	Y	Y	$
Prussia				
1871	—	Y	N	$$

Table 5.6
(continued)

	Examination system	Working req. or compul. lic.	Pats. for import or introd. of technology	Cost
Russia				
1871	—	Y	Y	$$$$
1899	Y	Y	Y	$$$$
Spain				
1848	N	Y	Y	$$$
1871	N	Y	Y	$$$
1899	N	Y	Y	$$
Sweden				
1848	—	Y	Y	—
1871	—	Y	—	$$
1899	Y	Y	N	$$
South and Central America				
Argentina				
1891	N	Y	Y	$$$$
Brazil				
1871	N	Y	Y	$$$
1891	N	Y	—	$$$$
1899	N	Y	N	$$$$
Br. Guiana				
1891	N	N	—	$$$$
Br. Honduras				
1891	N	N	—	$$$$
Chile				
1891	N	Y	—	$$$
Colombia				
1891	N	Y	—	$$$
Cuba				
1871	N	Y	Y	$$$
Ecuador				
1891	N	Y	Y	$$
Guatemala				
1891	N	Y	Y	$$$
Mexico				
1871	N	Y	N	$$$$
1899	N	N	N	$$$$

Table 5.6
(continued)

	Examination system	Working req. or compul. lic.	Pats. for import or introd. of technology	Cost
Peru				
1891	N	Y	N	$$$$
Uruguay				
1891	N	Y	—	$$$$
Venezuela				
1891	N	Y	N	$$$$
Others				
Barbados				
1891	N	Y	—	$$$
Canada				
1871	—	Y	N	$
1899	Y	Y	N	$
Fiji				
1891	N	N	—	$$
Hawaii				
1891	Y*	N	Y	$$
India				
1891	N	Y	—	$$$$
Jamaica				
1891	N	Y	—	$$$
Japan				
1899	Y	Y	N	$$
Liberia				
1891	Y*	Y	—	$$$
New South Wales				
1891	N	N	N	$
New Zealand				
1891	N	Y	N	$$
South Africa				
1891	N	N	—	$$$
United States				
1848	Y	N	N	$
1899	Y	N	N	$

Table 5.6
(continued)

Sources:
1848: John Kingsley and Joseph Piesson, *Laws and Practice of All Nations and Governments Relating to Patents of Inventions* (New York: Kingsley and Piesson, 1848).
1871: *United States and International Patent Office Manual* (New York: Fitch and Co., 1871).
1891: *Epitome of the World's Patent Laws and Statistics* (New York: The British and European Patent Agency, 1891).
1899: Arthur Greeley, *Foreign Patent and Trademark Laws* (Washington, D.C.: John Byrne and Co., 1899).
Note: "Y" and "N" indicate whether patent systems had the characteristics of the column headings ("—" indicates uncertain). Patent costs are the total official fees for the full patent life. $, $$, $$$, and $$$$ indicate cost intervals of $0–$100, $100–$250, $250–$500, and >$500, respectively.
* Although this country officially had an examination system, its patent office does not appear to have carried out a serious examination of applications for novelty or utility.

likely had difficulty mobilizing all forms of capital to invest in inventive activity, and thus it was perhaps not unreasonable for them to choose to rely on technological knowledge from abroad. Early industrializers, whose populations and industries were more familiar with the frontiers of technology, would not have been so inclined.

It should not be too surprising that those who came after tended to adopt patent systems that were oriented more toward securing flows of technology from outside the country, especially technological knowledge that was embodied in actual plant or production. Except for the United States, and a few relatively minor exceptions, nearly all societies had working requirements (or, in a few cases, compulsory licensing) as a central component of their patent systems. Similarly, the great majority of follower countries had provisions for so-called patents of importation—whereby those who were the first to introduce a new technology to the country (regardless of whether he or she was the inventor) from abroad could obtain a property right (typically lasting until the original foreign patent on the technology expired).[39] It was only late in the century, after international patent conventions—under pressure from the United States—urged countries to adopt the principle that only inventors had the right to a patent, that such awards of property rights to the importer of a technology were disallowed. Even then, many countries specified that the first applicant for a patent would be presumed to be the inventor.

Many argue that extremely underdeveloped countries should have no interest in maintaining a patent system, on the grounds that their citizens would not be very likely to make new contributions to technological knowledge, and thus the only impact would be to increase the flow abroad of licensing fees and other payments for the fruits of technology. This attitude, however, has an uncomfortable resemblance to that early nineteenth-century view that ordinary people could not be expected to produce any truly significant new technical knowledge. We have shown that the latter opinion was incorrect, and though circumstances today are very different, it is far from clear that strong patent systems in developing countries will fail to elicit a supply response in domestic invention. Moreover, one can also justify the implementation of an effective patent system if it were necessary to make substantial investments to migrate technologies from abroad or if the cooperation or assistance of the original inventor was an important factor in diffusion. Although Taiwan of the mid-1980s was and is hardly typical of most developing countries, in his careful investigation of the strong patent system suddenly imposed on the country by the United States, Lo offers persuasive evidence of a powerful response in expenditures on research and development, patenting by Taiwanese residents in the United States, and foreign direct investment.[40] More study is needed, but it seems to us premature to dismiss the idea that there could be some formidable benefits to developing countries from strengthening their protection of property rights to new technology, especially if such moves involved working requirements.

Notes

1. Especially after the Crystal Palace Exhibition of 1851, there was a continued fascination with issues about patent systems, including whether patent systems were desirable for all countries, how they should be designed, and (for individuals) how to make money from them. See Machlup and Penrose 1950 for an excellent account of one part of the debate. Mark Twain ([1889] 1997) was not far from the mainstream when he spoke through the "Connecticut Yankee in King Arthur's Court": "[T]he very first official thing I did, in my administration—and it was on the very first day of it too—was to start a patent office; for I knew that a country without a patent office and good patent laws was just a crab and couldn't travel any way but sideways or backways" (70).

2. The standard references for the early British patent system are Christine MacLeod 1988 and Dutton 1984.

3. 21 Jac. I. C. 3, 1623, Sec. 6.

4. See MacLeod 1999 for a discussion of how craftsmen in Britain had to rely on other methods of extracting returns from their ideas about how to improve on technical practice.

5. Since the patent could be filed in any of three offices in Chancery, searches of the prior art involved much time and inconvenience. It is hardly surprising that the defenders of

the early patent system included patent agents and patent lawyers. Patent fees also provided an important source of revenues for the Crown and its employees, and created a class who had strong incentives to block proposed reforms.

6. According to an editorial published in *Newton's London Journal* in 1862, "there can be no doubt that a large amount of property is bound up in patent rights, and that the utmost uncertainty exists as to the legal value of that property" (qtd. in Coulter 1991, 140). Other constraints on the market for inventions related to policies toward assignments. Ever vigilant to protect an unsuspecting public from fraudulent financial schemes on the scale of the South Sea Bubble, ownership of patent rights was limited to five investors (later extended to twelve).

7. The case law on licenses was more convoluted. See, for instance, *Lawes v. Purser*, 6 Ell. and Bl. 930, where a licensee refused to continue payments on the grounds that the patent was void. It was held that the licensee could not make such a defense as long as the contract for the invalid patent had been executed without fraud.

8. The markedly higher ratio of assignments to patents in the United States is all the more significant, both because the British figures are biased upward by the inclusion of licenses, and because the higher costs of obtaining a patent in Britain should, at least in principle (if screening by cost was a good substitute for screening by examination), have led to patents of higher than average quality.

9. Excellent assessments of such issues during the Enlightenment include Hilaire-Pérez 1994 and 2000.

10. The famous English textile inventor, John Kay, illustrates the asymmetries involved in individual bargains struck with state authorities. Kay settled in France because of promises to subsidize the transfer of technology and substantially aided in the diffusion of textile machinery. The Society for the Encouragement of Arts and Manufacturing in England promised him a generous award to return there, but then reneged once he was in London. Kay wrote early in 1761 to Prudaine de Montigny, Conseiller d'Etat in London, to explore the possibility of receiving French financial aid if he again immigrated to Paris. Later that same year, Kay wrote to M. de Brou, Intendant de Rouen, to complain that he was still not receiving the pension he had been promised.

11. See the Decret du 30 Decembre 1790, in the Code des Pensions, 45. Although the legal rhetoric implied that the primary intent of the legislation was to recognize the natural rights of inventors, the actual clauses led to results that were different and that reflected former mercantilist policies. In an obvious attempt to limit international diffusion of French discoveries, until 1844 patents were voided if the inventor attempted to obtain a patent overseas on the same invention. The first introducer of an invention covered by a foreign patent would enjoy the same "natural rights" as the patentee of an original invention or improvement, although the term would expire at the same time as any foreign patent on the item. In order to qualify for a patent of importation, the applicant had to have obtained practical knowledge of how the item worked through personal risk and effort, although he was not obliged to prove that the invention had been patented elsewhere or to even state its country of origin. The rights of patentees were also restricted if the invention related to items that were controlled by the French government, such as printing presses and firearms.

12. Since France during the ancien régime was likely the first country to introduce systematic examinations of applications for privileges, it is somewhat ironic that commentators point to the retention of registration without prior examination as the defining

feature of the "French system." In 1968 a partial examination system was adopted that was similar to the early British reforms along these lines, since it did not include a search for novelty, merely a test for accordance with the law: "[il] se situe a mi-chemin entre la libre deliverance et l'examen prealable . . . en effet, l'administration n'avait pas les moyens de pratiquer un tel examen." (Marcellin 1983, 21). The changes were made to give value to patents and to protect the interests of third parties. It was only in 1978 that an examination for novelty was introduced.

13. The law of 1844 only allowed for the publication of the full text of patents that were judged to be important. The attempt to obtain information was also inhibited by restrictions placed on access: viewers had to state their motives; foreigners had to be assisted by French attorneys; and no extract from the manuscript could be copied until the patent had expired. "C'est donc bien avec la loi de 1902 que le brevet a definitivement perdu son charactere de document d'archives" (Brevets d'Invention Français, 1791–1902, 12).

14. American legislators were concerned with ensuring that information about the stock of patented knowledge was readily available and diffused rapidly. As early as 1805 Congress stipulated that the secretary of state should publish an annual list of patents granted the preceding year, and after 1832 it also required the publication in newspapers of notices regarding expired patents. The Patent Office in Washington was a source of centralized information on the state of the arts, but it also maintained repositories throughout the country, where inventors could forward their patent models at the expense of the Patent Office. Rural inventors could apply for patents without significant obstacles, because applications could be submitted by mail, free of postage.

15. See Khan 1995 and 2005. Supreme Court Justice Joseph Story, the acknowledged intellectual property expert of the early courts, succinctly stated the dominant perspective in *Lowell v. Lewis* (15 F. Cas. 1018 [1817]): "[T]he inventor has a property in his invention; a property which is often of very great value, and of which the law intended to give him the absolute enjoyment and possession . . . involving some of the dearest and most valuable rights which society acknowledges, and the constitution itself means to favor."

16. Although the statutes proposed to grant patents for "new and useful" inventions, in practice the utility claim was never enforced. Courts declared that it was up to the market, not to administrators, to determine what was useful. In the 1817 case, *Lowell v. Lewis* (15 F. Cas. 1018), Joseph Story charged the jury that the utility of the invention "is a circumstance very material to the interest of the patentee, but of no importance to the public. If it is not extensively useful, it will silently sink into contempt and disregard." It was the role of the market, rather than the courts, to determine the ultimate success of the patent. This policy was continued by the Patent Office, which also did not attempt to gauge the social or technical value of an invention, deciding conflicting claims predominantly on the basis of novelty.

17. Dutton 1984; MacLeod 1988, 1999.

18. By the mid-1840s, for example, a number of national patent agencies had begun to publish periodicals (such as *Scientific American*) that popularized invention as a career path for the ambitious and talented. Over time, intermediation in this market for technology grew ever more articulated in a process not unlike the evolution of financial intermediaries. Patent agents and lawyers became increasingly specialized and were drawn into activities such as the provision of advice to inventors about the prospects for various lines of inventive activity, and the matching not only of buyers with sellers of patents but also of inventors with individuals seeking to invest in the development of new technologies.

As the extent of the market for technology expanded over the course of the nineteenth century, creative individuals with a comparative advantage in technology appear to have increasingly specialized in inventive activity. This tendency was likely reinforced by the increasing importance to inventors of specialized technical knowledge as technology became more complex. For evidence and more discussion, see Lamoreaux and Sokoloff 1996, 1999a, 1999b, 2001, 2003; Khan and Sokoloff 1993, 2001; and Khan 2005.

19. Khan and Sokoloff 1998.

20. Sokoloff 1988; Sokoloff and Khan 1990; Khan and Sokoloff 1993.

21. So testified Charles Few to the Select Committee on the Law Relative to Patents for Invention, on May 15, 1829. See British Parliamentary Papers 1968, vol. 1, 48.

22. Sokoloff 1988; Khan and Sokoloff 1993.

23. This claim is based on a presumption that the great inventors who were employees (a distinct minority as will be made clear) did take positive action to derive benefit. From the biographies we have read, there were only a few inventors who did not try to realize some returns from their efforts. Nearly all patented their inventions, but some did not. Only three of the cohort born after 1820 did not obtain patent protection.

24. In previous work, we showed that both overall inventors, and especially great inventors, were highly disproportionately concentrated in counties with access to water transportation (prior to railroads). See Sokoloff 1988 and Khan and Sokoloff 1993. For evidence that patent agents and other indicators of the market for technology came to be disproportionately concentrated in those areas (mostly in New England and the Middle Atlantic), see Lamoreaux and Sokoloff 1996, 1999a, 1999b, and 2003.

25. See Khan and Sokoloff 1993 for evidence that great inventors during the early nineteenth century were both more likely to be born in counties with low-cost access to broad markets and to migrate to such counties. Both factors contributed to the highly disproportionate concentration of great inventors in such counties.

26. See Lamoreaux and Sokoloff 1996, 1999a, 1999b for more discussion.

27. Joseph Ferrie very kindly provided estimates of the interstate migration rates from the random sample of the native-born population that he has collected from the late nineteenth-century public use census records, for groups at comparable ages. He agreed that the great inventors seem to have been much more geographically mobile than the general population. It is also worth noting how the foreign-born are disproportionately represented among the great inventors of the United States. This is consistent with the idea that the technologically creative in Europe were more likely to migrate across the Atlantic because of greater opportunities, some of them associated with the patent system.

28. See Cubberley 1920 for a discussion of how schools, and the backgrounds of the students who attended them, evolved over the nineteenth century.

29. Those classified as receiving only a primary education encompass a range from those who spent no time in school to those who attended school until about age twelve. Those who were identified as spending any years in an academy or who attended school after the age of twelve (but did not attend a college or seminary) were placed in the secondary school category. Those who spent any time at all in college were either counted in the college category, or—if they had attended a school with an engineering orientation or

followed a course of study in medicine or a natural science—in the engineering/natural science category.

30. See Snyder 1993, Figures 11 and 17. The rates of graduation from secondary school and college were markedly higher in the Northeast and East North Central regions, where inventive activity was disproportionately concentrated throughout the nineteenth century, however. See Cubberley 1920 and 1947 for discussion of the regional patterns in schooling.

31. The differences in patent systems had implications for how apprenticeship worked, and the effective rights of workers to technological improvements they generated, in Britain and the United States. See Fisk 1998 and MacLeod 1999.

32. Although a bit less striking, the inventors who had studied engineering or a natural science were also, for a time (the middle three birth cohorts), much more inclined to rely on sales or licensing of their inventions to realize income. This pattern might be explained as due to these inventors choosing to specialize in what their human capital gave them a comparative advantage in—inventive activity—and leaving it to others to carry out the commercial exploitation.

33. Many of the college-educated of the 1739–1794 birth cohort were evidently not so concerned with realizing a return from their inventions. Fourteen percent of the college educated, and more than one third of those who studied engineering or natural science, chose not to pursue returns to their inventions. This attitude, however admirable, was not shared by inventors who came from less privileged backgrounds.

34. Lamoreaux, Levenstein, and Sokoloff 2004.

35. Although deeply impressed with how well this result fits with the work of Lamoreaux and Sokoloff (1999a, 2003), some caution may be warranted. Because the *Dictionary of American Biography* (1928–1936) was originally prepared during the 1920s, our sample does not include as many great inventors born after 1865 as we would like. We would feel a bit more secure with more observations.

36. Lamoreaux and Sokoloff 1999a.

37. Machlup and Penrose 1950; Penrose 1951; Rosenberg 1969.

38. Gerschenkron explicitly identified the ability to draw on the technologies developed in those countries that had industrialized earlier as one of the greatest advantages of economic backwardness. See his classic essay in Gerschenkron 1976.

39. Many of the follower countries assessed very high fees for patents (especially those that are less developed) at rates that are all the more remarkable for the relatively low per capita incomes prevailing there. Indeed, many of the societies in Central and South America are distinguished for having among the highest fees in the world for patent protection, and the pattern holds across quite a range of national institutional heritages (e.g., Brazil, British Honduras, and Peru). These fees might have been so high because the elites in these extremely unequal societies did not find it in their interest to provide broad access to property rights in new technology. Given that the number of outliers in Central and South America, where inequality in wealth and political influence was extreme, as compared to, say, Spain and Portugal (which also had rather low per capita incomes), this hypothesis should perhaps not be immediately dismissed. It might also be noted that the influence of colonial heritage is not nearly so powerful as one might have expected. There was, for example, enormous diversity in the characteristics of the patent systems of the remaining British colonies. Another rationale is that it provided an effec-

tive way of deriving government revenue at the expense of foreigners (who in many of these countries accounted for well over half of all patents).

40. Lo 2004.

References

Brevets d'invention français. 1791–1902. un siècle de progrès technique/Ministère de l'Industrie et du Commerce.

British Parliamentary Papers. 1968. *Reports from Select Committees on the Law Relative to Patents for Inventions, 1829–72*. 2 vols. Shannon: Irish University Press.

Coulter, Moureen. 1991. *Property in Ideas: The Patent Question in Mid-Victorian England*. Kirksville, MO: Thomas Jefferson Press.

Cubberley, Ellwood P. 1920. *The History of Education*. Boston: Houghton Mifflin.

Cubberley, Ellwood P. 1947. *Public Education in the United States, A Study and Interpretation of American Educational History*. Boston: Houghton Mifflin.

Dictionary of American Biography. 1928–1936. New York: Scribner's.

Dutton, Harold I. 1984. *The Patent System and Inventive Activity during the Industrial Revolution, 1750–1852*. Manchester, UK: Manchester University Press.

Fisk, Catherine L. 1998. "Removing the 'Fuel of Interest' from the 'Fire of Genius': Law and the Employee-Inventor, 1830–1930." *University of Chicago Law Review* 65 (Fall): 1127–1198.

Gerschenkron, Alexander. 1976. *Economic Backwardness in Historical Perspective.* Cambridge, MA: Belknap Press.

Hilaire-Pérez, Liliane. 1994. "Inventions et Inventeurs en France et en Angleterre au XVIIIè Siècle." Ph.D. diss., Université de Lille, Lille.

Hilaire-Pérez, Liliane. 2000. *L'invention technique au siècle des Lumières.* Paris: Albin Michel.

Khan, B. Zorina. 1995. "Property Rights and Patent Litigation in Early Nineteenth-Century America." *Journal of Economic History* 55 (March): 58–97.

Khan, B. Zorina. 2005. *The Democratization of Invention: Patents and Copyrights in American Economic Development, 1790–1920.* New York: Cambridge University Press.

Khan, B. Zorina, and Kenneth L. Sokoloff. 1993. "Schemes of Practical Utility." *Journal of Economic History* 53 (June): 289–307.

Khan, B. Zorina, and Kenneth L. Sokoloff. 1998. "Two Paths to Industrial Development and Technological Change." In *Technological Revolutions in Europe, 1760–1860*, ed. Maxine Berg and Kristine Bruland, 292–313. Cheltenham: Edward Elgar.

Khan, B. Zorina, and Kenneth L. Sokoloff. 2001. "The Early Development of Intellectual Property Institutions in the United States." *Journal of Economic Perspectives* 15, no. 3: 233–246.

Khan, B. Zorina, and Kenneth L. Sokoloff. 2004. "Institutions and Democratic Invention in 19th Century America." *American Economic Review* 94 (May): 395–401.

Lamoreaux, Naomi R., Margaret A. Levenstein, and Kenneth L. Sokoloff. 2004. "Financing Invention during the Second Industrial Revolution: Cleveland, Ohio, 1870–1920." Unpublished working paper.

Lamoreaux, Naomi R., and Kenneth L. Sokoloff. 1996. "Long-Term Change in the Organization of Inventive Activity." *Proceedings of the National Academy of Sciences* 93 (November): 12686–12692.

Lamoreaux, Naomi R., and Kenneth L. Sokoloff. 1999a. "Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries." In *Learning By Doing in Markets, Firms, and Countries*, ed. Naomi R. Lamoreaux, Daniel M. G. Raff, and Peter Temin. Chicago: University of Chicago Press.

Lamoreaux, Naomi R., and Kenneth L. Sokoloff. 1999b. "The Geography of the Market for Technology in the Late-Nineteenth and Early-Twentieth Century United States." In *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, ed. Gary D. Libecap, 67–121. Stamford, CT: JAI Press.

Lamoreaux, Naomi R., and Kenneth L. Sokoloff. 2001. "Market Trade in Patents and the Rise of a Class of Specialized Inventors in 19th Century U.S." *American Economics Review Papers and Proceedings* 91 (May): 39–44.

Lamoreaux, Naomi R., and Kenneth L. Sokoloff. 2003. "The Decline of the Independent Inventor: A Schumpeterian Story?" Unpublished working paper.

Lo, Shih-Tse. 2004. "Strengthening Intellectual Property Rights: The Experience of the 1986 Taiwanese Patent Reforms." Ph.D. diss., University of California, Los Angeles.

Machlup, Fritz, and Edith Penrose. 1950. "The Patent Controversy in the Nineteenth Century." *Journal of Economic History* 10 (March): 1–29.

MacLeod, Christine. 1988. *Inventing the Industrial Revolution*. Cambridge: Cambridge University Press.

MacLeod, Christine. 1999. "Negotiating the Rewards Of Invention: The Shop-Floor Inventor in Victorian Britain." *Business History* 41, no. 2: 17–36.

Marcellin, Yves. 1983. *La Procédure française de délivrance des brevets d'invention*. Rosny-Sous-Bois: Editions Cédat.

Penrose, Edith Tilton. 1951. *The Economics of the International Patent System*. Baltimore, MD: Johns Hopkins University Press.

Rosenberg, Nathan, ed. 1969. *The American System of Manufactures: The Report of the Committee on the Machinery of the United States 1855, and the Special Reports of George Wallis and Joseph Whitworth 1854*. Edinburgh: Edinburgh University Press.

Sokoloff, Kenneth L. 1988. "Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790–1846." *Journal of Economic History* 48 (December): 813–850.

Sokoloff, Kenneth L., and B. Zorina Khan. 1990. "The Democratization of Invention during Early Industrialization: Evidence from the United States, 1790–1846." *Journal of Economic History* 50 (June): 363–378.

Twain, Mark. *A Connecticut Yankee in King Arthur's Court.* [1889] 1997. New York: Oxford University Press.

6 On the Efficacy of Reforms: Policy Tinkering, Institutional Change, and Entrepreneurship

Murat Iyigun and Dani Rodrik

6.1 Introduction

The conventional model of economic policy that inspired the wave of reform in developing and transitional economies during the last two decades comes with a standard list of prescriptions: establish property rights, enforce contracts, remove price distortions, and maintain macroeconomic stability. Once these things are done, economies are supposed to respond predictably and vigorously.

Recent experience around the world has not been kind to this vision of reform. Countries such as China, India, and Vietnam have embarked on high growth while retaining policies and institutional arrangements that are supposed to be highly inimical to economic activity (e.g., absence of private property rights, state trading, large amounts of public ownership, high barriers to trade). Meanwhile, countries that have enthusiastically adopted the standard institutional reforms—such as those in Latin America—have reaped very meager growth benefits on average with considerable variance in actual outcomes. This experience has raised doubts as to whether we have a good fix on what makes growth happen. As Al Harberger recently put it, "When you get right down to business, there aren't too many policies that we can say with certainty deeply and positively affect growth" (2003, 215; see also Rodrik 2003).

We develop a framework in this chapter that tries to make sense of this heterogenous experience with policy reform. Our starting point is the idea that a key obstacle to economic growth in low-income environments is an inadequate level of entrepreneurship in nontraditional activities. As a recent paper by Imbs and Wacziarg (2003) documents, countries grow rich by increasing the range of products that they produce, not by concentrating on what they already do well. Productive diversification requires entrepreneurs who are willing to invest in activities that are new to the local economy. Such entrepreneurship can be blocked both because the policy environment is poor in the conventional sense—namely, property rights are protected poorly, there is excessive taxation, and so on—*and* because markets do not generate adequate incentives to reward entrepreneurship of the needed type. Our chapter takes both obstacles seriously.

The central market failure that we consider in relation to entrepreneurship is an information externality. As in Hausmann and Rodrik 2003, we assume that production costs of modern, nontraditional activities are unknown and can be discovered only by making sunk investments. Once an entrepreneur discovers costs of a given activity, this information becomes public knowledge, prompting imitative entry with a lag (if entry is profitable). Hence, an entrepreneur provides a useful "cost discovery" function, but can reap at best only part of the gains from his effort. If he discovers a profitable activity, his profits are soon dissipated; if he makes a bad investment, he bears the full cost of his mistake. Under these conditions, entrepreneurship is underprovided and structural change is too slow.

We embed this model of entrepreneurial choice in a framework that allows policy reforms of different kinds. We assume the policymaker has access to two strategies, both of which have the potential to increase productivity but produce uncertain outcomes. The first is "policy tinkering," whereby the policymaker is allowed to draw a new policy from the preexisting "policy regime." The second is "institutional reform," whereby a policy draw is made from a different policy regime, at the price of imposing an adjustment cost on incumbent firms.[1] The latter is meant to capture more radical reforms that alter underlying institutional arrangements. Consider, for example, the difference between reducing the corporate tax rate and making a switch from import substitution to export orientation. The former is an instance of tinkering within an existing set of institutional arrangements. From our standpoint, its most important characteristic is that it operates neutrally between existing firms and new firms. If a reduction in corporate taxes increases the profitability of investment in the modern sector, it does so both for incumbent firms and for potential entrants. By contrast, a switch from one trade regime to another is not neutral: it imposes a cost on the incumbents, while new ventures (in export-oriented activities) are unaffected or helped.

While institutional reform engenders an adjustment cost, this cost also presents a subtle potential advantage over policy tinkering. Tinkering is unable to induce greater amounts of cost discovery and new entrepreneurship precisely because it does not affect the margin between old and new activities. Institutional reform can induce greater cost discovery where policy tinkering would fail to do so. Therefore, there are circumstances under which institutional reform will dominate policy tinkering, even when the shift in the policy regime itself does not confer any direct economic benefit.[2]

Our framework therefore yields new insights on the circumstances under which different types of policy reform—policy tinkering versus deeper institutional reforms—are likely to foster structural change and economic growth. We find that the relative benefits of institutional reform depend critically on the vigor of entrepreneurship in the modern sector of the economy. Institutional reform is likely to dominate policy tinkering only for *intermediate* levels of cost discovery. When prevailing levels of cost discovery (and the associated levels of productivity) are too low, policy tinkering is adequate to generate new entrepreneurship; when they are already high, institutional change is unable to stimulate additional entrepreneurship.

One empirical implication of our framework is that, conditional upon institutional reforms having been undertaken, we should observe a systematic relationship between the success of such reforms and the prevailing state of entrepreneurship in the modern sector. In particular, institutional reforms should produce a boost in economic activity in countries where the modern sector was languishing due to a lack of cost discovery attempts, but fail in places where a relatively productive modern sector already existed (thanks to a healthy dose of entrepreneurship). The main difficulty in testing this idea is that we do not have reliable and cross-nationally comparable indicators of entrepreneurship. In the empirical analysis that we carry out in the appendix, we use the level of self-employment in non-agriculture as our proxy. Controlling for levels of income, this can be viewed as a crude indicator of the propensity of people to start their own businesses. Using this proxy, we are able to provide some formal evidence that we think is supportive of our model. We show, in particular, that the success of institutional reform depends critically on the level of our proxy for entrepreneurial experimentation and cost discovery. Institutional reform appears to have worked when this proxy is indicative of low levels of prevailing entrepreneurship, and failed otherwise.

In sum, our approach yields a rich set of normative and positive implications. On the normative side, it helps to identify the circumstances under which different types of policy reform—policy tinkering versus deeper institutional reforms—are likely to foster structural change and economic growth. On the positive side, our model offers insights as to why institutional reforms have worked in a handful of countries and failed in many others.

The outline of the rest of the chapter is as follows. Section 6.2 lays out the basic economic environment. Section 6.3 describes the market equilibrium. The outcomes under policy tinkering and institutional reform are discussed in section 6.4. Section 6.5 presents the case where institutional reform has a clear-cut advantage over tinkering. Section 6.6 presents some evidence on three of the empirical implications of our framework. Finally, section 6.7 provides concluding observations.

6.2 The Model

We consider a model of a small open economy with two sectors, modern and traditional. These two sectors differ according to whether costs of production are known. The modern sector is made up of Ψ goods with uncertain costs, none of which is produced at time zero. We assume there are two factors that determine the cost of producing a modern-sector good. First, there is a *policy-specific* cost component, a. This variable, which is observable, represents the impact of the policy environment on entrepreneurial productivity. We assume that the distribution of a is uniform over the interval $[0, 2]$. Hence, $E(a) = 1$. Second, there is a *good-specific* cost, ψ_i. This variable, which is unobservable until production of good i starts, represents the productivity of good i for a given policy draw a. We assume that the ex ante distribution of ψ_i is uniform over the interval $[0, \Psi]$.

Therefore the cost of production in the modern sector can be written as

$$c_t^i = \frac{w_t}{a_t \psi_i}, \tag{1}$$

where c_t^i denotes the unit costs of producing good i when policy a_t is in effect at time t, w_t is the wage rate in period t, and $1/a_t\psi_i$ is the number of workers needed to produce a single unit of the good i . Modern-sector production uses only labor and has constant-returns-to-scale technology once productivity is known.

The justification for the uncertainty about costs of production in the modern sector is provided by the fact that production involves learning along various different dimensions. For instance, producing a good that has not been locally produced previously requires learning about how to combine different inputs in a given environment, figuring out whether the existing local conditions are conducive to efficient production, discovering the true costs of production, and so on (see Hausmann and Rodrik 2003). In addition, our framework captures the idea that some policy environments are better for entrepreneurship than others.

We note that the unobserved productivity parameters ψ_i is a property of individual goods, and not of entrepreneurs: all entrepreneurs who run firms producing good i will operate with productivity ψ_i.[3] We assume that each modern-sector firm is of a given size, fixed (by appropriate choice of units) to one unit of good i's output. Each entrepreneur can run one, and no more than one, modern-sector firm.

Discovering ψ_i requires setting up the firm and utilizing one unit of labor.[4] Let m_t denote the number of entrepreneurs who choose to establish firms in period t, which also equals the total amount of (sunk) labor investment in the same period. After firms are set up and labor is sunk, ψ_i's become known for those m_t goods in which investments have been made. Subsequently, all m_t entrepreneurs can produce a unit of the good and earn p (an exogenous price fixed on world markets).[5] During this inaugural production stage, which we call the "cost discovery" phase, there is no entry into the modern sector so that any entrepreneur who draws a cost less than or equal to p earns excess profits. (Even though p is fixed, so is output due to the assumptions that firm size is fixed and an entrepreneur cannot run more than a single firm.) This transitional period of monopoly profits can be motivated in one of two ways. It could be that it takes time for the ψ_i to become common knowledge. Alternatively, ψ_i can be immediately known, but it could take time for an "imitator" to set up a firm. Note that while some firms will make profits in the cost-discovery phase, the ex ante expected profits from starting a new firm would be zero in equilibrium. This is because the quantity of entrepreneurship, m_t, is determined endogenously.

Following the cost-discovery phase, production in the modern sector enters the "consolidation" phase, in which there is free entry into any preexisting modern-sector activity and excess profits are eliminated. The mechanism through which the latter happens is the upward adjustments in the wage rate w_t as labor is drawn toward the modern sector and modern-sector production expands. Since there are no diminishing returns to labor in the modern sector, we will have an extreme form of industry rationalization in this phase: all but the highest productivity modern-sector activity ceases to exist.

The productivity of the modern sector in the consolidation phase is the maximum from the m_t draws made by entrepreneurs, which will be itself conditional on the policy rule in effect, a_t. Let this maximum productivity be denoted by $\psi^{\max}(m_t)$. Since the ex ante distribution of ψ_i is uniform over $[0, \Psi]$ and the draws are independent, the expected value of the rank statistic $\psi^{\max}(m_t)$ has the simple form $E[\psi^{\max}(m_t)] = \Psi m_t/(1 + m_t)$. Note that $E[\psi^{\max}(m_t)]$ is increasing in m_t but at a decreasing rate. We assume that entrepreneurs (as well as policymakers) are risk neutral.

We close the model by describing production in the traditional sector. The traditional sector operates under constant returns to scale and employs labor and a fixed factor. It will be convenient to use a specific functional form, so we write the production function in the traditional sector as $y_t = (\bar{l} - s_t)^{\alpha}$, where $\bar{l}$ is the total labor force of the economy, s_t is employment in the modern sector, and α is the factor share of labor in the traditional sector. At any given time t, total employment in the modern sector equals the sum of workers employed in new entrepreneurial ventures (during the cost-discovery phase), m_t,[6] and the workers employed in previously established modern-sector firms (during the consolidation phase), e_t. That is, $s_t = m_t + e_t$. The diminishing marginal returns to labor in the traditional sector implies that the modern sector faces a positively sloped labor supply curve. Adjustments in wages will therefore play an important equilibrating role for our economy. The price of the traditional sector is fixed at 1 as the numeraire.

Economic activity extends over infinite discrete time. Every period t, $t > 0$, begins with an inherited policy a_{t-1} and a maximum known productivity in the modern sector $\psi_{t-1}^{\max}$ drawn in the preceding period. Then, on the basis of a_{t-1} and $\psi_{t-1}^{\max}$, the policymaker can make one of the following choices: (a) no new draw (*status quo*); (b) a new draw a_t from the existing policy regime (*policy tinkering*); or (c) a new draw b_t from a new policy regime (*institutional reform*). Like policy draws from the existing regime, draws from a new policy regime are uniformly distributed over the continuum $[0, 2]$. Hence $E(b) = E(a) = 1$.[7] But institutional reform imposes a cost on incumbent modern-sector activities, so that the productivity of the incumbent modern-sector activity following a regime change is $\phi b_t \psi_{t-1}^{\max}$ with $0 < \phi < 1$, whereas that following policy tinkering is equal to $a_t \psi_{t-1}^{\max}$.[8]

Within each time period, the complete sequence of events is as follows:

Stage 1: The government decides whether or not to make a new policy draw a_t or b_t.

Stage 2: Conditional on the policy (either a newly drawn one or the one inherited from the previous period), labor allocations ($e_t \geq 0$) and the new number of entrepreneurs ($m_t \geq 0$) are determined.

Stage 3: Conditional on labor allocations and entrepreneurship decisions, wages (w_t) are determined. If $m_t > 0$, new costs, ψ_i, are revealed. The highest modern-sector productivity attains $\psi_t^{\max}$. The market structure of any young industry that has just emerged is one of monopoly, whereas that of a preexisting industry is characterized by free entry and a competitive market.

We proceed by first defining the equilibrium levels of entrepreneurial activity, labor allocation, and the determination of wages. We then explore the socially optimal patterns of policy experimentation consistent with the market equilibrium.

6.3 Entrepreneurial Activity, Labor Allocation, and Market Equilibrium

We first note that in equilibrium e_t and m_t cannot both be strictly positive. If it pays to operate a preexisting modern-sector activity with the highest known productivity, it will not pay to start new entrepreneurial ventures with the expected level of productivity, and vice versa.

To see this, consider the relationship between the productivity of the incumbent modern-sector activity and the expected productivity of entrepreneurship under both policy tinkering and institutional reform. Under policy tinkering, suppose first that $\psi_{t-1}^{\max} \geq E(\psi) = \Psi/2$. Then, the productivity of the most efficient preexisting modern-sector activity—which will either be in or have gone through its consolidation phase—is higher than the expected productivity of new entrepreneurial ventures, and no one will find it optimal to experiment with new activities. In this case, all individuals prefer to work in either in the traditional sector or the preexisting modern sector (with the consequence that $e_t > 0$ and $m_t = 0$). If on the other hand $\psi_{t-1}^{\max} < E(\psi) = \Psi/2$, the expected entrepreneurial productivity draw exceeds preexisting productivity levels in the modern sector, and, in equilibrium, $e_t = 0$ and $m_t > 0$. A similar argument holds under institutional reform. In particular, if $\phi\psi_{t-1}^{\max} \geq E(\psi) = \Psi/2$, then the productivity of the most efficient preexisting modern-sector activity—despite the fact that it incurs an adjustment cost—is higher than the expected productivity of new entrepreneurial ventures. This leads to $e_t > 0$ and $m_t = 0$. But if $\phi\psi_{t-1}^{\max} < E(\psi) = \Psi/2$, then the productivity of the most efficient modern-sector activity after the reform is below the expected entrepreneurial productivity. Hence, in equilibrium, $e_t = 0$ and $m_t > 0$.

Given the policymaker's choice, the equilibrium wage rates can be derived easily. If policy tinkering is chosen when $\psi_{t-1}^{\max} \geq E(\psi) = \Psi/2$, we have $m_t = 0$, and w_t and e_t are determined by the following two equations:

$$w_t = \alpha(\bar{l} - e_t)^{\alpha-1}, \tag{2}$$

and

$$w_t = pa_t\psi_{t-1}^{\max}. \tag{2'}$$

The second equation ensures zero profits in the modern sector while, taken together, both equations represents labor market equilibrium. If, on the other hand, policy tinkering is done when $\psi_{t-1}^{\max} < E(\psi) = \Psi/2$, then $e_t = 0$, and w_t and m_t are determined by the following equations:

$$w_t = \alpha(\bar{l} - m_t)^{\alpha-1}, \tag{3}$$

and

$$w_t = pa_t\frac{\Psi}{2}. \tag{3'}$$

Equation (3) ensures expected profits are zero for entrepreneurial ventures in an ex ante sense, since expected profits for any individual entrepreneur are given by

$$\pi_t = p - \frac{2w_t}{a_t\Psi}.$$ [9]

If institutional reform is chosen when $\phi\psi_{t-1}^{\max} \geq E(\psi) = \Psi/2$, we have $m_t = 0$, and w_t and e_t are determined by equation (2) and

$$w_t = p\phi b_t \psi_{t-1}^{\max}. \tag{4}$$

If, on the other hand, institutional reform is undertaken when $\phi\psi_{t-1}^{\max} < E(\psi) = \Psi/2$, then $e_t = 0$, and w_t and m_t are determined by equation (3) and

$$w_t = pb_t \frac{\Psi}{2}. \tag{5}$$

As shown, at any point in time, the modern sector will be in either a cost discovery phase or a consolidation phase but not both. The productivity of the incumbent modern-sector activity, together with the policy choice, determines which of these phases the modern sector will be in. For sufficiently low levels of initial modern-sector productivity and prevailing wages, entrepreneurial activity / self-discovery would not be crowded out. Not so when the incumbent modern-sector productivity and wage rates are relatively high (in which case employment in the incumbent modern-sector activity would fully crowd out entrepreneurship).

In what follows, we explore optimal policy choice. In doing so, we focus solely on a second-best world where the policymaker has the same uncertainty about production costs as private entrepreneurs do.[10]

6.4 Optimal Policy Choice

At the beginning of each period $t > 0$, the policymaker observes the maximum productivity draw of the previous period $\psi_{t-1}^{\max}$ and, depending on the inherited policy draw a_{t-1}, decides whether or not to make a new policy draw a_t or b_t.

What is the policymaker's optimal decision? The easiest case to consider is the one where the inherited policy draw, a_{t-1}, exceeds its expected value, $E(a) = 1$. In that case, the policymaker would choose to maintain the status quo under all circumstances we examine, since there is nothing to be gained in expected value terms by making a renewed draw. Under the status quo, free entry reigns, all but the highest productivity firms close down, and imitation—together with the wage adjustment mechanism that accompanies it—drives profits from that activity down to zero.

Policy tinkering does not affect the margin between old and new activities, and hence does not influence the equilibrium level of cost discovery. But institutional reform can generate cost discovery where policy tinkering would fail to do so since the former reduces productivity (and wages) in preexisting activities. As a consequence, depending on the productivity of the incumbent modern-sector activity, $\psi_{t-1}^{\max}$, and whether the policymaker chooses to tinker, a_t, or reform, b_t, there are three other cases to consider: (a) $\psi_{t-1}^{\max} \geq \Psi/2\phi$ so that wages are too high to generate new entrepreneurial ventures even after major reforms are instituted; (b) $\psi_{t-1}^{\max} < \Psi/2$, which implies that wages are low enough that tinkering with existing policies is sufficient to entice new entrepreneurs; and (c) $\Psi/2\phi > \psi_{t-1}^{\max} \geq \Psi/2$, so that wages are too high to yield new entrepreneurship under policy tinkering but are low enough to entice entrepreneurs with institutional reforms.

We now turn to an examination of each of these cases.

(a) $\psi_{t-1}^{\max} \geq \Psi/2\phi$

In this region, wages are too high to warrant new entrepreneurial experimentation. Thus, labor is allocated between the traditional sector and the incumbent modern-sector activity only. That is, $m_t = 0$ and $s_t = e_t > 0$.

The equilibrium wage rate equates the marginal product of labor in the incumbent modern-sector activity to that of labor in the traditional sector, as indicated by equations (2) and (2′):

$$w_t = pa_{t-1}\psi_{t-1}^{\max} = \alpha(\bar{l} - e_t)^{\alpha-1}. \tag{6}$$

Using (6), we can solve for the level of employment in period t:

$$e_t = \bar{l} - \left(\frac{\alpha}{pa_{t-1}\psi_{t-1}^{\max}}\right)^{1/(1-\alpha)}. \tag{7}$$

Thus, with a policy a_{t-1} in place, the aggregate output of the economy will be given by $Y_t \equiv y_t + px_t$, where

$$y_t = (\bar{l} - e_t)^{\alpha} = \left(\frac{\alpha}{pa_{t-1}\psi_{t-1}^{\max}}\right)^{\alpha/(1-\alpha)}, \tag{8}$$

and

$$x_t = e_t a_{t-1}\psi_{t-1}^{\max} = a_{t-1}\psi_{t-1}^{\max}\left[\bar{l} - \left(\frac{\alpha}{pa_{t-1}\psi_{t-1}^{\max}}\right)^{1/(1-\alpha)}\right]. \tag{9}$$

Now consider the outcome when the policymaker decides to tinker and make a new draw a_t in period t. With the new policy in effect, the productivity of the incumbent activity would equal $a_t\psi_{t-1}^{\max}$. And as implied by equation (7), this would lead to a change in the level of employment in the incumbent modern-sector activity.[11]

Let $E(Y_t|a) \equiv E(y_t|a) + pE(x_t|a)$ denote the expected level of aggregate output associated with tinkering (i.e., a new policy draw a_t). Given that $E(a) = 1$, we establish the following:

$$E(y_t|a) = (\bar{l} - e_t)^{\alpha} = \left(\frac{\alpha}{p\psi_{t-1}^{\max}}\right)^{\alpha/(1-\alpha)}, \tag{10}$$

$$E(x_t|a) = e_t\psi_{t-1}^{\max} = \psi_{t-1}^{\max}\left[\bar{l} - \left(\frac{\alpha}{p\psi_{t-1}^{\max}}\right)^{1/(1-\alpha)}\right]. \tag{11}$$

Next consider the case where the policymaker decides in favor of institutional reform and makes a policy draw b_t. The equilibrium wage rate equates the marginal product of labor in the incumbent modern-sector activity to that of labor in the traditional sector:

$$w_t = p\phi b_t\psi_{t-1}^{\max} = \alpha(\bar{l} - e_t)^{\alpha-1}, \tag{6$'$}$$

Using (6′), we can solve for the level of employment in period t:

$$e_t = \bar{l} - \left(\frac{\alpha}{p\phi b_t\psi_{t-1}^{\max}}\right)^{1/(1-\alpha)} \tag{7$'$}$$

Let $E(Y_t|b)$, $E(Y_t|b) \equiv E(y_t|b) + pE(x_t|b)$ denote the expected level of aggregate output conditional on the policy regime change. Since $E(b) = 1$, we can establish the following:

$$E(y_t|b) = (\bar{l} - e_t)^{\alpha} = \left(\frac{\alpha}{p\phi\psi_{t-1}^{\max}}\right)^{\alpha/(1-\alpha)}, \tag{10$'$}$$

$$E(x_t|b) = e_t\phi\psi_{t-1}^{\max} = \phi\psi_{t-1}^{\max}\left[\bar{l} - \left(\frac{\alpha}{p\phi\psi_{t-1}^{\max}}\right)^{1/(1-\alpha)}\right]. \tag{11$'$}$$

Note that if it were not optimal for the policymaker to change policy in period t, it would also not be optimal to do so in any subsequent period, since the economic environment is assumed to remain unchanged. This suggests that the present discounted welfare associated with the status quo is given by $Y_t/(1-\beta)$ where β, $0 < \beta < 1$, denotes the time discount factor.

If instead the policymaker were to tinker (and make a new policy draw a_t) in period t, the outcome would be stochastic. From this period's vantage point, the expected value of the outcome in any subsequent period would be $E(Y_t|a)$, regardless of whether the policymaker makes additional draws down the line. This is due to the fact that, evaluated at $a_t = E(a) = 1$, $Y_{t+1}(a = 1)$ would equal $E(Y_t|a)$. Thus, the present discounted welfare associated with a policy change is equal to $E(Y_t|a)/(1-\beta)$. Based on the same argument, the present discounted welfare associated with institutional reform is equal to $E(Y_t|b)/(1-\beta)$.

An examination of equations (8)–(11), (10′), and (11′) reveals that

$$\frac{E(Y_t|a)}{1-\beta} > \frac{Y_t}{1-\beta} \qquad \text{and} \qquad \frac{E(Y_t|a)}{1-\beta} > \frac{E(Yt|b)}{1-\beta}. \tag{12}$$

Hence, when $a_{t-1} < 1$ and $\psi_{t-1}^{\max} \geq \Psi/2\phi$, we find that the policymaker would—instead of pursuing major reforms—just tinker with existing policies. This is due to the fact that institutional reforms are costly, and without new entrepreneurial experimentation they provide no additional benefit over policy tinkering.

(b) $\psi_{t-1}^{\max} < \Psi/2$

In this case, equilibrium wages are low enough that there is new entrepreneurial experimentation and employment in the incumbent modern-sector activity is driven to zero. Thus, labor is allocated between the traditional sector and new entrepreneurship only. That is, $e_t = 0$ and $s_t = m_t > 0$.

The equilibrium wage rate equates the expected marginal product of new entrepreneurial ventures to that of labor in the traditional sector, as in equations (3) and (3′):

$$w_t = \frac{pa_{t-1}\Psi}{2} = \alpha(\bar{l} - m_t)^{\alpha-1}, \tag{13}$$

Using (13), we can solve for the equilibrium level of expected entrepreneurial ventures:

$$m_t = \bar{l} - \left(\frac{2\alpha}{pa_{t-1}\Psi}\right)^{1/(1-\alpha)}. \tag{14}$$

With no change in policy, the aggregate output of the economy would equal $E(Y_t) \equiv y_t + pE(x_t)$, where

$$y_t = (\bar{l} - m_t)^{\alpha} = \left(\frac{2\alpha}{pa_{t-1}\Psi}\right)^{\alpha/(1-\alpha)}, \tag{15}$$

$$E(x_t) = m_t a_{t-1}\frac{\Psi}{2}. \tag{16}$$

This is a case in which there is no uncertainty with respect to the output of the traditional sector because no new policy draw is made and the number of new entrepreneurial ventures is observable ex ante. In contrast, there is uncertainty about the output of the highest modern-sector activity because, while the expected value of the economy-wide outcome of entrepreneurial ventures equals $E[\psi^{\max}(m_t)] = \Psi m_t/(1+m_t)$, it actual value is not observable ex ante.

At time $t+1$ free entry reigns, eliminating all but the highest productivity modern-sector activity. Thus, the expected level of aggregate output in all future periods, $E(Y_{t+1}) \equiv E(y_{t+1}) + pE(x_{t+1})$, equals

$$E(y_{t+1}) = (\bar{l} - e_{t+1})^{\alpha} = \left(\frac{\alpha}{pa_{t-1}\Psi}\frac{[1+m_t]}{m_t}\right)^{\alpha/(1-\alpha)}, \tag{17}$$

$$E(x_{t+1}) = e_{t+1}a_{t-1}E\{\psi^{\max}[m(a)]\} = \frac{a_{t-1}m_t\Psi}{1+m_t}\left[\bar{l} - \left(\frac{\alpha}{pa_{t-1}\Psi}\frac{[1+m_t]}{m_t}\right)^{1/(1-\alpha)}\right]. \tag{18}$$

Now consider the outcomes when the policymaker decides to tinker and make a new policy draw a_t. Since the expected value of the draw a_t equals $E(a) = 1$, both the equilibrium wage rate and the number of entrepreneurs would exceed those given by (13) and (14), respectively. With $E(Y_t|a) \equiv E(y_t|a) + pE(x_t|a)$ denoting the level of aggregate output associated with tinkering and the expected new policy draw, $a_t = E(a) = 1$, we can establish the following:

$$E(y_t|a) = (\bar{l} - m_t)^{\alpha} = \left(\frac{2\alpha}{p\Psi}\right)^{\alpha/(1-\alpha)}, \tag{19}$$

$$E(x_t|a) = m_t\frac{\Psi}{2}. \tag{20}$$

At time $t+1$ free entry eliminates all except the highest productivity modern-sector activity which ex ante attains $E\{a\psi^{\max}[m(a)]\} = \Psi m_t/(1+m_t)$. Thus, the expected level of aggregate output in all future periods, $Y_{t+1}(a=1) \equiv y_{t+1}(a=1) + px_{t+1}(a=1)$, equals

$$y_{t+1}(a=1) = (\bar{l} - e_{t+1})^{\alpha} = \left(\frac{\alpha}{p\Psi}\frac{[1+m_t]}{m_t}\right)^{\alpha/(1-\alpha)}, \tag{21}$$

$$x_{t+1}(a=1) = e_tE\{\psi^{\max}[m(a)]\} = \frac{m_t\Psi}{1+m_t}\left[\bar{l} - \left(\frac{\alpha}{p\Psi}\frac{[1+m_t]}{m_t}\right)^{1/(1-\alpha)}\right]. \tag{22}$$

Instead, if the policymaker opts out for institutional reform and makes a new policy draw b_t, the equilibrium wage rate is determined by the following equation:

$$w_t = \frac{pb_t\Psi}{2} = \alpha(\bar{l} - m_t)^{\alpha-1}. \tag{13'}$$

With (13′), we can derive the equilibrium level of entrepreneurship:

$$m_t = \bar{l} - \left(\frac{2\alpha}{pb_t\Psi}\right)^{1/(1-\alpha)}. \tag{14'}$$

The output of the economy will be given by $E(Y_t|b)$, $E(Y_t|b) \equiv E(y_t|b) + pE(x_t|b)$, where the components $E(y_t|b)$ and $E(x_t|b)$ are identical to equations (19) and (20), respectively. The expected output of the economy in all subsequent periods will equal $Y_{t+1}(b=1)$, $Y_{t+1}(b=1) \equiv y_{t+1}(b=1) + px_{t+1}(b=1)$, where the output of the traditional and the modern sectors are given by (21) and (22), respectively.

Based on equations (15)–(22), we can now state the following:

$$E(Y_t|a) + \frac{\beta[Y_{t+1}(a=1)]}{1-\beta} = E(Y_t|b) + \frac{\beta[Y_{t+1}(b=1)]}{1-\beta} > E(Y_t) + \frac{\beta[E(Y_{t+1})]}{1-\beta}. \tag{23}$$

In (23), the terms $\beta[Y_{t+1}(a=1)]/(1-\beta)$ and $\beta[Y_{t+1}(b=1)]/(1-\beta)$ are equal to one another. This is because, subsequent to the initial period in which monopoly rents accrue, the expected aggregate output of the economy would be equal under the two policy-setting regimes. The terms $E(Y_t|a)$ and $E(Y_t|b)$ are also equal, because policy draws from either regime generate the same amount of entrepreneurial experimentation.[12] As a result, we establish that the policymaker would just tinker with existing policies if $a_{t-1} < 1$ and $\psi_{t-1}^{\max} < \Psi/2$.

(c) $\Psi/2 \leq \psi_{t-1}^{\max} < \Psi/2\phi$

In this case, wages are low enough to warrant entrepreneurial experimentation under a reform, but they are not sufficiently low to generate it with policy tinkering. Thus, in order to determine the appropriate course of action, the policymaker would need to compare the expected aggregate output associated with institutional reform with the expected aggregate output associated with policy tinkering, both of which we discussed earlier.

In this case, for sufficiently high values of β, the following inequality would hold:

$$E(Y_t|b) + \frac{\beta[Y_{t+1}(b=1)]}{1-\beta} > \frac{E(Y_t|a)}{1-\beta} > \frac{Y_t}{1-\beta} \tag{24}$$

As we show in the appendix, we find that $\forall a_{t-1} < 1$ and $\Psi/2 \leq \psi_{t-1}^{\max} < \Psi/2\phi$ reform dominates tinkering for a sufficiently forward-looking policymaker who has a relatively high β. However, we cannot rule out the possibility of status quo for $\psi_{t-1}^{\max}$ in the neighborhood of $\Psi/2\phi$ for a shortsighted policymaker. The reason that the discount rate matters is this: under the status quo as well as tinkering, there are gains in the current period from consolidation in the modern sector as resources move from less profitable activities to the highest-productivity incumbent activity. Institutional reform generates (expected) gains in the future from higher cost discovery, but does so at the cost of giving up these current gains.

In sum, our results have the following implications. Policy tinkering dominates institutional reform when existing policies leave something to be desired and the modern sector is pretty unproductive (i.e., for $a_{t-1} < 1$ and $\psi_{t-1}^{\max} < \Psi/2$). In this case, the prevailing wage rate is low enough to entice new cost discovery even in the absence of institutional reform. Hence, given the adjustment costs involved and the possible loss of gains that arise during the consolidation phase in the modern sector, it would not be desirable to alter the economy's institutional arrangements. Similarly, policy tinkering dominates institutional reform when existing policies are not terribly desirable but the modern sector is quite productive (i.e., for $a_{t-1} < 1$ and $\psi_{t-1}^{\max} \geq \Psi/2\phi$). In this case, wages are high enough to stifle cost discovery even when institutional reform is attempted. Thus, given the adjustment costs involved with institutional reform and the absence of cost discovery gains, it is desirable to tinker with policies within the existing institutional framework. In contrast, provided that a policymaker is sufficiently forward-looking, institutional reform dominates policy tinkering when existing policies are undesirable and the modern sector is only moderately productive (i.e., for $a_{t-1} < 1$ and $\Psi/2 \leq \psi_{t-1}^{\max} < \Psi/2\phi$). In this case, the prevailing wage rate is low enough to entice new entrepreneurial experimentation under institutional reform, but too high to do so under policy tinkering. Hence, it is socially desirable to bear the adjustment costs and explore alternative institutional arrangements. These results are summarized in table 6.1.

As the table shows, the expected impact on welfare (and economic performance) of institutional reform varies with the quality of

Table 6.1
Summary of main implications

	Quality of preexisting policies				
	Lousy ($a_{t-1} < 1$)			Good ($a_{t-1} \geq 1$)	
	Low productivity $\psi_{t-1}^{max} < \Psi/2$	Intermediate productivity $\Psi/2 \leq \psi_{t-1}^{max} < \Psi/2\phi$	High productivity $\psi_{t-1}^{max} \geq \Psi/2\phi$	Low productivity $\psi_{t-1}^{max} < \Psi/2$	High productivity $\psi_{t-1}^{max} \geq \Psi/2$
Optimal policy	tinker	inst. reform	tinker	status quo	status quo
Cost discovery under optimal policy?	yes	yes	no	yes	no
Expected impact on welfare of					
tinkering	+++++	+++	++	−	
inst. reform	+++++	++++	+/−	−−	
Policy ranking	tinker inst. ref. s.q.	inst. ref. tinker s.q.	tinker s.q. inst. ref.	status quo tinker inst. ref.	

preexisting policies and the initial productivity of the modern sector. In our model, initial productivity is in turn determined by the inherited level of entrepreneurial experimentation. Note that even when it is not the dominant strategy, institutional reform can improve welfare in economies where the productivity of the modern sector is not too high. The same cannot be said with respect to economies where the modern sector is relatively productive; in those economies, policy tinkering would enhance welfare but institutional reform would undermine it. We test this idea in our empirical work in section 6.6.

6.5 Institutional Reforms with Large Productivity Impact

We have assumed so far that the expected productivity impact of institutional reform is no greater than that of policy tinkering (i.e., $E(b) = E(a) = 1$). We finally consider the possibility that $E(b) > E(a)$. This corresponds to a case where the policy regime can be unambiguously improved because existing institutional arrangements are exceedingly weak. Suppose, therefore, that $E(b) = \gamma > 1$. The expected productivity of an incumbent modern sector activity after a policy regime change now equals $\gamma\phi\psi_{t-1}^{\max}$. Thus, whether the cost of adjustment to a new policy regime change is high enough to offset the expected gain of a reform will be crucial. If the expected productivity impact of institutional reform were fairly large so that $\gamma\phi \geq 1$, then reform would not be costly on net to incumbent modern-sector activities. The government would then want to undertake institutional reform as long as $a_{t-1} < E(b) = \gamma$. If, on the other hand, the expected productivity impact of a reform were only moderately large so that $\gamma\phi < 1$, incumbent modern-sector activities would still suffer an expected loss—albeit a smaller one than that in the previous section—as a result of institutional reform.

What this suggests is that when there is no new entrepreneurial experimentation (as in section 6.4.1 where $\psi_{t-1}^{\max} \geq \Psi/2\phi$), institutional reforms with relatively small expected productivity gains (i.e., γ closer to 1) will still be dominated by policy tinkering. However, as long as wages are low enough to allow new cost discovery, institutional reform will now unambiguously dominate tinkering since incumbents are displaced by new entrepreneurial ventures anyway. In this latter case, the adjustment costs that incumbents would have incurred had they remained in business become irrelevant.

6.6 Empirical Evidence

The model we discussed earlier yields a rich range of empirical implications. However, testing these implications directly is rendered difficult by the absence of internationally comparable measures of entrepreneurship, which plays a key mediating role in our framework. The ILO provides some patchy cross-national data on self-employment.[13] In the absence of better proxies, we used this data to construct an index of entrepreneurial intensity (*ENTRAT*), which we compute by taking the ratio of self-employed individuals to total nonagricultural employment. This ratio can be calculated for more than fifty countries around the year 1990, and it varies from a low of 5 percent in Sweden to a high of 58 percent in Nigeria. We recognize that *ENTRAT* varies systematically with levels of development, so we control for per capita GDP in all our regressions to guard against spurious results. See table 6.2 for summary statistics and the correlation matrix for the variables used in what follows.

We use *ENTRAT* to test three of the implications of our model. First, our model implies that entrepreneurial experimentation is inversely related to the prevailing level of modern-sector labor costs. Second, economies with higher levels of entrepreneurship should generate more productive modern-sector activities, and therefore should experience higher rates of economic growth. Third, institutional reforms should stimulate economic activity the most in countries where prior levels of entrepreneurship have been too low to generate much high-productivity activity (or more precisely, the economic impact of reforms should monotonically decline in prior levels of entrepreneurship; see table 6.1).

Columns (1)–(3) of table 6.3 present our results on the first implication. Our measure of modern-sector labor costs is average unit labor costs in manufacturing (*ln ULC*), which we calculate by taking the ratio of wages to manufacturing value added per employee (both from the ILO). Since *ENTRAT* is measured around 1990, we compute *ln ULC* as an average for 1985–1989. As column (1) shows, *ln ULC* exerts a negative and statistically significant influence on *ENTRAT*, even after controlling for per capita GDP. Since labor costs are also related to the price level (see Rodrik 1999), we control for cross-country differences in the price level for consumption (*ln PC*) in column (2). In column (3), we add three regional dummies as additional controls. The estimated coefficient of *ln ULC* remains negative and statistically significant with

Table 6.2
Descriptive statistics

	Correlation matrix											
	Mean	S.D.	GROWTH	ENTRAT	ln ULC	ln PC	ln GDPCAP	FERT	SECM	REF.	ΔGRW	GOVT
GROWTH	.0151	.0209	1.00									
ENTRAT	.188	.109	−.278	1.00								
ln ULC	−1.126	.4375	−.042	−.680	1.00							
ln PC	−.675	.381	−.100	−.612	.513	1.00						
ln GDPCAP	8.38	1.13	.208	−.838	.607	.747	1.00					
FERT	1.53	2.07	−.260	.691	−.509	−.609	−.915	1.00				
SECM	4.98	8.03	.158	−.502	.399	.609	.604	−.577	1.00			
REFORM	.355	.481	−.420	−.175	.510	−.330	−.476	.475	−.391	1.00		
ΔGRW	−.0084	.0277	.372	.008	−.116	−.079	.012	−.158	.139	−.175	1.00	
GOVT	7.35	9.26	−.048	.126	.045	−.258	−.455	.486	−.155	.037	.116	1.00

Table 6.3
Main results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: ENTRAT			Dependent variable: GROWTH			Dependent variable: ΔGROWTH		
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	IV
ln ULC	−.0481*	−.0514*	−.0412*						
	(.0137)	(.0143)	(.015)						
ln GDPCAP	−.0842*	−.0884*	−.0635*	.0124*	.021*	.0181*	−.0112	−.0108	−.0196**
	(.0067)	(.0092)	(0.0087)	(.0027)	(.0079)	(.0077)	(.0077)	(.0069)	(.0086)
ln PC		.0165	−.0362***						
		(.0207)	(.0193)						
ENTRAT							−.0847	−.0143	−.1197**
							(.0579)	(.0552)	(.064)
ENTRAThat				.1098*	.2284*	.2151*			
				(.0359)	(.0916)	(.0892)			
LAAM			.0279**		−.0155*	−.0119**	−.0039	−.0033	−.0019
			(.013)		(.0061)	(.0060)	(.0100)	(.0093)	(.008)
SAFRICA			−.1032*		.0096	.0118	−.0239***	−.0646*	−.0566*
			(.0154)		(.0124)	(.0119)	(.0133)	(.0122)	(.012)
ASIA			−.0030		.0090	.0137**	−.0162	−.0165***	−.0191*
			(.0153)		(.0071)	(.007)	(.0105)	(.0093)	(.005)
SECM						−.0001			
						(.0003)			
FERT						−.0008			
						(.0018)			

GOVT						.0006 (.0004)			
REFORM							−.0089 (.009)	.0411* (.016)	.0351** (.018)
REFORM ∗ ENTRAT								−.2477* (.0719)	−.1912** (.086)
Observations	53	52	52	82	82	81	53	53	50

Note: Robust regression estimates, except for col. (9), which shows IV-GMM estimates. GROWTH is per capita GDP growth from 1990 to 2000. ΔGROWTH is the difference between growth rate in the 1990s and growth rate in the 1970s. See text for more details. Standard errors in parentheses. * significant at 1 percent; ** significant at 5 percent, *** significant at 10 percent.

both robustness checks. In addition, the fit of our most parsimonious specification, which includes unit labor costs and per capita income only, is remarkably high; its adjusted R-squared is 0.75.[14]

One possible concern with the specifications in columns (1)–(3) is reverse causation. Perhaps we are getting the effect of entrepreneurial intensity on labor costs, rather than vice versa. But theoretically this reverse relationship is positively signed rather than negatively signed. So if there is simultaneous-equation bias at play, it should work against us (i.e., the bias is in the direction of making the estimated coefficient less negative). Another potential concern involves the fact that labor costs might be a reflection of labor market inflexibility and the costs of formality (generating a source of omitted variable bias). Here too, the effect would have been a positive relationship—the more institutionalized the labor market, the greater the escape into self-employment—rather than the negative one that we find. In addition, Friedman et al. (2000) present cross-country data on the shares of the informal sectors in economic activity—all of which are from the late 1980s or early 1990s. Using these estimates as additional explanatory variables, we reran the specifications in columns (1)–(3). Accounting for the shares of the informal sector not only did not influence our main results but also yielded insignificant coefficients on the shares of the informal sector.[15]

We next turn to the relationship between entrepreneurship and subsequent growth. In our model, the higher the inherited level of entrepreneurship m_{t-1}, the higher is the productivity reached in the modern sector $\psi_{t-1}^{\max}$. We proxy $\psi_{t-1}^{\max}$ with economic growth. In order to enlarge our sample size (which was limited to fifty-three countries in the previous set of regressions), for this exercise we first use regression (3) to generate a predicted value for *ENTRAT* for more than eighty countries (*ENTRAThat*). Regressions (4)–(6) show that *ENTRAThat* is robustly correlated with subsequent growth during the 1990s. The first of these regressions (column (4)) is a bare-bones specification, to which we next add regional dummies (column (5)) and a number of standard growth determinants (fertility, male schooling, and government consumption, column (6)). Hence, the intensity of entrepreneurship—as predicted by labor costs, among other things—has a positive and significant impact on the subsequent rates of growth.[16]

Our third and most ambitious set of tests relates to the interaction between institutional reform and the preexisting level of entrepreneurship in determining economic outcomes. In our model, a high level of entrepreneurship m_{t-1} raises the (expected) productivity in the modern sector $\psi_{t-1}^{\max}$ but also lowers the return to institutional reform (see table 6.1). We now test this last implication.

To code our institutional reform variable, we rely on Wacziarg and Welch (2003), who have recently revised and updated the Sachs-Warner data set on the timing of major reforms. The original Sachs and Warner (1995) effort was aimed at identifying countries that had opened up their economies to trade and the timing of these reforms. However, their definition of trade reform is so broad and demanding (requiring adjustments in trade policies, macroeconomic policies, and structural policies) that it is quite well suited for our purposes (see Rodriguez and Rodrik 2000 for a discussion). Hence, in order to be classified as "open," a country needs to have not only suitably low levels of trade barriers, but also no major macroeconomic disequilibria (measured by the black-market premium for foreign currency), no socialist economic system, and no export marketing board. Since *ENTRAT* is measured around 1990, we code our *REFORM* variable as a dummy variable that takes the value of 1 if the country has undergone a Sachs-Warner-Wacziarg-Welch reform between 1985 and 1994 (inclusive). Our dependent variable is the change in growth between the 1990s and 1970s, Δ*GROWTH*. (We exclude the 1980s because of the pervasive effects of the Latin American debt crisis during that decade.) Our model implies that *REFORM* should have different effects on Δ*GROWTH* depending on the value of *ENTRAT*.

In column (7) we regress Δ*GROWTH* on *REFORM*, *ENTRAT*, per capita GDP, and a set of regional dummies. Note that the estimated coefficient on *REFORM* is negative (with a t-statistic around 1). This result reflects the disappointing outcome with institutional reforms in the 1990s, as discussed in the introduction (see also Rodrik 2003). In the next column (8), we interact *REFORM* with *ENTRAT*. The results are quite striking. The estimated coefficient on the interaction term is negative and highly significant. In addition, once the interaction term is included, the coefficient on *REFORM* turns positive and becomes statistically significant. So the impact of institutional reform turns out to be dependent on the level of our proxy for entrepreneurship. Those that benefited were the countries with very low levels of entrepreneurial intensity. In column (9) we repeat the exercise in an instrumental variables framework, to alleviate concern about the possible endogeneity of *ENTRAT*. We use as our instruments the determinants of *ENTRAT* used in column (3) and their interaction with *REFORM*.

The results are equally strong. Institutional reform had dramatically different effects depending on the preexisting levels of entrepreneurial intensity.

According to our results, institutional reform enhanced growth in countries where prevailing entrepreneurial intensity fell short of a certain cutoff level, and reduced growth elsewhere. Using the estimates of column (8), we find this cutoff value of *ENTRAT* to be 0.17 $(= .0411/.2477)$. This corresponds to the median value in our sample and is about the level observed in Malaysia. The countries in our sample that undertook institutional reform and where *ENTRAT* was below this level are South Africa, Tunisia, Trinidad and Tobago, Israel, and New Zealand. The average for Latin American countries is substantially above this cutoff at 0.27. Interestingly, India and China, two important cases of gradualist tinkering, were likely above this cutoff as well. While we do not have a value for *ENTRAT* for either of these countries, India's *ENTRAThat* is 0.30, and China (for which we cannot compute *ENTRAThat* due to missing labor cost data) would have had to have labor costs that are implausibly high (two orders of magnitude higher than India's) to fall below the 0.17 threshold. Hence this evidence suggests that both countries were better off not having undertaken deep institutional reform à la Latin America.

Encouraging as they are for our model, these results are obviously contingent on the reliability of our proxy for entrepreneurial intensity *ENTRAT* and are sensitive to the coding of *REFORM*. We end this section by providing a somewhat different type of evidence that does not rely on either of these variables. We simply focus on the experience in Latin America, where we know significant amounts of structural reform took place in the late 1980s and early 1990s. As an alternative to *ENTRAT*, we use productivity growth in the 1970–1980 period as a proxy for the strength of the cost discovery process and the vibrancy of entrepreneurship. (We ignore the 1980s once again, due to the special circumstances related to the debt crisis.) As before, we take the difference in the growth rates of GDP per capita between the 1990s and 1970s as a measure of the impact of institutional reform.

Figure 6.1 shows that there is a strong negative correlation in Latin America between TFP performance during the 1970s and $\Delta GROWTH$ $(\rho = -0.72)$. Countries that were experiencing rapid TFP growth in the 1970s (e.g., Brazil) reaped little gain in the 1990s, while those that had poor TFP growth performance (e.g., Chile) improved their performance. In line with the implications of our framework, the payoffs to

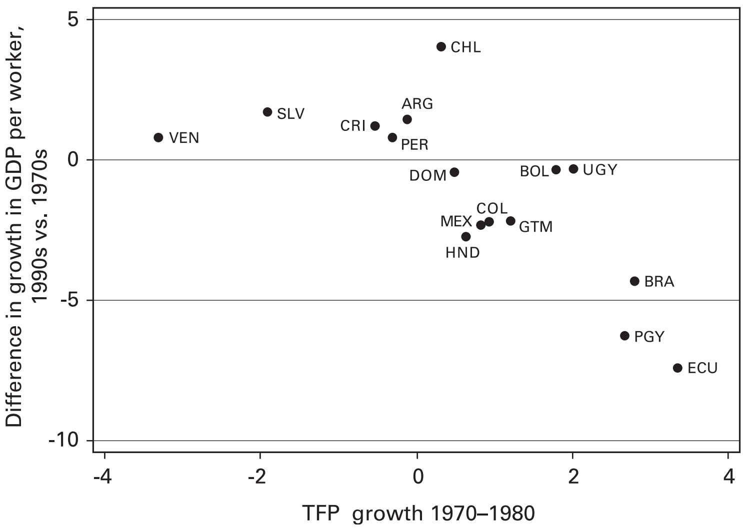


Figure 6.1
Relationship between prior productivity growth and impact of institutional reform in Latin America.
Source: Data on TFP and GDP per worker from Bosworth and Collins 2003.

institutional reform were greatest when it was likely to induce a new wave of entrepreneurship, namely, when the cost discovery process had run out of steam. And they were lowest when productivity performance was already satisfactory.

We end by emphasizing that these empirical results are suggestive rather than definitive. Our proxies for entrepreneurship are crude, and cross-national regressions are subject to a wide variety of well-known problems (see Durlauf et al. 2004 for a recent critical review). In the final analysis, there is no substitute for detailed empirical work focusing directly on the mechanisms that our theoretical work has identified as being important.

6.7 Concluding Remarks

We argued in this chapter that the interplay between policy choices and entrepreneurial incentives provides an important key to understanding recent patterns of economic performance around the world. The taxonomy we offer yields a rich set of normative and positive implications.

On the normative side, we find that optimal policy choice is highly contingent on initial conditions. When the quality of preexisting policies is high, status quo is the dominant policy choice regardless of the productivity level in the modern sector. But when the quality of preexisting policies leaves something to be desired, the optimal choice between policy tinkering and institutional reform depends critically on the level of productivity reached in the modern sector. And the relationship is not linear. Policy tinkering is the best choice when the modern sector is either (a) unproductive *or* (b) highly productive, while institutional reform is the best choice when (c) the productivity level is intermediate between these two. The reason is that only in case (c) does institutional reform have a clear advantage over tinkering: that is the case where institutional reform induces cost discovery while tinkering fails to do so. In case (a) tinkering is enough to generate cost discovery, while in case (b) neither tinkering nor institutional reform is able to do so.

Perhaps our most striking conclusion is a positive one: institutional reforms boost economic activity in countries where entrepreneurial activity is languishing, and they fail in places where entrepreneurial attempts at cost discovery are relatively vibrant. The available empirical evidence supports such a conditional relationship. Hence, recognizing the interplay between reforms and entrepreneurship may help resolve the puzzle of why institutional reforms have worked in a handful of countries while failing in others.

Our framework provides additional subtle insights on reform strategies and new ways to interpret recent experience with economic development. Consider, for instance, our results on policy tinkering. We find that policy tinkering works best when existing policies are demonstrably poor *and* the productivity of modern sector activities is extremely low. This seems to characterize the experience of some of the growth superstars of the last two decades fairly well. In particular, China (since 1978), India (since 1980), and Vietnam (since 1986) have scored spectacular economic gains with changes in institutional arrangements that fall far short of what most Western economists would have considered a prerequisite for success. In India, the changes in policy during the 1980s were barely perceptible. And even the more ambitious reforms of the 1990s are better described as gradualist tinkering than as deep institutional reform. China and Vietnam made considerable strides toward building a market economy while keeping the basic socialist institutional arrangements (including state ownership of key industries) intact.[17] All three countries started from a very low level, not just in terms of the market-friendliness of their policies, but also in terms of the productivity of their economies. Policy tinkering has a potentially very high return under these circumstances, as our model shows. But as the model also indicates, not all tinkerers will succeed; what matters is the actual policy draw.

Our model also provides a reason why Chinese-style gradualism may not have worked in the former socialist countries of Eastern Europe, and therefore rationalizes the deeper institutional reform and "shock therapy" that countries such as Poland and the Czech Republic undertook. Unlike China and Vietnam, Eastern European countries had built modern manufacturing sectors and were already high-wage economies. Tinkering would likely not have been enough to generate new entrepreneurship and structural change. The fact that economic performance in the former Soviet Union and Eastern Europe has turned out quite uneven is, of course once, again consistent with one of our central building blocks—the uncertainty with regard to policy outcomes.

We close by reiterating the central normative messages of this chapter. Productive transformation and policy reform are both subject to a great deal of uncertainty. Entrepreneurship depends both on good policy and on adequate rents. Policy tinkering and institutional reform both have their respective advantages. Appropriate strategies depend on initial conditions—namely, the quality of policies, the level of productivity in nontraditional activities, and the state of entrepreneurship. Reformers who internalize these lessons are likely to make good choices while those who don't are likely to be disappointed.

Appendix

Claim: $\forall a_{t-1} \in [0,1)$ and $\Psi/2 \leq \psi_{t-1}^{\max} < \Psi/2\phi$, $\exists \beta \in (0,1)$ such that

$$E(Y_t|b) + \frac{\beta[Y_{t+1}(b=1)]}{1-\beta} > \frac{E(Y_t|a)}{1-\beta}. \tag{A.1}$$

Proof: Equation (A.1) can be rewritten as

$$E(Y_t|b) - E(Y_t|a) + \frac{\beta[Y_{t+1}(b=1) - E(Y_t|a)]}{1-\beta} > 0.$$

$\forall a_{t-1} \in [0,1)$ and $\Psi/2 \leq \psi_{t-1}^{\max} < \Psi/2\phi$, institutional reform leads to new entrepreneurial experimentation, but policy tinkering does not.

Hence, the difference $Y_{t+1}(b=1) - E(Y_t|a)$ is strictly positive. The reason is that the expected policy draw under a reform equals one and the policy draw that helped determine the productivity of the incumbent modern-sector activity, a_{t-1}, is strictly less than one. In contrast, the difference $E(Y_t|b) - E(Y_t|a)$ is strictly negative in the limit when $\psi_{t-1}^{\max}$ is approaching $\Psi/2$ as well as when it is approaching $\Psi/2\phi$ (both of which indicate that the components (10) and (11) add up to more than the components (19) and (20)). Then, we can establish that the difference $E(Y_t|b) - E(Y_t|a)$ is negative due to the strict monotonicity of $E(Y_t|b) - E(Y_t|a)$ in $\psi_{t-1}^{\max}$. Given that the difference $Y_{t+1}(b=1) - E(Y_t|a)$ is strictly positive, $\exists \beta \in (0,1)$ such that (A.1) holds. ▪

Notes

Dani Rodrik thanks the Carnegie Corporation of New York for financial support. For useful comments and suggestions, we thank Lee Alston, Mushfiq Mobarak, and David N. Weil. The usual disclaimer applies.

1. We take a broad view of these costs. What we have in mind are not just the standard adjustment costs, but the loss incurred in the value of organizational capital accumulated under previous institutional arrangements. This includes, for example, the disruption in the relation-specific investments made by incumbents with their suppliers, with their customers, and with the government. A change in the rules of the game necessitates that these investments be reconstituted, and therefore imposes search and other transaction costs. Roland and Verdier (1999) explore these transition costs in the context of former socialist economies, arguing that their absence is one advantage of the more gradualist paths followed by China (see also Blanchard and Kremer 1997).

2. Our approach here has parallels with the work of Caballero and Hammour (2000), who emphasize the costs of institutional sclerosis and inadequate levels of creative destruction.

3. Hence uncertainty is associated with costs of production rather than with entrepreneurial talents. Once an entrepreneur discovers costs in a given sector, there is a large number of entrepreneurs who can emulate the incumbent. Some other models of industrialization emphasize instead the selection of talented entrepreneurs who can best undertake the innovations needed for modern production. See, for example, Acemoglu, Aghion, and Zilibotti 2002.

4. One way of thinking of this is that all entrepreneurs are self-employed.

5. Note that these are goods that are already being produced in other, more advanced countries. So saying that there are known, fixed prices is not at odds with the assumption that none of them is produced at home currently.

6. Remember that we normalized the number of workers needed to start a new firm to one.

7. More realistically, the policy experience of a country and the experience of other countries with similar socioeconomic and geographic attributes may influence the range of the policymaker's experimentation draws. For a discussion of the interplay between learning

and policy experimentation, see Mukand and Rodrik 2005. Furthermore, there may be cases where institutional reform would yield a clearcut advantage over tinkering so that $E(b) > E(a)$. We consider this case in section 6.5.

8. Of course, our qualitative results depend on a weaker form of this assumption: as long as incumbent firms bear *higher* adjustment costs when a new policy draw is made from a newly instituted policy regime, our main results remain intact.

9. As we elaborate here, our model generates an inverse relationship between entrepreneurial experimentation and the prevailing wage rates. For empirical evidence, refer to section 6.6. Also, see Iyigun and Owen 1998 and 1999 for some related discussion.

10. For an analysis of the full-information case, see Haussman and Rodrik 2003.

11. As we stated earlier, the new policy draw would not yield any new entrepreneurial ventures, no matter how large a_t is: that is because a_t shifts the productivity of actual and potential modern-sector activities in the same proportion, and does not affect their relative profitability.

12. This simply follows from the fact that $E(a) = E(b) = 1$.

13. These data are accessible in ILO 1995.

14. This R-squared refers to a conventional OLS regression, and not the robust regressions we report in table 6.2.

15. Thus, we do not report these results here.

16. We also explored the results of an instrumental-variables, GMM specification using our actual *ENTRAT* data (which consist of 53 country observations). While we do not report them here, these results were roughly similar to—but slightly weaker than—the ones we present in what follows: *ENTRAT* had a positive impact on subsequent growth in the analogs of columns (4)–(6), and the association between *ENTRAT* and growth was statistically significant at the 5 or 10 percent confidence level in two of the three specifications.

17. See Rodrik and Subramanian 2004; Qian 2003; and van Arkadie and Mallon 2003 on India, China, and Vietnam, respectively.

References

Acemoglu, Daron, Phillipe Aghion, and Fabrizio Zilibotti. 2002. "Distance to Frontier, Selection, and Economic Growth." NBER working paper no. 9066, July.

Blanchard, Olivier, and Michael Kremer. 1997. "Disorganization." *Quarterly Journal of Economics* 112, no. 4: 1091–1126.

Caballero, Ricardo J., and Mohammed L. Hammour. 2000. "Creative Destruction and Development: Institutions, Crises, and Restructuring." NBER working paper no. 7849, August.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R. W. Temple. 2004. "Growth Econometrics." Unpublished paper, University of Wisconsin, October.

Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton. 2000. "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries." *Journal of Public Economics* 76, no. 3 (June): 459–493.

Harberger, Arnold C. 2003. "Interview with Arnold Harberger: Sound Policies Can Free Up Natural Forces of Growth." *IMF Survey* (July 14): 213–216.

Hausmann, Ricardo, and Dani Rodrik. 2003. "Economic Development as Self-Discovery." *Journal of Development Economics* 72, no. 2 (December): 603–633.

International Labor Organization. 1995. "LABORSTA dataset, Table 2.d." Available online at ⟨http://www.ilo.org⟩.

Imbs, Jean, and Romain Wacziarg. 2003. "Stages of Diversification." *American Economic Review* 93, no. 1 (March): 63–86.

Iyigun, Murat F., and Ann L. Owen. 1998. "Risk, Entrepreneurship and Human Capital Accumulation." *American Economic Review* (Papers & Proceedings) 88, no. 2 (May): 454–457.

Iyigun, Murat F., and Ann L. Owen. 1999. "Entrepreneurs, Professionals, and Growth." *Journal of Economic Growth* 4, no. 2 (June): 211–230.

Mukand, Sharun, and Dani Rodrik. 2005. "In Search of the Holy Grail: Policy Convergence, Experimentation, and Economic Performance." *American Economic Review* 95, no. 1 (March): 374–383.

Qian, Yingyi. 2003. "How Reform Worked in China." In *Search of Prosperity: Analytic Narratives of Economic Growth*, ed. D. Rodrik, 297–333. Princeton, NJ: Princeton University Press.

Rodriguez, Francisco, and Dani Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." *NBER Macroeconomics Annual* 15, ed. Ben Bernanke and Kenneth S. Rogoff. Cambridge, MA: MIT Press.

Rodrik, Dani. 1999. "Democracies Pay Higher Wages." *Quarterly Journal of Economics* 114(3) (August): 707–732.

Rodrik, Dani. 2003. "Growth Strategies." In *Handbook of Economic Growth*, ed. Philippe Aghion and Steven Durlauf. Amsterdam: North-Holland.

Rodrik, Dani, and Arvind Subramanian. 2004. "From 'Hindu Growth' to Productivity Surge: The Mystery of the Indian Growth Transition." Unpublished paper, Harvard University, March.

Roland, Gerard, and Thierry Verdier. 1999. "Transition and the Output Fall." *Economics of Transition* 7, no. 1: 1–28.

Sachs, Jeffrey, and Andrew Warner. 1995. "Economic Convergence and Economic Policies." In *Brookings Papers on Economic Activity*, vol. 1, ed. William Brainard and George Perry, 1–95, 108–118. Washington, DC: Brookings Institution Press.

van Arkadie, Brian, and Raymond Mallon. 2003. *Vietnam: A Transition Tiger?* Canberra, Australia: Asia Pacific Press at The Australian National University.

Wacziarg, Romain, and Karen Horn Welch. 2003. "Trade Liberalization and Growth: New Evidence." NBER working paper no. W10152.

III Endogenous Institutions

7 The Role of Higher Education Institutions: Recruitment of Elites and Economic Growth

Elise S. Brezis and
François Crouzet

7.1 Introduction

Higher education institutions have seen much turmoil over the past few centuries. For long, universities were part of the religious establishment, and their main role was to teach liberal arts, philosophy, and theology, though some of them became famous for specific subjects, such as law in Bologna and medicine in Montpellier. Most university students, whose numbers were in any case few, were preparing for a career in the church, even after the Reformation. Then, in the nineteenth and early twentieth centuries, universities underwent reform; they generally became secular, and they started to teach new subjects, particularly the sciences. Still, a university degree was not necessary for a career, except in a few specific professions.

During the last half of the twentieth century, a dramatic change took place in higher education: The number of universities and colleges in the West rose, and the number of students increased even more. Concurrently with this democratization of higher education, universities became heterogeneous not only in their specializations but in their quality. When higher education is provided only to a fortunate few, there is no need for selection, and universities do not differ in their quality and prestige; yet when higher education is democratized and nearly 40 percent of the population attend colleges or universities, uniformity in their quality is impossible. There is, therefore, a distinction between, on the one hand, the elite universities (for which after World War II, selection became meritocratic) and, on the other hand, the rest.[1]

These changes in higher education that occurred in the twentieth century affected the way in which elites were recruited. While before World War II elites had rarely been educated at universities, after World War II they attended elite universities, and the criterion for their recruitment became graduation from elite schools. In consequence, post–World War II, elites were recruited through education in elite universities to which admission was conferred following success at meritocratic exams.

The purpose of this chapter is to examine the effects of this meritocratic recruitment of elites on class stratification and economic growth. We first show that recruitment to a university via a meritocratic method based on entrance exams does not lead to enrollment from all classes of society according to distribution or ability, nor does it necessarily lead to the admission of the most talented. Recruitment by entrance exam still encompasses a bias in favor of elite candidates, because this type of exam requires a pattern of aptitude and thinking that favors candidates from an elite background.

In this chapter, we show that even a slight cultural bias can lead to more than half of students enrolled in elite universities coming from an elite background. This cultural bias, which can be included in the typology of environmental factors, has a magnification effect on class stratification.[2] Therefore the resulting student body is a mostly homogeneous group that is not as open as it should be to the non-elite public, despite the meritocratic selection method of the elite universities. In other words, we show that an elite education leads to a "noncirculation of elites."[3]

Although meritocratic selection should result in the best being chosen to enter the top ranks of public service or business, the framework described herein allows us to show that elite schools and universities have a tendency to recruit in a nondiversified way, resulting in certain classes being overrepresented. Our model emphasizes that despite meritocratic recruitment, elite universities actually recruit from the "aristocracy," and we get a resulting "stratification" of recruitment. In consequence, the fact that over time individuals from the same background are accepted to elite universities is due not to cronyism, but to the system itself, despite the fact that it is meritocratic.

We then examine the consequences of meritocratic recruitment on economic growth. The recruitment of the elite affects the economy, just as it affects the quality of leaders: Having the best citizens as leaders enables efficient and correct choices. We show that the effects of meritocratic selection are dependent on the type of technological changes occurring in the country in question. During times of *innovation*—namely, minor changes in technology—the elite schools optimally fulfill their purpose, since the aptitude acquired at home by the children of the elite class is an advantage regarding the type of technology in use. These students therefore perform better on average than students recruited from the non-elite population. Thus, the noncirculation of elites in this case does not hamper growth.

But during times of *invention*, that is, the emergence of totally new technologies, aptitude acquired via family education is useless, and so lack of circulation of elites is detrimental to the adoption of new technologies. Therefore, in an era of invention, the recruitment of the elite from elite schools actually leads to a lower growth rate.

This chapter is divided into five parts. In section 7.2, we present a short survey of the evolution of elite recruitment in the Western world. In section 7.3, we analyze the effects of meritocratic exams on stratification; section 7.4 examines the effects of meritocracy on economic growth; and section 7.5 concludes.

7.2 Recruitment of the Elite in the Twentieth Century

The main shifts in the recruitment of elites took place in the second half of the twentieth century and are related to meritocratic selection. However, this was not the first time that changes occurred in the way elites were recruited. Before introducing the facts on recruitment in the twentieth century, we present a concise survey of the recruitment and training of elites in earlier times.[4]

7.2.1 Before the Industrial Revolution

Over the centuries, there have been changes in the way the Western world has recruited its elite. As early as Biblical times, Jethro, Moses' stepfather, suggested to him that he should find elite people by looking for "distinguished men, fearing God, liking truth and enemies of luxury" (Exodus 18:21). Later on, when the first king in the Bible was to be chosen, the Prophet Samuel ordered that the King should be selected by lot, in a random way (1 Samuel 10:20). The subject of selecting the elites had also been discussed by the Greek philosophers. Aristotle stressed that a city should be *ruled by the best* ("aristoi" in Greek), and that government should be in the hands of the most able members of society. They should be highly intelligent and educated (as well as brave and temperate) citizens. For Plato and Aristotle, the recruitment of the elite was a crucial element in determining optimal political structure.

Despite this enlightened view, over the centuries, recruitment of the elite was actually carried out via heredity, nepotism, and violence, and the word "aristocracy" came to describe the hereditary upper ruling class. Hereditary monarchy was for centuries considered the most legitimate means of recruitment for rulers, based on the assumption that morality and intellectual prowess are inherited, according to God's will. In traditional European societies until the nineteenth century, membership in the elite was mainly hereditary; noble birth was the rule. Moreover, highly born people were also generally wealthy, even though their wealth was mainly landed. In consequence, the upper elite was made up of large landowners, an *état de fait* that was normal in agrarian societies. Appointments to various state positions (including the armed forces and the church) were made either by patronage or by purchase. Patronage was a matter of family connections, favor, and intrigue.

Nevertheless, there were some channels through which new people emerged regularly into the elite: the favor of the sovereign or of some great lord; military prowess and exploits; amassing wealth through trade or, frequently, involvement in government finance (such as tax farming); and purchase of public offices.[5] Such upward channels involved some meritocratic elements; this was particularly pronounced within the Catholic Church, where some rose as a result of sheer intellectual prowess.

Education was not a necessary element for entering the elite, and there was no specific education for the elite, although most sons of the nobility had private tutors. From the seventeenth century onward, many of them were sent to "high schools" such as the English public schools, or in Catholic countries, colleges run by the Jesuits or the Oratorians. In these schools, pupils received a purely classical education.

7.2.2 *From the Eighteenth to the Twentieth Century*

Changes in the recruitment and training of elites first took place in the late eighteenth century, arising from the needs of modern states. Military schools and various other institutions, such as the French *grandes écoles*, were established for training experts, although only a small proportion of the elites went through these schools and universities at this period.

Then, in the late nineteenth century, in Europe as well as in the United States, two major economic changes had a strong impact. The first was the "second Industrial Revolution," namely, the rise of new industries, like chemicals and electricity, which were science based. The second was the rise of the "corporate economy" and of the Chandlerian managerial enterprise that led to the process by which salaried senior managers largely took over from capital owners and heirs of the founding families. Both engineers and managers needed specialized formal training, while the pioneers of industrialization and their heirs had had only on-the-job training. So there was a clear difference between the education of the leaders of the first and of the second Industrial Revolutions (Kaelble 1979, 29).

On the whole, the late nineteenth and early twentieth centuries saw major changes in the education of the elite, and institutions were reformed or created to provide such training. In most countries, the old "medieval" universities were reformed and expanded during this period, and many new ones were established.

In England, in the late nineteenth century, new or "red brick" universities had been established in provincial cities; they emphasized science and technology and also had schools of commerce. However, the proportion of business leaders who graduated from universities others than Oxbridge was low for a considerable period, and for well-born young men, universities such as Oxford and Cambridge were merely "finishing schools" where they had a good time and made useful connections. A large majority of business leaders—especially in the city, and somewhat less in industry and commerce—had attended public schools before entering university. Education at a major public school was more important in terms of social prestige than a university degree.

In France, the most original character of the system for recruiting and training elites was the role played by elite institutions—the *grandes écoles* (GE), and, more specifically, the *Ecole Nationale d'Administration*, best known by its abbreviation—ENA. The origins of the GE go back to the eighteenth-century military schools, and to the creation of special schools to train engineers that the state needed: the *Ecole des Ponts et Chaussées* (for civil road engineers) was established in 1715, and the *Ecole des Mines* (for mining engineers) in 1783. During the French Revolution, the need for more civil and military engineers was felt. In 1794, the *Ecole Polytechnique* was established.

During the nineteenth and the twentieth centuries, a large number of other GE were gradually established, mainly for training engineers, as *Ecole Centrale* in 1829 (a school to which, after 1840, a significant number of French industrialists sent their sons), but also for giving commercial training, particularly HEC (*Hautes Etudes Commerciales*) in 1881, which can be described as the first French business school. This proliferation resulted partly from deficiencies in the French university system.

During the nineteenth century and at the beginning of the twentieth, there were differences between countries in the importance of education and training of elites. In England, business leaders, most of whom had not had any higher education, were wary of university graduates. This suspicion reflected the cult of the "practical man,"[6] and therefore, most of the British economic elite was recruited and trained via the traditional channels of family connections and patronage, the so-called old boys' networks of those who had attended public schools.

In some respects this system survived into the twentieth century. A high, though decreasing, proportion of top British executives began their careers at the end of their secondary education; they were either "heirs," namely, members of the family that owned the firm and sons of other "good families," who received a top job at once, or men from a more modest background, who had risen within a firm. However, it became increasingly frequent for upper- and upper-middle-class young men to attend university before entering business. In consequence, the percentage of university-educated British executives was lower than in France and Germany, especially in the first half of the twentieth century, but increased over time. From the interwar period, large firms increased their intake of university graduates, many of them from the older universities (especially Cambridge, which had placed greater stress upon the teaching of science and even engineering).

As for the United States, until the end of the nineteenth century, the situation was not that different. Elite members were either heirs, children of rich men who inherited the family firm, or newcomers to the business world who made their fortunes through their own exertions. As in other countries, few of these "self-made men" were of truly proletarian origin, and most came from the lower middle class and the labor aristocracy. In any case, such men, ipso facto, had had very little formal education.

Most of the "heirs" had attended secondary school, had entered the family firm at about age sixteen, and learned on the job. It was widely believed that further studies—especially at a university or college—would be a waste of time and even bad for business, and that direct contact with the latter was desirable as early as possible. Therefore, higher education seemed irrelevant for most professional callings. Until the 1870s, young men who were to follow a business career did not attend universities.

From the turn of the twentieth century, there was a gradual change in the recruitment and training of elites.[7] An increasing number of large companies started to hire people who had received an academic education. In the 1890s, the need for managers who had undergone serious and thorough training was felt, and graduates from the new business colleges, which had been recently established, were hired in large numbers.[8] The founding of the Harvard School of Business Management in 1908 and the creation of the MBA sanctioned this development. During the same period, graduate schools of law and medicine were established in the major American universities.

7.2.3 The Second Half of the Twentieth Century

The recruitment of elites changed dramatically after World War II. In all countries, there was a desire on the part of politicians to "democratize" the elite, and, in consequence, there were significant reforms in the way the elites were recruited, as well as in their education.

After World War II, while a democratization of higher education took place, reflected by an enormous increase in the number of university students, an emergence of two channels of education occurred: one for the elite and the other for the rest. This was emphasized by University of California President Clark Kerr, who set the stage for university for all, but the elite university for the best.[9] In consequence, the elite was trained in elite colleges. Temin described this evolution in the United States: "I was able to identify the colleges attended by 454 CEOs of the Fortune 500 companies. All current business leaders on whom I could find information attended college and almost one-fifth graduated from the Ivy League" (1999a, 32).[10] By the early 1960s, those who had not attended college were precluded from becoming part of the business elite.[11]

In the United States, the changes took place mainly during the 1940s. Until then, there was a group of people who constituted the Establishment: they were male, white, and Protestant (mostly Episcopalian). They were the elite, their children attended the elite universities, and few others could attain any power.[12] Education at an Ivy League university was the entry ticket to the elites of all kinds (except the political elite, which was more of a melting pot), and before 1936, recruitment to universities was based on family and geography.

There was, therefore, a widespread desire to break the hold of this old elite and replace it with a new elite that would be made up of people from a broad range of backgrounds from all over the country, selected on the basis of talent and not birth. There was a need to replace the "aristocratic" and nondemocratic elite with a "brainy" one that would lead the country. This desire was already expressed by Thomas Jefferson more than a century before: "There is a natural aristocracy among men. The grounds of this are virtue and talents . . . while the artificial aristocracy is founded on wealth and birth."[13]

As already emphasized, this wish to find the "natural aristocracy" is not specific to the United States, and has equivalents in Europe, especially in France. However, the American meritocratic way of selecting the most intelligent in order to recruit the best public servants, and letting them run the country, is different from the French one. While France opted for the system of the *grandes écoles* already existing, which were based upon achievement exams, the United States adopted SAT exams.

The SAT, or Scholastic Aptitude Test (itself an adaptation of the Army intelligence test called the Army Alpha), was developed at Princeton University, and placed the emphasis for university admissions on aptitude instead of achievement.[14] The system was slowly adopted by all universities. It was adopted first by Harvard—run at that time by James Conant—who supported a selection process that would lead to the recruitment of his university elite from all social classes, and who felt that achievement tests were unfair to poor children because most had not attended good high schools. Therefore, he called for a system for choosing the elites that was not based on achievement. Harvard thus adopted the SAT for use as a scholarship test during the 1930s; its use then spread as a scholarship test for all Ivy League schools. It took twenty years for the SAT to become a requirement for all applicants to the University of California, and soon afterward to all universities. Standardized testing provided the basis for selection to elite universities.[15]

However, despite the wish to democratize selection, SAT scores were correlated with family education and wealth.[16] Meritocracy did not mean democratization and opportunity for all. The unrealized dream of the virtue of meritocracy as opposed to aristocracy, has been emphasized by Temin. He has shown that the U.S. economic elite is still overwhelmingly made up of white Protestant males, a significant number of whom were educated at Ivy League institutions. The picture has not changed significantly from the 1900s: "The American business elite comes from elite families" just as it does in France or Britain.[17] The fundamental irony of the American meritocracy is that the system finally favored the elite's children. The wish that America would become a classless society through the use of aptitude tests did not come true: meritocracy led to aristocracy.

In order to reduce stratification, the United States elaborated a selection process for minorities, trying to advance the best persons under affirmative action, which can be seen as a "patch" on meritocracy to make it run better. In consequence, meritocracy is a sort of particular system of picking people for the elite, based on one set of abilities. Affirmative action is trying to twist the dials a bit to get more minority representation in the meritocratic elite.

In France, despite a different system of recruitment, the situation is somewhat similar. The GE and especially the ENA play a role similar to the Ivy Leagues universities in the United States: they are elite schools, and very selective. France presently has 302 elite engineering schools with 59,000 students, and 226 commerce or business schools with 64,000 students. These figures may be compared with the million and a half students in universities, which have no entry exam and admit anyone who has graduated from high school (*baccalauréat*).

At the competitive exam (*concours*) to enter a *grande école*, the number of candidates accepted every year is fixed.[18] Moreover, one does not sit for the *concours* just after high school; students first go to specialized schools (*classes préparatoires*) where they are only accepted if they had good grades in high school or at the *baccalauréat*. They study intensively at the *classes préparatoires* for one to four years, after which they take the entry exam for one or several of the *grandes écoles*.[19] Thus, of the half a million people who succeed in the *baccalauréat* each year, 36,000 are accepted in the *classes préparatoires*, among whom only 25,000 will eventually enter a GE in the next few years.

The GE have, over time, become increasingly important to the recruitment of the French business elite. From 1920 onward, over 50 percent of a sample among the leaders of French industry had graduated from engineering schools, and the percentage had reached 70 percent in 1939. According to Lévy-Leboyer (1979, Table 6, 160–161), among a cohort of business leaders over the period 1912–1979, 29 percent of them had graduated from the *Ecole Polytechnique*.

A very specific GE, which has over time become the most elitist of the elite schools, is the ENA. This school is the main channel for recruiting the elite. Forty-seven percent of the heads of the two hundred largest French companies in 1993 came from the civil service (and have been through the ENA). In 1997, 55 percent of the leaders of French CAC 40 firms came from the civil service.[20] It is also the entry to the political elite. Indeed, from 1980 onward, 35 percent of ministers had attended ENA.[21]

The main goal of ENA when it was created was to make changes in the recruitment of the French bureaucratic elites. At the end of World War II, the government and, more precisely, General Charles de Gaulle saw a need to change the recruitment and training of civil servants. It was thought that since recruitment would be meritocratic, only the best would be selected. They would come from *all* classes, in contrast to the previous system for recruiting that was restrictive from a social point of view. So the goals of recruitment to the ENA was social openness, diversification of intellectual origins, and developing a new elite that would be chosen for its talent rather than its link to the elite in power.

The selection for the ENA is more drastic than for the other GE: students have to jump over two fences—an entry exam and a final exam, and the last one is decisive. The department in which an alumnus (*énarque*) begins his employment, from the most prestigious ministries to the least attractive, depends on his or her ranking in the final competitive exam at the end of the second and last year.

As for the entrance exam, the early design to democratize recruitment was embodied in the creation of two separates exams. One is for "students," that is, graduates coming directly from university (generally from law and the humanities) or from an IEP (*institut d'études politiques*). The major IEP, which is situated in Paris, is usually known as *Sciences-Po*. The quality of teaching is high, and *Sciences-Po* is really the main channel toward ENA. After graduating, would-be candidates spend a year or two in special classes, where they are coached for the ENA exam and which are the equivalent of the *classes préparatoire*s for the other GE. As for the written exam, it tests students on broad general culture, though writing some papers on subjects like economics or international relations is required. At the oral exams, the ability to speak brilliantly about a subject one knows nothing about is crucial!

The second exam is reserved for candidates who have spent some years (at least five) in the lower ranks of the civil service; but they must have a university degree, and generally they have been coached for one year in special classes (like the candidates for the first exam). This exam is separate but not very different from the first; it was intended to afford an opportunity for those with a more modest background than alumni from *Sciences-Po* to enter the ENA.[22]

A comparison of ENA recruitment with other GE shows that the ENA recruits approximately one hundred students each year, while *Polytechnique* recruits almost five hundred. Although other GE have a specific technical curriculum, the ENA focuses more on recruiting than on training the elite. ENA students spend their first year working as interns in some public agency (e.g., a regional administration or an embassy). They then return to school for the second year, where the emphasis is again on humanities and social sciences (though in recent years the teaching at the ENA has been mildly "technicized").[23]

As for the democratization of recruitment, after World War II the first few promotions were open to all classes and open to reform. At that time, the ENA was synonymous with innovation and new blood in the administration, and there was a feeling that only the best were chosen. Twenty years later, however, it was apparent that recruitment was sociologically and geographically narrow. The proportion of students in the ENA whose parents belonged to the elite (8 percent of the population) was 44 percent in 1950, and rose to 63 percent in 1980.[24] Thus, starting in the 1970s, an auto-recruitment of the ruling class has taken place, since 8 percent of the population supplies 63 percent of the ENA students, or the next generation of rulers.

Altogether, France has a system that is drastically selective and highly elitist, a system in which the selection becomes even more severe over time. The number of GE students is very low compared to the total number of students in universities, and it is stable while the number of students in universities has increased annually.

In conclusion, in all countries, the education and training of the bureaucratic and business elite, until recently, had little relevance to their future working career. Education at the French GE and at English public schools or Oxbridge imbued a strong feeling of belonging to the elite and laid the foundations for vast networks of relationships.

From World War II onward, the path to elite positions has required attendance at an exclusive school or university, in which recruitment is based on meritocracy. The two main differences between the French and American recruitment systems lay in the method of selection chosen. The first difference is in the type of exam: achievement tests

(France) versus aptitude tests (the United States). The second difference is in the number of times selections are made, and the relative number of recruits.[25]

In the United States, university applicants take the SATs, and those earning the highest scores are usually admitted to the elite universities. Of 2,000 colleges, 50 are considered elite colleges (including the Ivy League). In contrast, in France, of 450,000 students who obtain the *baccalauréat*, only 36,000 enter the *classes préparatoires*, from which only 10,000 will reach the first rank of GE in the next couple of years. So in the United States, the relative number of such "favorites of fortune" is higher than the number of graduates of the ENA and the GE.

However, both systems ultimately lead to a very narrow recruitment process. Both countries tried to react to this narrowing and stratification. The American reaction to its recruitment system was affirmative action for minority group members, whereas the French reaction to its system was the "second entry exam" for admission to the ENA, and open access to universities.

In section 7.3, we examine the reasons why elites are auto-recruited. Temin (1999b) has proposed three causes for auto-recruitment, which he terms "the stability of the elite": discrimination, signaling, and education. Temin rejects the first two,[26] and he concludes that unequal access to education might explain the demographic stability of the elite.[27]

In the next section, we show that recruitment to elite universities by meritocratic exams might be the cause of this stability of elites. We also analyze how different recruitment methods such as SAT or achievement tests affect stratification.

7.3 The Effects of Meritocratic Recruitment on Social Stratification

In this section, in order to examine the effects of meritocratic exam on the intergenerational mobility of elites, we incorporate elements specific to recruitment in countries like the United States and France. In France, there is an entrance exam to the elite schools, based on very broad subjects rather than on specific technical knowledge; they are some sort of achievement tests. In the United States, entrance exams are tests of ability and not achievement: the SAT.

As we have shown in section 7.2, the raison d'être of elite schools is to recruit the most capable students. If information were perfect, the exact value of a given applicant would be known, and elite schools would then choose the best candidates. However, since the information available is imperfect, the best approximation is performance in the entrance exams.[28] We will show that these meritocratic exams lead to class stratification.

We define $I \in [0, 1]$ as the minimum grade necessary to be accepted to the school. If the grade α_i of student i is greater than I, he or she is accepted to the elite school:

$$\alpha_i > I. \tag{1}$$

The performance of a student on the test is based on two elements. The first is his or her ability; more able students get better grades at their exams. We assume that the ability α_i for all students, whatever the social class, is uniformly distributed on $[0, 1]$.[29]

The second element takes into consideration that tests are not perfectly objective, but reflect a culture related to the milieu of the elite with which the examiners for a school are associated. Therefore, students with an equivalent ability, but who are born to the elite and raised in this milieu, will perform better on tests.

The grade of student i who is not part of the elite milieu corresponds to his inherent ability, while the grade of a student from a family in the elite incorporates not only his or her ability, but also the cultural background of his or her family—the inside knowledge specific to the elite milieu, which we define as f.[30] Without loss of generality, we assume that the relation is linear; the grade the student receives is therefore

$$\alpha_i = \begin{cases} a_i & \text{for student } i \text{ outside the elite system,} \\ a_i + f & \text{for student } i \text{ being raised in the milieu.} \end{cases} \tag{2}$$

Since for the whole population, the success is only due to ability, the percentage of accepted students from the entire population denoted γ_{P} is $1 - I$:

$$\gamma_{\mathrm{P}} = 1 - I = \lambda \tag{3}$$

where $\lambda = 1 - I$. λ is a factor that represents the tightness of enrollment. We show that λ affects the size of the stratification effect.

For the students of the elite milieu, f affects the percentage of accepted students, γ_{E}, which is

$$\gamma_{\mathrm{E}} = 1 - I + f = \lambda + f. \tag{4}$$

Let us define β as the ratio of the percentage of the elite children in the elite school over the percentage of elite in the total population. β is in fact the parameter that measures the amount of auto-recruitment and stratification in the economy. When β is 1, then the percent of children from the elite milieu in these elite schools is equal to the percentage of the elite in the population, which means that there is no auto-recruitment and the system is totally democratic. When β is greater than 1, there is auto-recruitment; and the bigger β, the greater the stratification effect in this economy. We now show how meritocratic exams affect β.

Since β is defined as the ratio of the percentage of the elite children in the elite school to the percentage of elite in the total population, it is in fact equal to the ratio of γ_{E} to γ_{p}. Therefore,

$$\beta = \gamma_{\mathrm{E}}/\gamma_{\mathrm{p}} = (1 - I + f)/(1 - I) = 1 + \frac{f}{\lambda} \tag{5}$$

Equation (5) shows that β is a function of f and λ. As explained earlier, β is the parameter that measures the amount of auto-recruitment and stratification in the economy; when β is greater than 1, we get a decrease in the diversity of elites and an auto-recruitment. This framework permits us to show that a very small cultural bias, f, will lead to a strong effect on class stratification, as emphasized in the next proposition:

Proposition 1 A school for elites based on meritocracy leads to class stratification. The parameter of auto-recruitment, β, is related to cultural bias, f, and tight recruitment level, $1 - \lambda$. An increase in one of them leads to an increase in class stratification. Even a small cultural bias brings about the result that children born in the elite are represented by much higher percentages than their ratio to the population.

Proof Homogeneity and stratification is measured by β; the bigger β, the greater the stratification effect in this economy. Since f/λ is greater than 0, β is greater than 1, which means that despite meritocratic exams, stratification exists.

Moreover, from equation (5), the higher the f, the greater the β. On the other hand, the higher the I, the lower the λ and therefore β is bigger.

This proposition states that stratification is a consequence of the advantage to the students raised in the elite milieu due to their cultural background, f. To give a sense of magnitude to our parameters: for an f of 0.07, the milieu gives an advantage of 7 percent (which does not seem a large number, since it seems very reasonable to assume that children raised in the elite get an advantage of around 10 percent); this will lead to $\beta = 8$ (by assuming that $I = 0.99$, which is the case in France).

A stratification effect, β, of 8 means that the percentage of children from the elite milieu who are accepted is 8 times higher than the percentage of children from the total population. In other words, if the elite represent 8 percent of the population, then the elite milieu will supply 64 percent of the students in the elite schools. This matches perfectly the facts found for France, since in section 7.2 we have shown that 8 percent of the population supplies 63 percent of the ENA students, which corresponds to a β of 8. So a small advantage for the elite milieu of 7 percent leads to a major auto-recruitment effect in France.

This simple model shows that the fact that, over time, individuals from the same milieu are accepted to a school for elites is not due to cronyism, but to the system itself, even if it is meritocratic. Elite schools freeze the circulation of elites. Auto-recruitment and stratification are not due to some favoritism, but to imperfect information on the true value of students.

No system can be perfect when there is imperfect information on the genuine talent of people. Recruitment by education and exams automatically advances those who are educated inside the system. Thus, under imperfect information, selection of students through tests leads to a bias, namely, for the same objective ability, students who are not part of the elite milieu will not be accepted, while a student of the milieu will be.

Proposition 1 also permits us to compare the levels of stratification in France and the United States. Indeed, the exams in France are based more on achievements and knowledge tests than the SAT tests in the United States. It means that in France f is higher than in the United States. Moreover, λ is higher in the United States, since in France recruitment levels are tighter. We can therefore conclude that the system chosen in the United States leads to a lower β, that is, to lower auto-recruitment than in France.

This effect of stratification and auto-recruitment due to meritocratic exams takes place only over time, and equation (5) represents the value of stratification in steady state. However, when the system of meritocracy is put in place and is a "new system," then there is no stratification. When new schools are emerging, there is no bias in favor of the elite, and therefore in this case we get that β is equal to 1.

The fact that a bias does not exist when a new system of recruitment occurs may explain why after World War II, there was higher social mobility than in the 1930s. However, in the 1980s and 1990s, as shown by proposition 1, there is much less social mobility toward the top.

In this section, we have analyzed the effect of meritocratic recruitment on stratification and formation of elites. We now analyze the impact on output.

7.4 Meritocracy and Economic Growth

We assume that the quality of the elites has an influence on the level of output, since they are in power positions and make decisions affecting the economy. We therefore assume that output is a function of the factors of production: capital, K, and labor, L; the technology level, A; and the average quality (that we term "value") of the elites, $\bar{V}$, as displayed in equation (6):[31]

$$Y = A\bar{V}F(K, L). \tag{6}$$

So the productivity level is a function of the value of the elites, $\bar{V}$, and of technological progress, A. Technological progress can be due to a change in techniques strictly speaking, but it also includes changes in processes of production, business culture, and methods of management. The evolution over time of technological progress takes two different forms: innovation and invention.[32] Innovation occurs in the context of a given technology; it leads to an increase in productivity based on the current technology and infrastructure (bureaucratic, technocratic). In this type of progress (built on the same structure), the value of students who come from the elite milieu has a value added, f, since they already are familiar with this structure. We can therefore write that the value of a student i in a time of innovation, V_i^n, is a function of its ability as well as the education and culture received in its family environment. Without loss of generality, we assume that the influence of the milieu, f, enters V_i^n linearly:

$$V_i^n = \begin{cases} a_i & \text{for } i \text{ outside the elite system,} \\ a_i + f & \text{for } i \text{ being raised in the elite milieu.} \end{cases} \tag{7}$$

The other type of progress is inventions. While innovations are based on previous technology, major breakthroughs that change the nature of technology fundamentally require that one start anew and most previous learning is lost. This means that the culture the elite has assimilated in his home is no longer useful (in some cases it could even be counterproductive, but not in this model). So the value of a student i in periods of invention, V_i^v is a function only of its ability (and not of its family environment), so that

$$V_i^v = a_i \qquad \text{for all } i. \tag{8}$$

Thus, in periods of innovation the students′ value is distributed on $[I, 1+f]$, while in periods of invention it is distributed on $[I-f, 1]$ (since the students′ ability is, in all cases, distributed on $[I-f, 1]$). The average value of elites in periods of invention[33] and innovation is, respectively,

$$\begin{aligned} \bar{V}^n &= (1+f+I)/2 \qquad \text{for innovation,} \\ \bar{V}^v &= (1-f+I)/2 \qquad \text{for invention.} \end{aligned} \tag{9}$$

The interpretation of equation (9) is that during periods of innovation, but not of technological revolution, the students from the elite milieu contribute an average value of $(1+f+I)/2$, which is a higher value than the average population accepted in the school $(1+I)/2$. Those from the elite milieu increases the average value of the elite in times of innovations, and this results in a higher output (or growth rate). By contrast, during periods of inventions, that is, of technological revolutions, the home culture is not useful, and only pure ability has an effect on output. The students from the elite milieu reduce the average ability and therefore reduce the average value of the elites. We summarize this effect in proposition 2.

Proposition 2 When the world faces innovations, the best elite is the one coming from the elites′ schools; but when the world faces inventions and big changes, diversity of elites is optimal. Auto-recruitment is, therefore, bad for growth, and elite schools are not optimal. Non-circulation of elites resulting from elite schools hampers growth during periods of invention, while it enhances it in times of innovation.

7.5 Conclusion

Over the centuries, there were changes in the way elites were recruited. From the end of the nineteenth century, the ticket to set foot into the elite was to enter an elite university, and until World War II, access to these elite universities was largely restricted to the upper class. After World War II, entry to elite universities was achieved through meritocratic recruitment and was not dependent anymore upon wealth; the best were chosen.

The idea of meritocracy made inroads, and new blood entered elite universities in the United States, Oxbridge in the United Kingdom, and the GE and the ENA in France. Consequently, the first postchange elite was recruited in a diverse way, by successful performance on exams. For the first generation after these changes in recruitment, elite universities were enabled to choose the best, and they provided an opportunity for some who did not belong to the elite milieu to enter the best schools.

In succeeding generations, however, exams have not permitted opportunity for all, as shown by our model. In the second postchange generation, the children of the elite enter the elite schools in greater proportions, due to a cultural bias. In other words, whenever a new system is introduced, the nascent class system is destroyed, yielding a fluid, mobile society. However, from the second postchange generation on, the children of the elite again have an advantage. Our model has shown that meritocratic exams lead to an auto-recruitment of elites, resulting in a stratification effect. Meritocratic choice is therefore not equivalent to equal opportunity, since success in exams is correlated with family wealth and education.

This stratification effect exists in France, but also in the United States. Over time, recruitment either through an SAT or a *concours* leads to a stratification effect. We have shown that the stratification effect will be greater in France due to achievement and knowledge-based tests versus the American SATs, and also due to the fact that in France, recruitment levels are narrower. However, the two levels of recruitment that exist in France lead to an opposite effect. Indeed, as counterintuitive as it appears, the double system of the *baccalauréat* and two years later (or more) the *grandes écoles* entry exams is actually a superior system of recruitment. Despite these differences, over the years, there has occurred some convergence in the way the Western world recruits and educates its elite; with the exception of Germany, the elites are recruited through elite schools and from elite families.

The second part of our chapter has checked the effect of this type of recruitment on economic growth. We have shown that these systems work very well in times of minor changes in technology, and they do allow for economic growth. However, during times of major technological change, the system of elite recruitment can actually cause a slowdown in the adoption of new technologies. Presumably, the best situation would entail periodic changes in the types of exams, causing the circulation of the elite to widen. However, for the French system to accept recruitment to the GE by exams such as the SAT would surely demand another revolution!

In conclusion, this chapter has shown that the stratification effect in the recruitment of elites is due to the entry exams to universities. The democratization of the universities has led to education of masses but not to a "democratization" of the elites. The policies, which were adopted after World War II to widen the recruitment of elites, were at first a success, but over time there was a perverse stratification effect and the circulation of elites receded, with the specified consequences for the economy.

Notes

We wish to thank Philippe Aghion, David de la Croix, Oded Galor, Zorina Khan, Gerard Roland, Kenneth Sokoloff, Peter Temin, and the editors of this book, as well as participants at the Cesifo workshop in Venice for their helpful comments.

1. Today, a degree affects not only the remuneration and career path of students, but also the prestige of the university or college from which they graduate.

2. In this chapter, we do not include heredity (emphasized by Herrnstein and Murray (1994)) as one of the factors leading to class stratification, but only environmental factors as cultural bias.

3. "Circulation of the elite" is an expression coined by Vilfredo Pareto in 1902, who claimed that the elite, in recruiting itself, chose subjects of increasingly mediocre caliber: "Merely a slowing down of this circulation may have the effect of considerably increasing the number of degenerate elements within the classes still possessing power, and—by contrast—of increasing the number of elements of superior quality within the subject classes ... The decadence originates from the fact that the elite, in recruiting itself, chose subjects of increasingly mediocre calibre" (1965, Vol. 1, Introduction).

4. For a more detailed description, see Brezis and Crouzet 2004.

5. Such rises were generally crowned by ennoblement, thanks to which descendants of the "new men" were—after a time—fully integrated into the elite.

6. See Rubinstein 1993.

7. The change was gradual, and clerkship remained the usual form of apprenticeship for a business career.

8. As early as the 1870s, an increase took place in the number of colleges and students. The number of undergraduates rose from 52,000 in 1870 to 238,000 in 1900, and that of graduate students in doctoral programs from 50 to 6,000.

9. Kerr wanted UC-Berkeley to become a highly selective, world-class university with a star faculty that would train the elite of California, which would be selected by testing, while the other colleges would accept all other applicants.

10. Of the 800 chief executives running the largest U.S. public companies in 2003, 87 had MBAs from the three top business schools—Harvard, Stanford, and Wharton.

11. On the other hand, in a recent development, the Information Technology Revolution (ITR) has offered opportunities to new entrepreneurs such as Bill Gates to rise quickly in the business world without graduating (but also for some, to fall fast later on). Yet unlike the tinkerers of the Industrial Revolution, such entrepreneurs need a strong scientific education.

12. See Miller 1949 and 1950. Taussig and Joslyn (1932, 240) have shown that in 1930, 80 percent of the business leaders came from the top 7 percent of the population.

13. A letter to John Adams in 1813 (see Cappon 1959).

14. It was developed by Carl Brigham, a psychologist (see Brigham 1923). It is a type of general intelligence test, and the verbal portion is similar to an IQ test. It was created and administered by the Educational Testing Service (ETS), a privately funded, nonprofit organization. For more details on the history of the SAT, see Lemann 1999.

15. There was a fierce debate about the success of ability tests as the basis for meritocracy. For the argument against such tests, see Hoffmann 1962, and also Nairn and Nader 1980, while Gardner (1995) presented a plea in favor of these tests. See also Jensen 1989.

16. Research on the variables affecting SAT results is numerous. See, in particular, Bouchard and McGue 1981, Neal and Johnson 1996, and also Hernnstein and Murray 1994.

17. Temin 1999a, 33. Although the percentage of workers entering the elite class in the United States in the 1960s, however, was double what it was in Britain, France, and Germany (Blau and Ducan 1967). As noted by Temin (1999a, 32) and Kingston and Lewis (1990, 111): "Approximately one quarter of 1986 college freshmen at highly selective universities come from families with incomes over $100,000, that is, from the extreme upper tail of the income distribution." It should be noted that this lack of change in the economic elite occurred despite the fact that the makeup of the political elite has changed markedly over the century.

18. In all these schools, the number of entrants was, and is, not large: the students admitted per year in the five biggest engineering schools were 320 in 1860, and 1176 over the period 1919–1932 (see Lévy-Leboyer 1979, 152).

19. Moreover, there is a hierarchy among GEs; the first-rank schools are usually in Paris. There is a wide gap in prestige and also in job opportunities for graduates between GE of the first rank—Polytechnique, Centrale, HEC, and provincial commerce schools. There is also a hierarchy between *classes préparatoires* for the main GE. Actually those that are located in three or four big high schools on the Left Bank in Paris supply a large majority of students who succeed at the exams for GE of the first rank. The fate of students who fail at the concours is to enter university, where they do well thanks to their intensive work in a *classe préparatoire*. Moreover, groups of engineering schools have a common exam, and candidates who do not well enough to be accepted to the top GE can enter the less prestigious ones.

20. See Baverez 1998. It includes the shares of the forty most important firms in France, the French Dow-Jones.

21. Presidents Giscard d'Estaing and Jacques Chirac and Prime Ministers Laurent Fabius, Michel Rocard, Alain Juppé, and Lionel Jospin also went to ENA.

22. As a matter of fact, and generally speaking, persons who entered the ENA through the bureaucrats' exam were less successful in their later careers than the ones taking the students' exam, because they did not do as well on the final comprehensive exam (which is the same for all students). Presently 60 percent of entrants are "students," and 40 percent are "civil servants."

23. It is commonly remarked that the internship during the first year is very useful, while students do not learn much during their second year at the school itself.

24. See Gaillard 1995, 105–108. However, each graduating class includes a few persons from modest backgrounds, some of whom go on to brilliant careers.

25. Another main difference is the tuition paid. While in the United States, tuition at elite universities can run around $100,000, in France, not only are the universities and *grandes écoles* almost free, but in the *grandes écoles* with a vocation of serving the state (*Polytechnique*, ENS, and ENA) students are paid! Actually, many of them serve the state for a few years and afterward enter the business world.

26. Temin also rejects the possibility raised by Taussig and Joslyn (1932) of a fundamental inequality of native endowments.

27. Indeed, primary and secondary education, especially in public or state schools, has fallen into crisis in all Western countries, owing to ill-conceived reforms, the breakdown of discipline, and the low quality of many teachers (see Temin 2002). This markedly restricts opportunities for bright young people from modest backgrounds to excel in their studies, win scholarships, and attend university. In France today, only children from middle-class or even upper-middle-class families can obtain a good high school education.

28. Moreover, tests also display a reliability problem, namely, that there is similarity in a given subject's exam scores on different runs of the exam. We discuss this problem later.

29. As mentioned earlier, the bias is only due to cultural background. We are aware that some empirical results show that ability is not uniformly distributed (Herrnstein and Murray 1994), and some theoretical models explain why effort, and therefore ability, would be different in the different social classes (see Durlauf 1999; Arrow, Bowles, and Durlauf 2000). However, the assumption that ability is uniformly distributed is often adopted in models on mobility; see, for instance, Galor and Tsiddon 1997.

30. The problem of reliability of exams can be incorporated in the parameter f. Exams such as those in France are subject to reliability problems higher than the SAT, due to subjectivity problems. Moreover, students who are not "outstanding" but on this particular day felt well would be accepted, while some more brilliant were not, because it was not the subject in which they excelled, or it was not the right day. This problem is less acute in the United States.

31. See Brezis and Crouzet 1999.

32. We use the typology formulated by Arrow. The effects of these different types of technology on the economy were already analyzed in other models (see Brezis, Krugman, and Tsiddon 1993; Galor and Tsiddon 1997). However, they were not incorporated in an analysis of the recruitment of elites.

33. In equation (9), the average value of elites is given only for students belonging to the elite milieu, and we did not take into consideration the other students, since their average value, during periods of inventions or innovations, is always $(1 + I)/2$.

References

Arrow, K., S. Bowles, and S. Durlauf, eds. 2000. *Meritocracy and Economic Inequality*. Princeton, NJ: Princeton University Press.

Baverez, N. 1998. "Etrange capitalisme à la française." *Sociétal* 17 (March): 30–31.

Blau, P. M., and O. D. Duncan. 1967. *The American Occupational Structure*. New York: Wiley.

Bouchard, T. J., and M. McCue. 1981. "Familial Studies of Intelligence." *Science* 212: 1055–1059.

Brezis, E. S., P. Krugman, and D. Tsiddon. 1993. "Leapfrogging in International Competition: A Theory of Cycles in National Technological Leadership." *American Economic Review* 83, no. 5: 1211–1219.

Brezis, E. S., and F. Crouzet. 1999. "Elites Schools, Circulation of Elites and Economic Development: The ENA case." In *Elites, Minorities and Economic Growth*, ed. E. S. Brezis and P. Temin, 235–249. New York: Elsevier, North-Holland.

Brezis, E. S., and F. Crouzet. 2004. "Changes in the Training of the Power Elites in Western Europe." *Journal of European Economic History* 33, no. 1: 33–58.

Brigham, C. C. 1923. *A Study of American Intelligence*. Princeton, NJ: Princeton University Press.

Cappon, Lester J. 1959. *The Adams-Jefferson Letters: The Complete Correspondence between Thomas Jefferson and Abigail and John Adams*. Chapell Hill: University of North Carolina Press.

Durlauf, S. 1999. "The Memberships Theory of Inequality: Ideas and Implications." In *Elites, Minorities and Economic Growth*, ed. E. S. Brezis and P. Temin, 161–178. New York: Elsevier, North-Holland.

Gaillard, J.-M. 1995. *L'ENA, Miroir de L'Etat*. Bruxelles: Complexe.

Galor, O., and D. Tsiddon. 1997. "Technological Progress, Mobility, and Growth." *American Economic Review* 87: 363–382.

Gardner, J. 1995. *Excellence: Can We Be Equal and Excellent Too?* New York: Norton.

Herrnstein, R. J., and C. A. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hoffmann, B. 1962. *The Tyranny of Testing.* New York: Crowell-Collier Press.

Jensen, A. R. 1989. "Raising IQ without Raising Intelligence?" *Developmental Review* 9: 234–258.

Kaelble, H. 1979. "L'évolution du recrutement du patronat en Allemagne comparée à celle des Etats-Unis et de la Grande-Bretagne depuis la Révolution Industrielle." In *Le patronat de la seconde industrialisation*, ed. M. Lévy-Leboyer, 15–36. Paris: Les éditions ouvrières.

Kingston, P. W., and L. S. Lewis. 1990. "Undergraduates at Elite institutions: The Best, the Brightest, and the Richest." In *The High Status Track*, ed. P. W. Kingston and L. S. Lewis, 105–120. Albany: State University of New York Press.

Lemann, N. 1999. *The Big Test. The Secret History of the American Meritocracy*. New York: Farrar, Straus and Giroux.

Lévy-Leboyer, M., ed. 1979. *Le patronat de la seconde industrialisation*. Paris: Les éditions ouvrières.

Miller, W. 1949. "American Historians and the Business Elite." *Journal of Economic History* 9(2): 184–208.

Miller, W. 1950. "The Recruitment of the American Business Elite." *Quarterly Journal of Economics* 64, no. 2: 242–253.

Nairn, A., and R. Nader. 1980. "The Reign of ETS: The Corporation that Makes Up Minds." *Today's Education* 69, no. 2: 58–64.

Neal, D. A., and W. A. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104: 869–895.

Pareto, V. 1935. *The Mind and Society*. New York: Harcourt Brace.

Pareto, V. 1965. *Les systèmes socialistes*. Genève: Droz.

Rubinstein, W. D. 1993. *Capitalism, Culture and Decline in Britain, 1750–1990*. London: Routledge.

Taussig, F. W., and C. S. Joslyn. 1932. *American Business Leaders: A Study in Social Origins and Social Stratification*. New York: Macmillan.

Temin, P. 1999a. "The American Business Elite in Historical Perspective." In *Elites, Minorities and Economic Growth*, ed. E. S. Brezis and P. Temin, 19–40. New York: Elsevier, North-Holland.

Temin, P. 1999b. "The Stability of the American Business Elite." *Industrial and Corporate Change* 8 (June): 189–212.

Temin, P. 2002. "Teacher Quality and the Future of America." *Eastern Economic Journal* 28: 285–300.

8 Growth and Endogenous Political Institutions

Matteo Cervellati,
Piergiuseppe Fortunato, and
Uwe Sunde

8.1 Introduction

Democratic societies are associated with higher levels of economic development and faster income growth than nondemocratic societies. While this correlation is often taken as a stylized fact, relatively few theoretical contributions investigate the relationship between the evolution of political institutions and the level of economic development or economic growth. The observed positive correlation indeed raises the question of whether democratic systems exhibit inherently more efficient characteristics and are therefore more conducive to economic development than nondemocracies, or whether the adoption of more democratic political institutions is only a by-product of economic development.

Several empirical contributions found that higher economic development apparently serves to stabilize democratic systems (see Przeworski and Limongi 1997; Przeworski et al. 2000). Some empirical evidence seems to indicate that there is also a positive effect of economic development on the probability as well as on the timing of democratization in addition to the positive effect of development on the stability of democracies (see Barro 1999; Boix and Stokes 2003). Recent empirical results by Acemoglu et al. (2005) cast doubt on the robustness of this effect, however. They come to the conclusion that there is no evidence for a causal effect of income on democracy when taking country specific or historical determinants of both economic and political development into account. While democratic structures in turn seem conducive and indeed important for economic development, it has also proved difficult to establish a direct link between the adoption of democratic institutions and subsequent economic performance. This is documented by the findings of La Porta et al. (1999) and Glaeser et al. (2004) that cast doubt on a direct effect of democracy on economic growth. On the other hand, empirical results reported by Minier (1998) and Papaioannou and Siouraounis (2004) come to the conclusion that democratic structures have a causal effect on subsequent growth.

Rather than considering a direct link, several contributions have tried to establish the role of democratic political institutions for development through different indirect channels. In fact, democratic systems are characterized by a complex set of rules governing social interactions and the resolution of conflicts of interests, which potentially play an important role in shaping the state's interventions in the marketplace. Democratic structures, the argument goes, provide a better environment for economic activity, and institutions are in fact an important underlying factor in explaining the huge differences in incomes and growth across countries. Tavares and Wacziarg (2001), for example, report evidence that democratic countries exhibit faster growth than nondemocratic countries because they provide better environments for investment—in particular, in human capital. Also the returns to human capital in terms of growth seem higher in democracies than in countries with less democratic regimes. There is also evidence that good institutions are the major determinant for economic performance in terms of income levels. Rodrik, Subramanian, and Trebbi (2004), for example, show that once institutions are controlled for, the degree of openness has no direct effect on incomes, while geography has at best weak direct effects. These results are in line with other empirical studies like Mauro 1995, Hall and Jones 1999, and Acemoglu, Johnson, and Robinson 2001, and suggest that countries with better political institutions, more secure property rights and a well-functioning system of checks against government's power will invest more in both physical and human capital and will use these factors more efficiently to produce a greater level of income. Good institutions and democracy, however, seem to mutually reinforce each other, as is suggested by the empirical findings of Rigobon and Rodrik (2004).

In search for the precise channels and institutional details that determine economic outcomes, a substantial literature analyzes the economic consequences of the political institutions arising under democracies. This includes, among others, investigations of the effects of the political system, the role of voting systems (majoritarian vs. proportional) or of the form of state (unitary vs. federal) or the form of government (presidential vs. parliamentary), to name a few, as well as their implications for various governmental activities and economic performance in general; see, for example, the recent books by Persson and Tabellini (2000, 2003) and Alesina and Glaeser (2004) for surveys of theories and empirical evidence. Several findings emerge. The rules governing the aggregation of conflicting interests—in a word, the political institutions—are not neutral. These rules have a first-order effect on the economic performance and the growth possibilities of a community as a whole, but also on the relative well-being of its various members. Institutions appear to exhibit a large degree of persistence and path dependence; compare, for example, the discussion of the differences between the United States and continental Europe provided by Alesina and Glaeser (2004), who document the long-term effects of the early constitutional stages.

Once these findings are acknowledged, the next logical questions that arise are why not all countries adopt democratic regimes, under which conditions countries democratize, and which conditions are the driving forces for democratization. The question of why countries democratize(d) has triggered substantial research effort, particularly in the political sciences, but it appears far from being a settled issue. Recent economic theories of democratization highlight different channels of transition from oligarchy to democracy. In a series of articles, Acemoglu and Robinson (2000, 2001, 2003, 2004) emphasize the role of coups and suggest that the threat of a revolution might have been crucial in inducing incumbent elites to give up their monopoly of political power and extend the franchise to larger groups of the population.[1] Democracy essentially serves the role of a commitment device since under oligarchy, the elite cannot credibly commit to future redistribution. The elite receives no intrinsic gains from democratization, but it is forced "from below" to concede power and, eventually, redistribute to the poor. Another line of research highlights the productive function of democratic government and argues that it was actually in the interest of the elite itself to democratize. Lizzeri and Persico (2004) show that in some cases, like England, democratization might have actually been in the elite's own interest. The reason is that the provision of public goods, or the prevention of inefficient rent-seeking and corrupt behavior, was easier under democracy than under oligarchy, as a consequence of the stronger checks and balances, and of the possibility of spreading responsibility across more shoulders. An alternative argument for why the elite might prefer to give up their monopoly in political power, based on superior possibilities of property rights protection under democracy, is provided by Gradstein (2006). Bourguignon and Verdier (2000) propose a model in which democracy provides better incentives to accumulate growth-enhancing human capital, inducing the oligarchic elite to release political power and trigger a democratic transition for efficiency reasons. These theories of democratization "from above" all emphasize that certain factors or public goods, such as property rights protection or education, become more important as an economy develops and are more easily provided under democracy.[2] In contrast to the previously mentioned line of argument, however, the latter papers provide arguments for the elite implementing democracy "from above" to reap the benefits of that form of government but without being, strictly speaking, under a serious treat of a revolution or coup.[3] Taken together, the theories of democratization "from below" implicitly focus on the redistributive role of democracies, while theories of democratization "from above" stress the productive role of public goods and service provision as driving forces behind democratic transitions. Economic development is a major determinant for democratic transitions in both lines of literature.

This chapter provides a simple theory of endogenous political institutions based on the interaction between the intertwined processes of economic development and democratization. The transition to democracy is seen as an endogenous event. While it is certainly true that the process of institution building is incremental over time, it is possible to identify key periods for the formation of political institutions in the history of each country. In a long-run perspective such as the one adopted here, democratization can therefore be interpreted as a unique event characterized by the abolition of oligarchic states. We propose a simple dynamic model of democratization that illustrates the different effects of political institutions on the primary functions of the public sector, and that can generate both types of democratization, democratization "from below" and "from above." The basic idea underlying our approach is that democratization is essentially about the provision of productive public goods and the redistribution of incomes. In particular, as argued in the literature, the extension of the franchise and democratization can occur because it is in the elite's own interest, a democratization "from above," or because the elite are forced to democratize through the threat of open conflict, a democratization "from below." How democratization proceeds depends primarily on the economic environment, the level of development, and development history, and is therefore to a large extent endogenous. The main differences between oligarchy and democracy concern public-goods provision and redistribution. In particular, the fact that decisions about redistribution and public-goods provision are made by different groups of interest under both regimes implies different public policies. Since our theory focuses on the dynamic emergence of different transition regimes, we abstract from modeling constitutional details. These are the focus of several other recent contributions.[4] We rather turn our attention to the two main traditional functions of governments, namely, public-goods provision and income redistribution. The model predicts a permanent bidirectional feedback mechanism between political institutions and economic development. In terms of public policies, democracies are predicted to create environments that are more favorable for economic activities (i.e., are more efficient in the public-goods provision) than the ones implemented under oligarchies. Democracies, on the other hand, implement more progressive income redistribution.

This chapter is therefore at the intersection of several branches of literature. Apart from contributing to the recent literature on democratization and the forces that drive the transition that was mentioned earlier, our model contributes to the literature on the transition between different regimes of long-term growth. The mechanism driving our results is based on a positive feedback effect between economic development and the democratic transition, where the acceleration in growth derives from the more efficient provision of productive public goods under democracy. This channel is complementary to the channels based on fertility in the model by Galor and Weil (2000) and based on human evolution in Galor and Moav 2002, the role of life expectancy in Cervellati and Sunde 2005, and the role of public education in Galor, Moav, and Vollrath 2004. For an exhaustive survey on this literature, see Galor 2005.

The chapter proceeds as follows. Section 8.2 lays out the basic economic framework and the potential for political conflict. Section 8.3 analyzes the model and presents the major result, a characterization of the interactions between economic and political development of an economy. Section 8.4 discusses the model implications in a historical perspective and presents some empirical findings corroborating the theory, while section 8.5 discusses some of the simplifying assumptions. Finally, section 8.6 concludes and points at directions for future research.

8.2 The Basic Framework

This section presents the economic environment and the decision problem faced by individuals with different factor endowments. We then introduce and discuss the potential for political conflict that arises in this economy.

8.2.1 Economic Environment

We consider an economy that is populated by an infinite sequence of subsequent generations t of individuals i. Each individual has one parent and one offspring, and there are no fertility decisions to be made. Consequently, there is no population growth over generations, with the size of each generation being $L_t = L$. During their life, individuals inelastically supply one unit of labor on the labor market and earn in exchange a competitively determined wage for their labor input. We abstract from labor-leisure choices. Moreover, individuals are endowed with physical capital, which they inherit as bequests from their parents. A fraction $0 < \gamma < 1/2$ of individuals is also endowed with land, while all the land is distributed equally among landowners. Individuals maximize their utility, which is logarithmic in consumption c and bequests b,[5]

$$u_t^i = u(c_t^i, b_t^i) = (1 - \beta) \log c_t^i + \beta \log b_t^i. \tag{1}$$

All individuals therefore optimally choose to spend a constant fraction $(1 - \beta)$ of their individual income y_t^i on consumption, such that $c_t^i = (1 - \beta) y_t^i$, while they bequeath the rest of their income to their offspring, hence $b_t^i = \beta y_t^i$. To keep things simple, we assume that bequests can only be invested into physical capital K, and that, conversely, capital can only be created through investing the bequests of the preceding generation. There is no other possibility to invest resources in capital formation. At the end of a generation's lifetime, its capital fully depreciates. Consequently, the capital stock available to an individual corresponds to his parents' bequests, such that $k_t^i = b_{t-1}^i = \beta y_{t-1}^i$. Land resources are ready to use for production for their owners. Moreover, land does not depreciate. Land is bequeathed from generation to generation.[6] Individuals use all their factor endowments for the generation of income by supplying them to the production process and selling them on the respective factor markets. Individual incomes are thus determined by the respective endowments and the corresponding factor prices realized on the competitive factor markets. For notational convenience, we denote aggregate variables by uppercase letters, and individual variables by lowercase letters. Consequently, the aggregate resources available in the economy during the existence of generation t are labor input L, an aggregate capital stock $K_t = B_{t-1} = \int b^i_{t-1}\, di$, and land N. Also, we introduce the following notation for average per capita variables: average individual incomes $y_t = Y_t/L$, average capital endowment $k_t = K_t/L$, and average land endowment $n = N/L$.

The economy is fully competitive, and all resources are employed in the production of a single commodity Y according to a production technology exhibiting constant returns to scale of the form

$$Y_t = [(1 + G_t)A_t K_t + N]^{\alpha} L^{(1-\alpha)}. \tag{2}$$

Besides the resource inputs, production is affected by a productivity index A_t, which reflects the technological state of the art of production, and by a productivity-enhancing public good G_t, which reflects, for example, infrastructure. Public-goods provision is discussed in more detail in section 8.2.2. Technological progress, as implied by the production function, relatively favors capital-intensive production as opposed to land-intensive production. This is expressed by the fact that productivity of physical capital in the form of A changes over the course of generations, while that of land remains constant and is normalized to 1. To keep the model simple, and since we are not interested in analyzing the determinants of productivity growth, we assume that technological innovations arrive only with the birth of a new generation. The process of technological progress is exogenous according to[7]

$$\frac{A_t - A_{t-1}}{A_{t-1}} = a_t = a > 0 \qquad \forall t. \tag{3}$$

The production function is formally equivalent to the production of a homogeneous commodity in two distinct sectors, one employing exclusively land resources together with labor, and the other exclusively physical capital together with labor.[8] Since the economy is competitive, all factors are paid according to their marginal products. For convenience, we normalize population size to 1 in what follows, such that $L_t = 1\ \forall t$. Hence, equilibrium factor prices in terms of wages, capital rents and land rents, in the economy are given by

$$w_t = (1 - \alpha)[(1 + G_t)A_t k_t + n]^{\alpha}; \tag{4}$$

$$r_t = \alpha[(1+G_t)A_t k_t + n]^{\alpha-1}(1+G_t)A_t; \tag{5}$$

$$\text{and } \rho_t = \alpha[(1+G_t)A_t k_t + n]^{\alpha-1}, \tag{6}$$

respectively. The production technology is therefore able to replicate the permanent growth in capital stocks and incomes experienced by most countries in the Western world. Moreover, while the implied income share of labor is stable over generations, as was the case in history, the incomes generated by capital grow at the expense of the incomes generated by land over the course of development; see also Acemoglu and Robinson 2003. Individual incomes, which can be allocated optimally to consumption and bequests, are determined by the individual resources employed in the production process and the respective rents accruing to them. Hence, all individuals earn a labor income plus a capital income. Those individuals i belonging to the fraction γ of the population owning land, which we denote in the following by $i \in E$ and refer to as the "landlord elite," additionally earn income from renting their land to the production process. Note that due to the equal distribution of land among the elite, every landowner has land resources of $n^E = n/\gamma$. On the other hand, members of the group without land, the landless people or "proletariat," $i \in P$, have no land, so $n^P = 0$, and hence also enjoy no incomes from land resources. Individual gross incomes can thus be written as

$$y_t^i = w_t + r_t k_t^i + \rho_t n_t^i \qquad \text{with } i \in \{E, P\}. \tag{7}$$

Substituting with the expressions for equilibrium factor prices given by conditions (4), (5), and (6), and denoting effective physical capital as $\tilde{k}_t(G_t)$, with

$$\tilde{k}_t(G_t) := (1+G_t)A_t k_t, \tag{8}$$

income of individual i, $i \in \{E, P\}$, can be expressed as

$$y_t^i = (\tilde{k}_t(G_t) + n)^\alpha \left[(1-\alpha) + \frac{\alpha \tilde{k}_t(G_t)}{\tilde{k}_t(G_t) + n}\frac{k_t^i}{k_t} + \frac{\alpha}{\tilde{k}_t(G_t) + n} n^i\right]. \tag{9}$$

This immediately implies that average per capita income in the economy can be calculated as $y_t = (\tilde{k}_t(G_t) + n)^\alpha$.

8.2.2 *Institutions and the Public Sector*

Next, consider the role of the state. The main purpose of the chapter is to provide a simple model that allows us to characterize the dynamic interdependencies between economic development and political development in terms of democratization. Political decisions are essentially made along two dimensions, the size and the structure of the state in the form of the budget and its use. The total budget is given by tax revenues R. Political decisions also always affect the use of this budget, which is subject to the fundamental trade-off between efficiency and equity. Efficiency-enhancing activities of the state are represented by the provision of a public good G, which enters the production function (2) in the form of higher productivity of physical capital. On the other hand, the state can pursue equity-driven activities, condensed as purely nonproductive lump-sum redistribution in form of transfers T, which are distributed equally among the population. We assume that there are no inefficiencies affecting either public-goods provision or redistribution, in the sense that neither of these two uses of tax revenues implies a waste of income. Rather, every unit of income used for public-goods provision produces one unit of public good, and likewise for redistribution. The budget must be balanced for every generation, since there are no capital markets allowing for intergenerational loans and debt. The budget is financed by proportional income taxation, implying a budget of the state for a given generation of individuals of $\tau Y_t \geq G_t + T_t$.

Note that we abstract from timing issues regarding production, taxation of income, and public-goods provision or redistribution. Rather, this formulation is meant to highlight the role of the size and structure of the public sector for individuals, while they themselves have to decide about both dimensions. Meanwhile, we neglect intergenerational issues, since they do not add fundamental insights to the main argument of our chapter.[9] In the following, the tax rate τ required to finance the public sector, as well as the amounts of redistributive transfers T and public goods G to be provided by the public sector, are determined as the outcome of a political process to be specified next. Of course, given τ and G and the respective total production outcome Y, the size of the public sector τY as well as the size of the redistributive component of the public sector T are determined residually, so that by choosing two variables the size and structure of the public sector are fully determined.

8.2.3 *Political Conflict and Timing of Events*

Size and structure of the public sector are chosen by the respective group of the population that is in power. Hence, power itself is defined as the possibility to decide upon issues such as public-goods provision and redistribution. Public-sector variables are essentially determined by the median voter of the respective electorate. Individuals are only heterogeneous with respect to whether they own land or not, and hence there are only two political regimes: oligarchy, where one group of individuals has exclusive political power, while the other group has no vote; and democracy, where all individuals, regardless of their status with respect to land ownership, enjoy suffrage. Despite having exclusive decision power, we assume that an oligarchic elite cannot forcefully tax and expropriate the politically subordinate class. Hence, if the elite desire a budget for some purpose—for example, the provision of productive public goods—they can only finance the required tax revenues themselves, but not force nonelitist people to participate. A crucial feature of democracy is the fact that the rules of the "democratic game" are fixed and known to everyone, in particular when it comes to making collective decisions, such as the size and structure of the state. The distinction to oligarchy in this respect is that the ruling oligarchic elite set the rules themselves, and hence can change them unilaterally, for example, deciding autonomously on the amount of public-goods provision. This is not possible under democracy.

Under a landlord oligarchy, the elite can determine T as well as the optimal level of public-goods provision G, both of which landlords have to fully finance themselves. Hence, the elite completely determine the public sector. Under democracy, on the other hand, the constitution sets the rules for redistribution and public-goods provision, and the levels of public-goods provision G and redistribution T are determined by majority rule. Since $\gamma < 1/2$, this means that it is essentially chosen by a member of the landless, the group of the median voter.

Following the historical experience, we assume that initially political suffrage was confined to the landowning elite only, implying an oligarchy of landowners. Of course, there are possibilities to change the political regime. Clearly, the respective ruling elite can offer to give up exclusive political power and extend the suffrage to other individuals as well.[10] On the other hand, if this is not the case, the politically excluded may try to obtain power by going to open conflict and violently challenging the ruling elite. To model this possibility, we adopt a "guns model," according to which the winner of an open conflict, if it arises, is determined by the group with preponderance in fighting power. Fighting power is determined by all the resources, persons, and physical capital that are available to a specific group. In the current context, there are only two observationally distinct groups, where the landlord elite are able to unleash a total conflict power of $\gamma(K_t^E)$, while members of the proletariat are able to set free a fighting power of $(1-\gamma)K_t^P$.[11] Note that realizing fighting power effectively and credibly does not require any investments. Rather, the resources can be thought of as being fully reversible, leading to conflict potential that can be mobilized instantaneously and costlessly in case an open conflict occurs. Consequently, the outcome of an open conflict depends on the sign of the following "guns condition":

$$\gamma k_t^E \gtreqless (1-\gamma)k_t^P. \tag{10}$$

In other words, the elite prevail with their political will if they have more conflict potential, namely, if the left-hand side is larger than (or equal to) the right-hand side, while the landless people or proletariat enforces their desired political system if the opposite is true. Note that, since they constitute the majority and since expropriation of rents and discriminatory taxation is ruled out, the people face no trade-off between a populist oligarchy or democracy as the elite do between elitist oligarchy and democracy.

The timing of events faced by every generation t, in the example of an oligarchy of the landlord elite, can be summarized as follows:

1. Birth, inheritance, and investment of bequests;
2. Elite: decision about defending oligarchic status quo or making a democratic offer;
3. People: decision about agreement or disagreement to decision of the elite;
4. Conflict resolution;
5. Implementation of political regime and public policy, production;
6. Consumption and bequest decision, death.

After birth, and the realization of bequests and investment, the respective elite can either decide to remain in power and opt for the status quo, or decide to make a democratic offer. This offer implies an extension of suffrage to the respectively politically excluded group. Under democracy, where nobody is excluded from political participation, the entire electorate makes a decision to keep democracy or to move to an oligarchy. Under oligarchy, the politically disenfranchised people, on the other hand, can then either choose to accommodate the elite's proposal, or choose to challenge it by going to open conflict.[12]

Once the potential conflict is resolved, the consequential political system materializes, and the associated decisive voter makes his or her decision about the public policy to be implemented. Then production takes place under this system, in particular, under the resulting taxation, and the public good provision and redistribution schemes that are implemented. Eventually, people consume or bequeath their remaining net income, and die. This completes the description of the model.

8.3 Development, Democratization, and Their Interdependencies

This section first establishes some basic results concerning the dynamics of the development process and provides an analysis of the decision problems faced by members of the different groups in the economy. Using these results, we then turn to the characterization of the processes of economic and political development, and highlight their interdependencies by considering development as the succession of generations and their political and economic decisions within an evolving environment.

8.3.1 The Provision of Public Goods

From the exogenous productivity growth given by (3), and the fact that capital is only created through bequests, it follows that both incomes and capital endowments are increasing from generation to generation. This is true regardless of the political regime and regardless of whether landowners or landless are concerned. The first useful result concerns the evolution of capital endowments of landowners and the landless, which asymptotically converge, regardless of the political environment and the level of public good provision. As an index of relative inequality in capital endowments, consider the ratio of individual *i*'s capital endowment to the average capital endowment per head in the economy,

$$\lambda_t^i := \frac{k_t^i}{k_t}, \qquad i \in \{E, P\}. \tag{11}$$

We can then state the following result:

Lemma 1 For any $\{\gamma, n, a\}$, $\lim_{t\to\infty} \lambda_t^E \searrow 1$ and $\lim_{t\to\infty} \lambda_t^P \nearrow 1$.

Proof Using the expressions for average per capita income, and the expressions for equilibrium factor prices, one has

$$\lambda_t^i = \frac{y_{t-1}^i}{y_{t-1}} = (1-\alpha) + \frac{\alpha n^i}{\tilde{k}_{t-1}(G_t) + n} + \frac{\alpha \tilde{k}_{t-1}}{\tilde{k}_{t-1}(G_t) + n} \lambda_{t-1}^i. \tag{12}$$

The initial conditions $k_0^E = k_0^P = 0$ imply that $\lambda_t^E > 1$ and $\lambda_t^P < 1$ $\forall t > 0$. Relative inequality in capital endowments of family i converge to a steady-state value

$$\lambda_*^i = \frac{(1-\alpha)(\tilde{k} + n) + \alpha n^i}{\tilde{k}(1-\alpha) + n},$$

which depends on the steady-state value of $\tilde{k}$. Due to unbounded technical progress, incomes and capital endowments increase over generations, implying $\lim_{t\to\infty} \tilde{k}_t = \infty$. Since land is fixed, using l'Hôpital's rule, this implies that $\lim_{t\to\infty} = \lambda_*^i = (1-\alpha)/(1-\alpha) = 1,\ i \in \{E, P\}$, which proves convergence. Moreover, condition (12) implies that

$$\lambda_t^P = (1-\alpha) + \frac{\alpha \tilde{k}_{t-1}}{\tilde{k}_{t-1}(G_t) + n} \lambda_{t-1}^P \qquad \text{with } \frac{\partial \lambda_t^P}{\partial \tilde{k}_{t-1}} = \frac{\alpha n}{(\tilde{k}_{t-1}(G_t) + n)^2} \lambda_{t-1}^P > 0.$$

However, since $\gamma \lambda_t^E + (1-\gamma)\lambda_t^P = 1$ $\forall t$, this implies also that $\partial \lambda_t^E / \partial \tilde{k}_{t-1} < 0$, which proves the directions of convergence. ∎

The following comparative statics results are useful for later reference:

Lemma 2 Everything else equal, the relative capital endowments of landlords λ^E adapt as follows to changes in the environment:

(i) $\partial \lambda_t^E / \partial a < 0$ and $\partial \lambda_t^E / \partial A < 0$, (ii) $\partial \lambda_t^E / \partial \gamma < 0$, (iii) $\partial \lambda_t^E / \partial n > 0$.

Proof The results follow from taking partial derivatives of condition (12), and because $\partial A / \partial a > 0$. ∎

From a certain point during the development process onward, economic development and public-goods provision are complements, in the sense that from a certain level of development onward, it is efficient to invest in infrastructure, where the efficient level of public-goods provision is denoted by G^*.

Lemma 3 G^* increases with the level of development.

Proof From the expression of average income, and the fact that marginal benefits of public good provision have to equal marginal costs of 1, G^* can be derived to be

$$G_t^* = \alpha^{1/(1-\alpha)}(A_t k_t)^{\alpha/(1-\alpha)} - \frac{A_t k_t + n}{A_t k_t}, \tag{13}$$

which can be shown to be strictly increasing in both A_t and k_t. ▪

Moreover, this implies the following result:

Lemma 4 There exists a unique generation $\underline{t}$: $G_t^* = 0\ \forall t < \underline{t}$ and $G_{t'}^* > 0\ \forall t' \geq \underline{t}$.

Proof The necessary condition for $G^* > 0$ is

$\alpha((1+G)A_t k_t)^{\alpha-1} A_t k_t \geq 1.$

The result follows since for low levels of development, namely, k and A, this condition is not (necessarily) satisfied, while monotonicity of development and hence growth in A and k ensures that there must be a unique generation t for which the condition eventually holds. ▪

What sort of public sector would be implemented prior to $\underline{t}$? Under oligarchy, the elite would produce without bothering to set up an infrastructure itself, whose marginal costs amount to $1/\gamma$. But even under democracy, no group, landlords or the landless, would endorse public-goods provision. The landlords, since it would be inefficient, and the landless, essentially for the same reason: implementing a scheme of direct redistribution would benefit them more. Intuitively, public goods complement technology in the production process, so if technology is not sufficiently advanced, the provision of public goods is not worthwhile. This leads to the following result:

Lemma 5 Public goods are provided only if it is overall efficient to provide at least some public goods.

Proof Net income of individual i from a purely redistributive state is $y^i + \tau(y - y^i)$. Hence, landless individuals $i \in P$ enjoy a net gain from redistribution since $y^P \leq y$, while landlords suffer a net loss. Now consider public-goods provision. For $t < \underline{t}$, the marginal benefit from providing public-goods provision is lower than the marginal cost, implying lower net individual income $y^i(G)(1 - \tau(G)) = y^i(G) - G$ for any individual i when a positive amount of G is provided, compared to $G = 0$. Thus, $G_t|_{t<\underline{t}} = 0$ under landlord oligarchy, as well as under democracy, since in that case the median voter prefers redistribution to public-goods provision. ▪

8.3.2 *Political Economy Equilibrium with Inequality*

We now turn to the analysis of the different possible transition regimes to democracy. First, consider the possibility that democratization arises when the disenfranchised landless are powerful enough to implement democratization by force, namely, when

$$\gamma k_t^E < (1-\gamma) k_t^P, \tag{14}$$

provided that it is in their interest to have a democracy.

Proposition 1 There is a unique generation $\hat{t}$: $\gamma k_t^E \lesseqqgtr (1-\gamma) k_t^P \Leftrightarrow t \gtreqqless \hat{t}$.

Proof The result follows directly from dividing condition (10) by k_t and applying lemma 1. ∎

This result reflects the possibility for a transition toward democracy under the threat of revolution, as studied by Acemoglu and Robinson (2001, 2003).

Since under democracy the decisive voter belongs to the group of people, to determine the equilibrium outcome under democracy we need to characterize the preferred levels of public-goods provision and redistribution by the people. The decision problem of the median voter, who happens to own no land, is as follows:

$$\max_{\{\tau, G, T\}} (1-\tau) y_t^P(G) + T \text{ s.t. } T + G - \tau y_t(G) \le 0, \quad \text{and} \quad \tau - 1 \le 0. \tag{15}$$

The solution of this problem implies the following results:

Proposition 2 The emerging democratic regime is characterized by $\hat{\tau} = 1$, $\hat{G} = G^*$ as $\partial y(\hat{G})/\partial \hat{G} = 1$, and, consequently, $\hat{T} = y(G^*) - G^*$.

Proof The results follow directly from the Kuhn-Tucker conditions of problem (15). ∎

This result illustrates that democratization implies full equalization of incomes.[13] Moreover, a democratic regime provides the efficient level of public good. To illustrate the impact of a change in the political regime, compare this to the level of public goods implemented by the elite under oligarchy. Note also that, even though landlords can decide autonomously about financing and providing public infrastructure G for themselves, by the very fact that G is a public good, they cannot exclude the landless from using that infrastructure and from benefiting in terms of income.

Without loss of generality, let us now consider a landlord oligarchy. While the landless have no influence on the creation and structure of a public sector, the elite, since public-goods provision is productive, face a trade-off between providing it themselves while retaining exclusive political power, and giving up power in exchange for a broader financial (tax) basis available for the public good. Even though starting from an oligarchic system in which the landlord elite monopolize all political power and strictly dismiss the possibility of releasing power, the members of the elite eventually change their mind and extend the franchise. This is shown in the following two results. The first one, which reflects arguments made by Lizzeri and Persico (2004), implies that eventually the elite benefit more from giving up political power in exchange for more efficient production, than from retaining power, regardless of its potential superiority in terms of conflict power. Let G_t^E denote the level of public goods the elite would provide if the landlords were to finance it fully by themselves, while the level of infrastructure alternatively provided under democracy with universal financing is G_t^*. Then we have the following result:

Proposition 3 There exists a unique generation $\check{t} \geq \underline{t}$:

$$y_t^E(G_t^E) - \frac{G_t^E}{\gamma} \gtreqqless y_t(G_t^*) - G_t^* \Leftrightarrow t \lesseqqgtr \check{t}. \tag{16}$$

Proof Note that for $t < \underline{t}$: $G_t = 0$, so the elite's income under democracy is strictly lower than under oligarchy, whenever $\tau > 0$, and (16) does not hold. However, also note that the marginal income gain of a member of the elite with respect to public good provision is positive. Consider now the elite's most preferred level of G under oligarchy, which is given by $G_t^E = \arg\max\left[y_t^E(G_t)\left(1 - G_t/\gamma y_t^E(G_t)\right)\right]$. This implies, given $t > \underline{t}$, a tax rate faced by a member of the elite of $\tau_t^E = G_t^E/\gamma y_t^E(G_t^E)$, which must be compared to the respective tax rate faced under democracy. Under democracy, the tax base is larger, so the taxes required to finance a given amount of public-goods provision is lower. Using the result of lemma 1, in the limit income inequality vanishes and every member of society earns the average income. In this situation, every member of the society would choose the optimal level of public good provision G^*, and there is no role for redistribution T. Condition (16) therefore becomes $y_t(G_t^E) - G_t^E/\gamma \leq y_t(G_t^*) - G_t^*$.

From optimality of G^*, it follows that in the limit this condition is always satisfied for any level of G_t^E. Hence, in the limit the elite are always better off under democracy, since the people have to contribute taxes to finance the public good. The monotonicity statement follows from the monotonicity of the convergence of incomes. ▪

Proposition 4 Apart from one special case, the provision of public goods under oligarchy is never efficient.

Proof Consider the problem of the elite for public-goods provision under oligarchy: $\max_{G^E} y_t^E - G_t^E/\gamma$. Noting that

$$y_t^P = y_t \left((1-\alpha) + \frac{\alpha \lambda_t^E (1+G) A_t k_t + n/\gamma}{(1+G) A_t k_t + n} \right) = y_t C_t^E(G),$$

the first-order condition of the elite's optimization problem reads

$$y_t'(G) \stackrel{!}{=} \frac{1}{\gamma} - y_t \frac{(C_t^E)'(G)}{C_t^E(G)}, \tag{17}$$

where

$$(C_t^G)'(G) \equiv \left(\lambda_t^E - \frac{1}{\gamma} \right) \left(\frac{\alpha A_t k_t n}{((1+G) A_t k_t + n)^2} \right). \tag{18}$$

Note that the sign of the last expression depends on the relationship between λ_t^E and $1/\gamma$. For the elite to provide an efficient level of public goods, it would have to hold that $\lambda_t^E = 1/\gamma$, which is a knife-edge result. ▪

Note that, in general, the elite underprovides public goods under oligarchy. This is, in particular, the case for a sufficiently large level of development, that is, as $\lambda_t^E \searrow 1$. The only scenario in which the elite would overprovide public goods is with low levels of development, where most capital is in the hands of the elite, who then would benefit disproportionately by public-goods provision. The last result implies predictions about the role of inequality for the efficiency of oligarchic political institutions that differ drastically from those under democracy. Democratic institutions display efficiency regardless of the distribution of resources, while the inefficiency of oligarchic institutions depends crucially on inequality.[14] In particular, the public-goods provision under oligarchic regimes declines with the degree of land inequality and increases with the degree of capital inequality. This reflects the different priorities of the elite under these scenarios. If the main source of income of the elite derives from natural resources, the elite has little incentives to provide public goods, while in contrast, incentives for public-goods provision are larger if the elite is more capitalistic.

As a corollary, it is possible to characterize the conditions under which a particular transition regime arises by referring to the timing of the respective transitions.

Corollary 1 An economy eventually democratizes and experiences a democratization "from above" when $t \nearrow \check{t} \leq \hat{t}$, while it experiences a democratization "from below" when $t \nearrow \hat{t} < \check{t}$.

An implication of this result is that a strong and entrenched elite is unlikely to propose a democratic transition. Given our assumption about technical progress, there will eventually be a democratic offer in this economy, which entails democratization "from above," if it is in the elite's own interest to extend the suffrage, or democratization "from below," if the people have become powerful enough to challenge the elite's political predominance. We can now analyze how the dynamics of economic and political development are affected by observable characteristics of the economy. In particular, consider the comparative statics of the timing and the consequential regime of a democratic transition with respect to technological progress a (or level of development A) by studying their impact on the crucial state variables, the stock of physical capital k available in the economy.

Proposition 5 Everything else being equal, faster technological progress a and a more industrialized structure of the economy implied by a higher level of development A lead to a sooner democratization.

Proof Note that, following lemma 2, $\partial\lambda^E/\partial a < 0$ and $\partial\lambda^E/\partial A < 0$, such that a transition from above is less likely. Moreover, consider the effects of a higher level of development on inequality to see that

$$\frac{\partial \lambda}{\partial A} = \frac{\partial}{\partial A}\left(\frac{\lambda^E}{\lambda^P}\right) = \frac{\partial}{\partial A}\left(1 + \frac{\alpha n/\gamma}{A(1+G)k + n}\right) < 0,$$

making a transition from below according to condition (10) more likely, that is, decreasing $\hat{t}$. Finally, note that A or a, respectively, increases the level of effective physical capital $\tilde{k}$, and hence increases the desirability of public-goods provision according to lemma 4, leading to a lower $\check{t}$, and therefore unambiguously to a sooner transition regardless of the transition regime. ▪

This result implies that if economic development is fast, oligarchies tend to be less stable and disintegrate sooner. This is in line with empirical findings by Boix and Stokes (2003) that economic development speeds up the arrival of democratization.

The previous discussion illustrates that different political regimes exhibit different features in terms of public-goods provision. In particular, democratic political institutions are efficient regardless the inequality in the society, while on the contrary the efficiency of oligarchic political institutions depends on inequality. Elites that are more entrenched in natural resources tend to provide fewer public goods than capitalistic elites. The transition from oligarchic political institutions to democratic institutions can either arise under the threat of conflict by the people, or be initiated by the elite. Finally, the transition to democratic institutions is facilitated by higher levels of development and a more dynamic technological environment. Taken together, these findings are recorded in the following:

Theorem 1 Democratic regimes are overall more efficient than oligarchies. Democratization is initiated by the people, that is, "from below," if and only if $\hat{t} < \check{t}$. Democratization is initiated by the elite, that is, "from above," if $\check{t} \leq \hat{t}$. Democratization happens earlier the larger A and a.

Proof The first claim directly follows from propositions 2 and 4. The second and third claims follow from propositions 1 and 3 and corollary 1. The last claim is a result of proposition 5. ▪

Several aspects of this result need to be put into perspective. First, the clear-cut results are obtained under the standard assumption of a pure (one head one vote) and direct (no party representation) democracy. This implies that the preferences of the median voter, who is a member of the people, are perfectly represented in public policies. In the current framework, the consequence of this assumption is full redistribution under the democratic regime and the efficient provision of public goods. In reality, however, the aggregation of preferences is heavily influenced by institutional arrangements that curb the possible choices that can be made by the electorate. Thus, the details of the institutional frame and the democratic structures can introduce a gap between de jure and de facto political power when the preferences of all groups in society, not only those of the median voter, are reflected in public policy. This observation is particularly relevant, if the preferences of different groups in society do not coincide. In the current framework, the pretax income of the elite is still larger than that of the poor, possibly inducing different preferences about redistribution issues. A question one might want to ask in this context is whether the elite, when contemplating offering democracy, may try to influence the design of democratic institutions in such a way as to limit redistribution. For example, consider the possibility that the elite can, prior to voting, restrict the policy space by imposing an upper limit to unproductive redistribution $\overline{T}$. The median voter's maximization problem (15) has to be modified by respecting an additional constraint that $T \leq \overline{T}$. In this case, denote the optimal choice of public-goods provision selected by the median voter by G_t^P, which may in general differ from G_t^*, depending on $\overline{T}$. Knowing this, the question then becomes whether the elite would find it profitable to impose a binding limit, namely, $\overline{T} \leq y(G_t^*) - G_t^*$. To address this question, note that efficient public-goods provision under democracy implies full redistribution and thus a net income of $y(G_t^*) - G_t^*$ for a member of the elite (and likewise a member of the people). Compared to that, if $\overline{T} \leq y(G_t^*) - G_t^*$, an elitist individual could earn an income of

$$y_t(G_t^P, \overline{T})C_t^E\left(1 - \frac{G_t^P + \overline{T}}{y_t(G_t^P, \overline{T})}\right) + \overline{T},$$

where

$$C_t^E \equiv \left[(1-\alpha) + \frac{\alpha(1+G_t^P)A_t k_t \lambda_t^E + \alpha\dfrac{n}{\gamma}}{(1+G_t^P)A_t k_t + n}\right], \tag{19}$$

when limiting redistribution to $\overline{T}$. Simplifying these expressions, the elite will prefer to limit redistribution to $\overline{T} < y_t(G_t^*) - G_t^*$ if

$$C_t^E(y_t(G_t^P) - G_t^P - \overline{T}) + \overline{T} > y_t(G_t^*) - G_t^*. \tag{20}$$

In the current framework, it is not possible to identify the conditions under which this condition holds since G_t^P cannot be determined analytically. Nonetheless, due to monotonicity of C_t^E in λ_t^E and n/γ, it is possible that for a sufficiently high level of inequality in income, land endowments, or both, the condition is satisfied. In this case, the elite would actually benefit from imposing a limit on redistribution if they could.[15] Note also that under this scenario the democratization process would arise at the same time or even earlier than under the benchmark case, since the opportunity cost for initiating a democratic transition is lower for the elite. Taken together, while the main results on democratic transitions stated earlier remain unaffected, the precise framework of political institutions adopted under democracy may actually depend on the transition and the level of inequality at the time of the transition.

Another noteworthy issue in this respect is the double cost the elite face during the transition to democracy. On the one hand, members of the elite have to pay taxes, which partly go to finance redistributive transfers. On the other hand, the elite lose the power to decide about issues such as public-goods provision, since the residual decision rights are taken on by the median voter under democracy. The case of democratization "from below" shows that the latter fact constitutes the real cost of democratization, while taxes and implicit redistribution in terms of proportional taxation are only second-order effects. Rather, taxation can very well be to the elite's advantage since universal taxation under democracy allows them to share the costs of public-goods provision. This is illustrated in particular when the elite voluntarily offer full redistribution in order to achieve an efficient level of public-goods provision. Finally, land resources make no clear prediction on the timing of democratization, while the size of the elite seems to be ambiguous with respect to the transition regime. The following section briefly discusses these implications in the light of historical evidence.

8.4 The Historical Context and Empirical Relevance

The predictions of our model are broadly in line with historical and empirical evidence. As a first attempt to substantiate this claim, we test whether the three main implications stand out in the cross-country data set collected by Persson and Tabellini (2003, 2004).[16] The predictions are (a) that countries with more democratic structures are better developed, which is a direct consequence of lemma 1, and propositions 1 and 3; (b) that countries with more democratic institutions have a larger state, which follows from corollary 1, lemma 3, and proposition 2;[17] and (c) that more democratic countries redistribute relatively more, which follows again from proposition 2. Since the data set is confined to democracies of different quality of democratic institutions, and contains no nondemocracies, we use variation in the quality of democratic institutions to test the model implications. In particular, we test the three predictions by testing the correlation of an index of the quality of democratic institutions with (a) log per capita income, denoted *lyp*, (b) the ratio of central government spending (including social security) as percentage of GDP, denoted *cgexp*, and (c) social security and welfare spending as percentage of GDP, denoted *ssw*, respectively. To do this, we run a system of seemingly unrelated regressions to account for endogeneity and potential error correlation across equations. The quality of democratic institutions, namely, the extent of democratization, is measured using the *polityIV*-index, which assigns to each country an integer score ranging from −10 to 10 with higher values associated with better democratic structures. Alternatively, and as a robustness test, we use the *gastil*-indices of political rights and civil liberties, which range from 1 to 7 with lower values associated with better democratic institutions. In the regressions, we control for several alternative factors potentially influencing the level of development, and for size and structure of the public sector.[18]

For reasons of brevity, table 8.1 contains only the coefficient estimates for the index of democratic quality. Detailed results for the other coefficients are available upon request. The first set of regressions displays a significant positive coefficient of *polityIV* on all three dependent

Table 8.1
Empirical estimates

	Dependent variable		
	lyp	cgexp	ssw
polityIV	0.040*	0.874*	0.462*
	(0.018)	(0.437)	(0.222)
N	58	58	58
R^2	0.935	0.771	0.858
gastil	−0.146*	−2.905*	−1.976**
	(0.060)	(1.336)	(0.659)
N	59	59	59
R^2	0.934	0.776	0.868

Notes: Results from SURE estimations. Other explanatory variables are a constant, *lpop*, *loga*, *prop*1564, *prop*65, *trade*, *yrsopen*, *oecd*, *avelf*, *federal*, *con*2150, *con*5180, *con*81, *age*, *latitude*, *africa*, *asia*, *laam*, *col_uka*, *col_espa*, and *col_otha*. In the equations for *cgexp* and *ssw*, additional explanatory variables are *lyp*, and *pres*, *maj*. See also note 18. Standard errors are in parentheses. Significance at 5 percent and 1 percent level is indicated by * and **, respectively.

variables, indicating that more democratic institutions are associated with higher levels of development as measured by GDP per capita, and larger and more redistributive states.[19] We refrain explicitly from interpreting these correlations as causal, since, in the light of the theoretical results obtained before, political institutions can be expected to endogeneous, at least in the long run. These correlations nevertheless indicate, however, that the three theoretical implications are not refuted by a first glance at the data. The results are also robust with respect to the measure of democratic quality, as indicated by the results using *gastil* as a measure. Again, better democratic institutions are associated with higher incomes, and with larger and more redistributive states. Further tests reveal that the hypothesis that democratic institutions have no impact on any of the dependent variables can be rejected at the 1 percent level. Without a doubt, more empirical work would be required to rigorously test the implications of the theory, and to show that these are in line with the historical experience. However, the concepts of democratization "from above" and "from below" are difficult to implement in an empirical analysis, since creating a respective dummy for each country, for example, would eventually always imply a somewhat subjective judgment. Nevertheless, from the results presented so far we conclude that the model by and large fits the historical facts.

The main implications of the model presented in section 8.3 also correspond well with the findings of earlier empirical contributions. Recently, the interdependencies of democratization, sustainability of democracy, and economic development have received a revived research interest among economists and political scientists. Some contributions found that economic development apparently serves to stabilize democratic systems (see Przeworski and Limongi 1997; Przeworski et al. 2000). Other empirical evidence seems to indicate that there is a positive causal effect of economic development on the probability that a country democratizes as well as a positive effect of development on the stability of democracies (see Barro 1999; Boix and Stokes 2003). Economic development, reflected by technological progress, can be seen as the driving force behind democratic transitions in the model presented earlier. Recent empirical findings cast doubt on the causal effect of income on democracy, however. Using variation across countries and over time, Acemoglu et al. (2005) find no evidence for a causal effect when controlling for country fixed effects representing factors that affect both economic and political development simultaneously. The model presented before is also consistent with these findings, because the ultimate determinant of whether and how a country democratizes is inequality in resources, rather than the mere size of the economy. Economic development affects the distribution of factors and incomes, and, through this, drives political development both in terms of the occurrence of a democratic transition as well as its type. Acemoglu and Robinson (2004, chap. 2) provide a survey of the cross-country evidence on the patterns of democracy. Their main conclusions are that richer countries are more likely to be democratic and that democracies are more redistributive than oligarchies with an increase in redistribution following democratization. All these facts are in line with the theoretical predictions. In particular, richer countries, which are on a higher level of technological and therefore economic development, experience a democratic transition sooner. Moreover, the model implies that redistribution primarily arises under democracy—in particular, if the extension of the franchise happened under the pressure of conflict from the formerly disenfranchised.

The model is also compatible with the findings by Glaeser et al. (2004), that cast doubts on the view that political institutions, in particular democratization, are a prerequisite for growth, but rather advocate the idea that human and physical capital accumulation are the engines of development, which in turn facilitates the improvement of crucial institutions such as property rights protection. In fact, the previous model suggests that democratization, rather than being a prerequisite, is a consequence of development, but might be crucial for the adoption of growth-enhancing policies, such as the provision of public goods like schools or police. Evidence on the growth effects of democratization by Minier (1998) and Tavares and Wacziarg (2001) is in line with this view. Their findings suggest that the advantage of democracy mainly materializes in terms of more efficient accumulation and use of human and physical capital under democratic rule. This view is also in line with other evidence that suggests that economic development, together with the political institutions in place, determine the size of the public sector as well as its structure in terms of infrastructure and transfers, such as unemployment benefits, health care, and retirement pensions (see Boix 2001). He also finds that public sectors are significantly larger under democratic than under oligarchic regimes. The main predictions of our model are also supported by the findings of Boix (2001, 2003). Different regressions of the size of the public sector with a large panel data set of countries reveal not only a larger public sector in democracies, but in particular a negative effect of the size of the share of the agricultural sector in total production as well as a positive effect of per capita income on the size of government (see Boix 2001, table 3). The differences between public policies implemented under democratic and nondemocratic regimes are still not fully settled in the literature, however, as a recent paper by Mulligan, Gil, and Sala-i-Martin (2004) documents. Their findings suggest that public policies are not so different among democracies and nondemocracies. According to their findings, the main differences lie rather in policies that affect competition for political power. On the other hand, the model provides a rationalization for the findings by Rodrik, Subramanian, and Trebbi (2004) and Rigobon and Rodrik (2004) in that democratic regimes adopt better institutions and policies, and therefore exhibit larger incomes, while good institutions and democratic structures themselves are more likely in more developed economies.

In a more historical context, Alesina and Glaeser (2004, chap. 2) argue that the U.S. and European countries experienced entirely different transitions toward democracy. With regard to the transition scenario corresponding to the theoretical model here, the United States as well as the United Kingdom could be seen as having experienced democratic transitions "from above," which were mainly initiated and determined by ruling classes and landowners; see also the description in Lizzeri and Persico (2004). On the other hand, France, and also Germany, experienced transitions that were characterized by substantial pressure on the elites to extend the franchise.[20]

Some important issues remain. After their critique of the work of Przeworski and Limongi 1997 and Przeworski et al. (2000), Boix and Stokes (2003) claim that the most puzzling yet unanswered questions regarding the links between economic development and democratization concern the findings of different effects of economic development on the propensity to democratize in different historical contexts. In particular, they raise the question why early during industrialization oligarchies appear to have been less stable to democratization than at later stages of development, and why economic development after World War II seems to have mainly helped stabilize democracies rather than induce democratization of nondemocratic regimes. According to our model, democratization can occur under very distinct scenarios, and countries that democratized early are likely to differ in several other respects than the level of economic development from those economies that democratized later, or have not democratized yet altogether. In particular, the distribution of power among oligarchs and the proletariat might differ substantially due to different capital accumulation histories, land resources, and group sizes, thus making comparisons only in terms of level of development difficult. While these issues deserve further investigation, the model can rationalize these empirical observations.

8.5 Discussion

Before concluding, we briefly discuss the implications of the simplifying assumptions we make. The adoption of a subsequent generation framework with a joy-of-giving representation of individual utility gives rise to myopic behavior as the members of a given generation do not internalize the consequences of their choices on subsequent generations. This appears to be a reasonable assumption for considering long-run processes as done in this analysis. Moreover, introducing forward-looking agents would complicate the analysis, while not changing the main results, as long as some discounting is incorporated.

The adoption of exogenous technological progress is of no relevance. Introducing endogenous innovations, with physical or human capital representing the engine for growth, could be accomplished without altering the main findings. The assumption that technological progress is not neutral with respect to all factors of production, and augments capital rather than land, is, together with the assumption of logarithmic utility and the accumulation of capital through bequests, crucial for the result that inequality declines with the level of development and vanishes in the limit. This result also implies that eventually democracy emerges. Assuming heterogeneous bequest behavior in the society, with the elite bequeathing relatively less capital due to their bequests of natural resources as was the case historically would tend to strengthen our results. On the other hand, if the elite would transfer a larger fraction of their incomes to their offspring, income might not converge fully, but the income share generated from land ownership would still decrease over the course of generations, which is sufficient for our results as long as the provision of public goods exclusively by the elite is sufficiently costly, or if that were not possible at all.

Further issues resulting from the fact that technological progress is also not neutral with respect to the relative position of the two groups in society are left aside in the current analysis. Endogenizing technological progress also invites questions concerning the political aspects of technology adoption. On the one hand, as studied by Cervellati and Fortunato (2004), the elite could resist innovation in order to block structural change and the income loss associated with modernization. In that model, the degree of resistance to innovation is not constant but declines with the level of inequality and development, and disappears eventually. Similarly, the elite might block modernization to delay or avoid democratic transitions and the associated loss of political power (see Mokyr 2000).

Following the literature on democratic transitions initiated by the elite—for example, Bourguignon and Verdier 2000, Lizzeri and Persico 2004, and Gradstein 2006—a democratic transition can deliver a gain in efficiency for society as a whole, but in particular for the elite. In our model, this is represented by the possibility of providing a public good at lower costs under democracy than under oligarchy. In a more general framework, this efficiency improvement under democracy could also be obtained as equilibrium outcome. A related line of research (Cervellati, Fortunato, and Sunde 2004), explores the possibility that the higher efficiency of democratic systems arises in the form of a self-sustained social contract as the outcome of a strategic game of rent-seeking investments played between and within different social groups.

In terms of fiscal policy, we consider the possibility for progressive redistribution under democracy, but exclude taxation and regressive redistribution under oligarchy. This is done for simplicity, since the key feature of the analysis is that democracy is *relatively* more progressive, reflecting an opportunity cost for democratization on part of the elite. Also for simplicity, taxation implies no distortions, which leads to corner solutions for taxation in the voting game, namely, full redistribution under democratization. Finally, and in line with previous contributions like Acemoglu, Johnson, and Robinson 2004, we consider a simple conflict technology, a "guns model." The outcome of a conflict is deterministic, implying that (potentially costly) conflict never arises in equilibrium, such that democratic transitions essentially arise under a shadow of conflict. Also, we do not consider incentives for triggering violent conflict and endogenous investments in conflict activity, as is done, for example, by Grossman (2001) or Esteban and Ray (1999). While the role of each of these assumptions is clear, a simultaneous consideration of potential for regressive and distortionary redistribution, together with a conflict game with endogenous effort choice, would require a more in-depth analysis. This is the case, since both the dynamic evolution of inequality and the relative conflict potential of the different groups of society would then depend on the particular specification adopted. While the main mechanism would still work, the joint consideration of these elements appears to be of particular interest in light of the analysis of the structure of the public sector under the different transition regimes, and of the precise conditions under which they may arise.

8.6 Conclusion and Future Research

We conclude by summarizing the focus and the main results of the current analysis, which represents a first step in modeling the interrelation between development and political institutions. We have provided a simple dynamic model of economic and political development that is able to reproduce several recent theories about the endogenous transition toward democracy and the determinants of the design of constitutions within a single framework. The main mechanism implies that economic development is a prerequisite for a democratic transition. Moreover, depending on the economic environment, this transition is triggered either by the ruling elite in the initial oligarchy, leading to a democratization "from above," or by the initially disenfranchised people, whose threat to go to open conflict and mount a revolution initiates a democratic transition "from below." The results also indicate that in equilibrium democratic political institutions are generally more efficient than oligarchies, and arise sooner the faster the process of economic development.

To make our argument as transparent as possible, we provided a model, which is stylized in several respects. We close by providing a discussion of the limitations of this analysis. The deeper investigation of these issues constitutes our agenda for future research. First, we categorized democratic transitions into the classes "from above" and "from below," which have the advantage of highlighting the main mechanism. However, this conceptualization is not trivial to implement empirically in the sense that it is difficult to exactly categorize historical democratization processes into these two classes. The next step on the research agenda therefore is to provide a taxonomy characterizing the precise conditions under which the different types of transition occur. This would deliver theoretical predictions which then could be tested empirically.

Another limitation of the current framework is that, in order to emphasize the role of the transition, the analysis concentrated on the main functions of the state. The constitutional details and political institutions, such as voting systems and so forth, that are instrumental for these outcomes have not been studied, however.[21] Exploring the relationship between the type of democratic transition and the constitutional details implemented in its occurrence may deliver further insights into the observed differences in political institutions and economic performance across countries. In a similar vein, for simplicity we adopted a reduced-form view of the allocation of decision power about constitutional details. A further investigation of the forces leading to different bargaining positions in the phase of constitutional design would shed more light on the different institutional aspects fixed in different institutions and could potentially deliver more precise empirical predictions about the actual policies to be expected to prevail in different systems: for example, the size of the public sector, type of public goods provided, kind of redistribution schemes in place, progressivity of fiscal systems, and so forth. Future work should therefore analyze the dynamic emergence of political institutions and the precise channels through which the particular structure of the public sector is obtained, to open what essentially remains a black box in the current analysis.

In order to highlight the impact of the democratization process on institutional design, we concentrated on "bang-bang" democratization episodes, with only two possible states, oligarchy and universal franchise. In reality, democratic transitions are generally gradual processes of the extension of the franchise. The consideration of the forces behind gradual extensions might deliver additional insights. In particular, democratic extensions can be thought of as reducing conflict pressure without the need to give away more political influence than necessary. Also, it is likely that the institutions emerging from a gradual extension depend on the actual path of democratization. For example, different public goods may be provided depending on the preferences and needs of the groups of society that are newly enfranchised.

Also, the consideration of more than two groups of individuals may change the predictions of the model. On the one hand, questions about cohesion, group formation, and coalitions could arise. On the other hand, this modification together with the potential for gradual extensions of the franchise would provide a much richer picture of democratization experiences.

Finally, to highlight the feedback among economic development, democratization, and institutional arrangements, we concentrated on two very distinct forces behind democratization, democratization in the interest of the elite and democratization under the threat of conflict. Other forces behind democratic transitions, such as enlightenment views emphasizing the role of liberal or emancipative values and the spread of knowledge, the emergence of an economically important but politically disenfranchised middle class, or issues of competition and conflicting interests within the ruling elite, have been neglected for simplicity. While we believe that these issues could be addressed and would enrich the picture of democratic transitions and their effects on arising institutions, we also believe that the main idea of the current model could be preserved.

Notes

The authors would like to thank David de la Croix, Matthias Doepke, Theo Eicher, Cecilia Garcia-Penalosa, Andreas Leukert, Omar Licandro, Ken Sokoloff, Davide Ticchi, and Gerald Willmann, as well as seminar participants at the CESifo Venice Summer Institute 2004; ESEM/EEA 2004 Meetings in Madrid; SIE 2004, Bologna; VfS Conference 2004, Dresden; and ASSET 2004, Barcelona, for helpful comments and discussions. The usual disclaimer applies. Financial support from IZA is gratefully acknowledged.

1. This "social conflict" view is also emphasized in Acemoglu, Johnson, and Robinson 2005.

2. See, for example, Gradstein 2004 for a dynamic model of the emergence of property rights protection in the context of economic development.

3. An argument for why the extension of the franchise can be in the interest of those holding oligarchic power, which builds on an internal struggle between members of the elite, can be found in Llavador and Oxoby 2005. In Ticchi and Vindigni 2004b, the extension of the franchise can be in the interest of the elite can if this helps in motivating the masses to provide effort in wars.

4. Aghion, Alesina, and Trebbi (2004) study the optimal constitutional size of a minority required to block legislation, and conversely the size of the supermajority needed to pass legislation. Ticchi and Vindigni (2004a) address the choice between a majoritarian and a consensual democracy made by a rich elite and show that the elite is more likely to choose a majoritarian democracy the larger the (exogenous) degree of inequality. These studies, which are inherently static in nature, provide a deeper economic understanding of the reasons for the emergence of different political institutions.

5. This formulation of the utility function is not crucial for the main insights but simplifies the analysis considerably. It is noteworthy, however, that the development dynamics of the economy, as shown later, essentially depend on the distribution of factor endowments, and hence the decision on consumption and bequest, which in reality may differ across different groups of the society.

6. For simplicity, assume that there is no market for land, so no land is ever traded. This assumption is without loss of generality. In fact, as will become clear here, due to the tim-

ing of events even allowing for land markets would not change the results since land markets are implicitly included in the model through the rents land generates.

7. Endogenizing the rate of technical progress would not affect the main argument.

8. Also Acemoglu and Robinson (2003) rely on the specification used for the production technology in equation (2).

9. One could argue that also in reality, by projecting future budgets, democratically elected governments adhere to a similar reverse timing with respect to production, taxation, and spending the tax revenues on redistribution or public goods, which affect the production process itself.

10. We exclude the possibility of discretionary extension of suffrage to particular persons, and assume that it can only be done regarding entire groups. In other words, apart from land ownership, there is no potentially "unobservable" heterogeneity of individuals.

11. Including land resources as a means of generating conflict potential analogously to physical capital would lead to identical results.

12. Note that this description of timing is without loss of generality. In particular, the same timing holds under democracy, where the "ruling elite" consists of all people populating the economy. Under the current setting, the median voter in equilibrium does not opt for oligarchy, for example, of the people, since expropriation of land resources is prohibited, and democracy provides the cheapest way to finance public good provision. See also the discussion that follows.

13. Just as the landless established a redistribution scheme that allows them to appropriate equally all land rents and capital rents accruing to the landowners under democracy, we could alternatively allow for expropriation of factor endowments—in particular, land and capital. We do not allow for discriminatory taxation.

14. This result is driven by the assumption of nondistortionary taxation. With distortions, both efficiency of public good provision and redistribution would depend on inequality under democracy. Nevertheless, the results on *relative* efficiency and redistribution of oligarchy and democracy would still hold.

15. The possibility of imposing restrictions requires further assumptions and arguments about the conditions under which this possibility arises, and to what extent restrictions can be implemented.

16. The data used for the analysis are a cross-country data set for eighty-five countries. See Persson and Tabellini 2003 and 2004 for a detailed description of the data, the variables, and the data sources. The data are available at http://rincewind.iies.su.se/perssont/datasetselectoralrules.htm.

17. Note that this result is driven by the fact that a higher level of development implies both a higher G_t^*, irrespective of the political system, and a higher likelihood of democracy, which in turn implies efficiently high G_t^* as well as a higher level of redistribution T_t.

18. In particular, we include log of population size (*lpop*), population structure (the shares of population in working age and retirement age, *prop*1564 and *prop*65, respectively), the sum of exports and imports as share of GDP (*trade*), log total factor productivity (*loga*), an index of openness to trade measuring the fraction of years during 1954–1990 during which a country was open to trade (*yrsopen*), economic institutions fostering

development (*oecd*), ethno-linguistic fragmentation (*avelf*), federal state structures (*federal*), constitutional inertia measured by dummies (*con*2050, *con*5180, *con*81) to account for constitutional fashions and historical context, and the age of each democracy as defined by the fraction of the last two hundred years of uninterrupted democracy (*age*). To control for unobservable influences of geographic location, we control for latitude as measured by absolute distance from equator *latitude* and introduce dummies for continental location *africa* for Africa, *asia* for Eastern and Southern Asia, and *laam* for Southern and Central America including the Caribbean. Historical factors are taken into account by controlling for colonial history (*col_uka*, *col_espa*, *col_otha*). All these variables, together with the respective index for democratic quality, are used as explanatory variables. The equations for size of public sector (*cgexp*) and the size of social redistribution (*ssw*) contain as additional explanatory variables the level of development (*lyp*) as well as indicators for majoritarian voting (*maj*) and presidential systems (*pres*). This takes account of the results found by Persson and Tabellini (2004) that details of democratic institutions are likely to have an impact on public policy.

19. This result is robust with respect to different specifications.

20. While the French democracy essentially goes back to the revolution of 1789, extension of the franchise in Germany was associated with several waves of social unrest, as was the case for the revolution in 1848, the socialist movement that led Otto von Bismarck to introduce the welfare state, and the revolution in 1919 to mention just the most prominent milestones of the transition.

21. This is done by the literature studying the economic effects of constitutional and institutional details mentioned in section 8.1.

References

Acemoglu, D., S. Johnson, and J. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91, no. 5: 1369–1401.

Acemoglu, D., S. Johnson, and J. Robinson. 2005. "Institutions as the Fundamental Cause of Long-run Growth." Forthcoming in *Handbook of Economic Growth*, ed. P. Aghion and S. Durlauf. Amsterdam: Elsevier.

Acemoglu, D., S. Johnson, J. Robinson, and P. Yared. 2005. "Income and Democracy." Mimeo., MIT.

Acemoglu, D., and J. Robinson. 2000. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics* 4: 1167–1199.

Acemoglu, D., and J. Robinson. 2001. "A Theory of Political Transitions." *American Economic Review* 91, no. 4: 938–963.

Acemoglu, D., and J. Robinson. 2003. "A Theory of Political Development." Mimeo., MIT.

Acemoglu, D., and J. Robinson. 2004. *Economic Origins of Dictatorship and Democracy*. Unpublished book manuscript. Available at http://web.mit.edu/daron/www.

Aghion, P., A. Alesina, and F. Trebbi. 2004. "Endogenous Political Institutions." *Quarterly Journal of Economics* 2: 565–612.

Alesina, A., and E. L. Glaeser. 2004. *Fighting Poverty in the U.S. and Europe: A World of Difference*. New York: Oxford University Press.

Barro, R. J. 1999. "Determinants of Democracy." *Journal of Political Economy* 107, no. 6: S158–S183.

Boix, C. 2001. "Democracy, Development, and the Public Sector." *American Journal of Political Science* 45, no. 1: 1–17.

Boix, C. 2003. *Democracy and Redistribution*. Cambridge: Cambridge University Press.

Boix, C., and S. C. Stokes. 2003. "Endogenous Democratization." Mimeo., University of Chicago.

Bourguignon, F., and T. Verdier. 2000. "Oligarchy, Democracy, Inequality and Growth." *Journal of Development Economics* 62, no. 2: 285–313.

Cervellati, M., and P. Fortunato. 2004. "Resistance to Reforms, Inequality and Development." Mimeo., Universitá di Bologna.

Cervellati, M., P. Fortunato, and U. Sunde. 2004. "Hobbes to Rousseau: Inequality, Institutions, and Development." IZA discussion paper 1450.

Cervellati, M., and U. Sunde. 2005. "Human Capital, Life Expectancy and the Process of Development." *American Economic Review* 25, no. 5: 1653–1672.

Esteban, J., and D. Ray. 1999. "Conflict and Distribution." *Journal of Economic Theory* 87: 379–415.

Galor, O. 2005. "From Stagnation to Growth: Unified Growth Theory." Forthcoming in *Handbook of Economic Growth*, ed. P. Aghion and S. Durlauf. Amsterdam: Elsevier.

Galor, O., and O. Moav. 2002. "Natural Selection and the Origin of Economic Growth." *Quarterly Journal of Economics* 117, no. 4: 1133–1192.

Galor, O., O. Moav, and D. Vollrath. 2004. "Land Inequality and the Origin of Divergence and Overtaking in the Growth Process: Theory and Evidence." Mimeo., Brown University.

Galor, O., and D. N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90, no. 4: 807–828.

Glaeser, E., R. LaPorta, F. Lopez-De-Silanes, and A. Shleifer. 2004. "Do Institutions Cause Growth." *Journal of Economic Growth* 9: 271–303.

Gradstein, M. 2004. "Governance and Growth." *Journal of Development Economics* 73, no. 2: 505–518.

Gradstein, M. 2006. "Inequality, Democracy, and the Emergency of Institutions." Forthcoming in *Economic Journal*.

Grossman, H. 2001. "Make Us a King: Anarchy, Predation, and the State." *European Journal of Political Economy* 18: 31–46.

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, no. 1: 83–116.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization* 15, no. 1: 222–279.

Lizzeri, A., and N. Persico. 2004. "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform.'" *Quarterly Journal of Economics* 119, no. 2: 707–765.

Llavador, H., and R. Oxoby. 2005. "Partisan Competition, Growth and the Franchise." *Quarterly Journal of Economics* 120, no. 3: 1155–1185.

Mauro, P. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110, no. 3: 681–712.

Minier, J. A. 1998. "Democracy and Growth: Alternative Approaches." *Journal of Economic Growth* 3, no. 3: 241–266.

Mokyr, J. 2000. "Innovation and Its Enemies: The Economic and Political Roots of Technological Inertia." In *A Not So Dismal Science*, ed. M. Olson and S. Kähkönen, 61–91. Oxford: Oxford University Press.

Mulligan, C. B., R. Gil, and X. Sala-i-Martin. 2004. "Do Democracies Have Different Public Policies than Nondemocracies?" *Journal of Economic Perspectives* 18, no. 1: 51–74.

Papaioannou, E., and G. Siouraounis. 2004. "Democratization and Growth." Mimeo., London Business School.

Persson, T., and G. Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

Persson, T., and G. Tabellini. 2003. *The Economic Effect of Constitutions*. Cambridge, MA: MIT Press.

Persson, T., and G. Tabellini. 2004. "Constitutional Rules and Fiscal Policy Outcomes." *American Economic Review* 94, no. 1: 25–45.

Przeworski, A., M. Alvarez, J. A. Cheibub, and F. Limongi. 2000. *Democracy and Development: Political Institutions and Well-being in the World, 1950–1990*. New York: Cambridge University Press.

Przeworski, A., and F. Limongi. 1997. "Modernization: Theories and Facts." *World Politics* 49: 155–193.

Rigobon, R., and D. Rodrik. 2004. "Rule of Law, Democracy, Openness and Income: Estimating the Interrelationships." NBER working paper 10750.

Rodrik, D., A. Subramanian, and F. Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9: 131–165.

Tavares, J., and R. Wacziarg. 2001. "How Democracy Affects Growth." *European Economic Review* 45, no. 8: 1341–1378.

Ticchi, D., and A. Vindigni. 2004a. "Endogenous Constitutions." Seminar paper 726, Institute for International Economic Studies, University of Stockholm.

Ticchi, D., and A. Vindigni. 2004b. "On Wars and Political Development. The Role of International Conflicts in the Democratization of the West." Leitner working paper 2004-02.

9 The Road from Agriculture

Thorvaldur Gylfason and
Gylfi Zoega

The great economist Arthur Lewis emphasized the distinction between traditional agriculture and urban industries. In his view, savings and investment originate solely in the latter, while vast pools of under-utilized labor can be found in the traditional sector (Lewis 1954). In this chapter we aim to fill a gap in his analysis by constructing a model of rational behavior in the traditional sector. We want to think of farmers as rational agents and so explain economic backwardness not in terms of history or mentality but rather in terms of a model with maximizing behavior. Our aim is to show that the level of technology in agriculture in each country will not, in general, coincide with the "frontier" technology of the most advanced economy. In particular, each country has an optimal "technology gap" that separates it from the frontier. In our analysis, the size of this gap turns out to depend on factors that are exogenous to most economic models and seldom subject to change, such as farm size reflecting geography, the fertility of the land, the ability of farmers to digest and take on new technologies, and the rate of time preference. Most surprisingly, perhaps, the distance from the technology frontier turns out to depend on the position of the frontier itself; the more advanced the frontier technology, the larger the optimal distance that maximizes the value of land from the frontier. We bring cliometric evidence from our native Iceland to bear on this issue. Further, we attempt to quantify the relationship between structural change and growth by considering the change in the share of agriculture in value added and of migration to cities as independent determinants of economic growth within a cross-country growth regression framework.

The share of agriculture in employment and value added has fallen relentlessly around the world over the past hundred years. Until the end of the nineteenth century, an overwhelming part of the workforce was engaged in agriculture everywhere. In 1960, almost half the labor

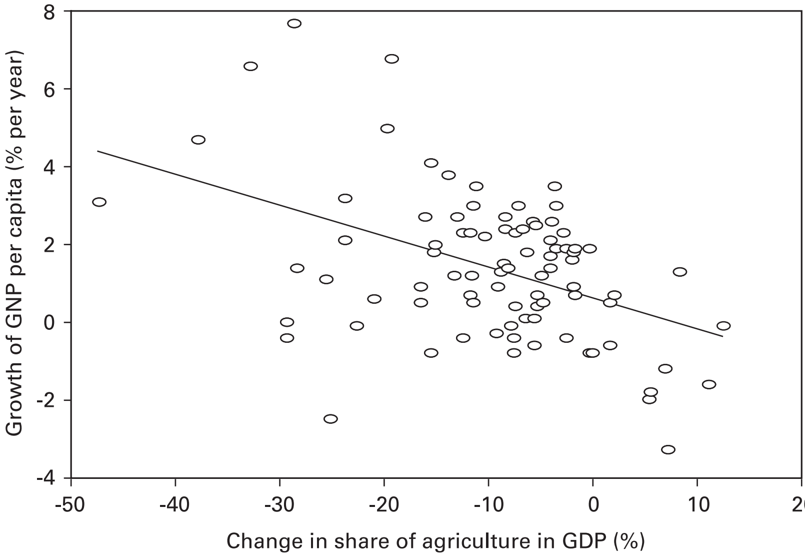


Figure 9.1
Structural change and growth, 1965–1998.

force in low-income countries was still employed in agriculture, but this ratio continues to fall: today almost a fourth of the labor force in low-income countries works on the land, less than 10 percent in middle-income countries, and less than 2 percent in high-income countries. To illustrate the relationship that motivates this study, we show in figure 9.1 data from eighty-six countries, some rich and some poor, in the period from 1965 to 1998.[1]

The figure shows the relationship between per capita economic growth along the vertical axis and structural change as measured from right to left along the horizontal axis by the decrease in the share of agriculture in value added from 1965 to 1998. Each country is represented by a single dot in the figure: the average growth rate over the sample period and the structural change from the beginning to the end of the period. The figure shows that a decrease in the share of agriculture by 13 percentage points from one country to another is associated with an increase in annual per capita growth by 1 percentage point.[2]

In a recent study, Temin (1999) argues that a relationship similar to that in figure 9.1 can account for the growth performance of fifteen European countries over the period 1955–1995. In particular, he argues that the migration of labor from rural to urban areas helps explain the postwar "Golden Age" of European economic growth, including the differences in growth rates during this period and the end of the high-growth era in the early 1970s.[3] Not all countries have handled this dramatic transformation of their economic structure as well. In extreme cases, the development was actively resisted, as witnessed originally by the institution of slavery that in some places continued well into the second half of the nineteenth century. The resistance to change took other, milder forms as well: for example, farmworkers in Iceland were throughout the nineteenth century prevented by law from leaving their employers, a form of serfdom that significantly delayed the transformation of the Icelandic economy from agriculture to industry.

This chapter adds to an expanding literature on the long-run sectoral implications of economic growth.[4] While we emphasize endogenous technological adoption at the farm level, other contributions have emphasized human capital accumulation. Galor and Moav (2003) model the transition from a rural agricultural society to an urban industrial society by showing how the complementarity of human and physical capital in industry generates an incentive for industrialists to support educational reforms. Human capital accumulation also plays an important part in the transition in Tamura 2002. In Galor and Weil 2000, skill-biased technical progress raises the rate of return on human capital, which causes human capital to grow, hence creating steady-state growth. Jones (1999), in contrast, argues that increasing returns to the accumulation of technology and labor sustains growth. We do not dispute the importance of human capital for the transition but, instead, want to describe some of the determinants of endogenous technological adoption in agriculture.

We argue that the extent of the transition from an agrarian to an industrial economy depends not only on the access of industrial producers to unlimited supplies of rural labor (Lewis 1954) and on productivity developments and availability of work in urban areas (Kaldor 1966; Harris and Todaro 1970), but also on farm size reflecting geography, the fertility of land, and the ability of farmers to adopt new technology. In this we are perhaps in part motivated by the experience of Iceland, an island in the far North Atlantic where agriculture was the main economic activity for centuries, supporting a population that lived on the margins of subsistence. Harsh climate, unfertile soil, small disparate plots of arable land, and a population not familiar with foreign cultures or languages hampered economic development for almost a thousand years. It is difficult to conceive of any form of institution building that could have helped inject dynamism into the agricultural economy.

9.1 Efficiency Gains in Agriculture and Growth

In this section we describe the behavior of farmers with regard to the adoption of new technology. Our aim is to endogenize the extent of allocative as well as organizational efficiency gains, both of which are important sources of economic growth.[5] We model the economy as consisting of two sectors, a rural agricultural sector and an urban manufacturing sector. Unlike Lewis, we assume that farmers engage in maximizing behavior. We are interested in decisions about the adoption of new labor-saving technology as well as in the implications of those decisions for economic growth in a two-sector world.[6]

9.1.1 Sectors

Agricultural output is produced with land and labor. Land is a fixed factor that limits the maximum feasible production. The land is split up into different farms that differ in size and fertility. The distribution of size and fertility is exogenous to our model and assumed to depend solely on geography and climate. In contrast, urban industrial output is not constrained by any fixed factor. Instead, output is produced with labor using a constant-returns technology. Individuals in our model are either farmers (i.e., owners of land), farmworkers, or urban dwellers. An individual may move between these three states: higher farm profits induce workers to become farmers, higher rural wages create an incentive for becoming a farmworker and for people to move from urban to rural areas, and higher urban wages pull workers to the cities.

9.1.2 Markets

There is perfect competition in the market for industrial goods, agricultural goods, and labor in the two sectors. Individuals differ in their preferences for rural versus urban labor. When relative wages in urban areas rise, more people decide to migrate from the farms to the cities, but not everyone will move. It follows that expected wages in the two sectors do not have to be equal. Cultural differences as well as education, peer pressure, and family considerations may also create an attachment to either rural or urban living.

As in Harris and Todaro 1970, the relative price of agricultural output in terms of manufacturing goods is a decreasing function of agricultural output and an increasing function of manufacturing output: $P_A/P_M = p = p(Y^A/Y^M)$, with $p' < 0$. This assumption captures the demand side of our model; we do not model consumption choices.

9.1.3 Utility

Preferences are separable in the utility of income, on the one hand, and the utility from living in rural/urban areas, on the other hand. Utility of income is homogenous and linear in income, while workers are heterogeneous in terms of the utility of residence. Farmers maximize the present discounted value of future utility using an exogenous and fixed-rate-of-time-preference r. For simplicity, we assume infinite horizons. At the same time, they compare this value to the present discounted value of working on other farms and switch between owning land and working for others when the latter gives higher future utility.

9.1.4 The Production Technology

We assume a Leontief production function in agriculture and a linear production function in urban industry:

$$Y_t^A = \min[A_t N_t^A, FL], \tag{1}$$

$$Y_t^M = B_t N_t^M. \tag{2}$$

Y^A denotes the level of output of agricultural produce and Y^M, modern urban output; A denotes the level of labor-augmenting technology in agriculture and B, technology in manufacturing. N^A is the number of workers in agriculture and N^M, in manufacturing. L is arable land and F denotes the fertility of the soil. It follows that if the number of effective labor units AN^A is up to the task, sustainable farm output is FL. There are constant returns to scale in industry but sharply diminishing returns in agriculture once we hit the capacity of land.[7]

The production frontier consists of two linear segments, HE and EI, as shown in figure 9.2. The distance OH in the figure equals FL, the maximum output possible in agriculture. The slope of the segment EI equals the ratio of marginal labor productivities in the two sectors, $-A/B$. At point E, modern output is shown by the distance OC and farm output by OH $= FL$, and total output at world prices is shown by

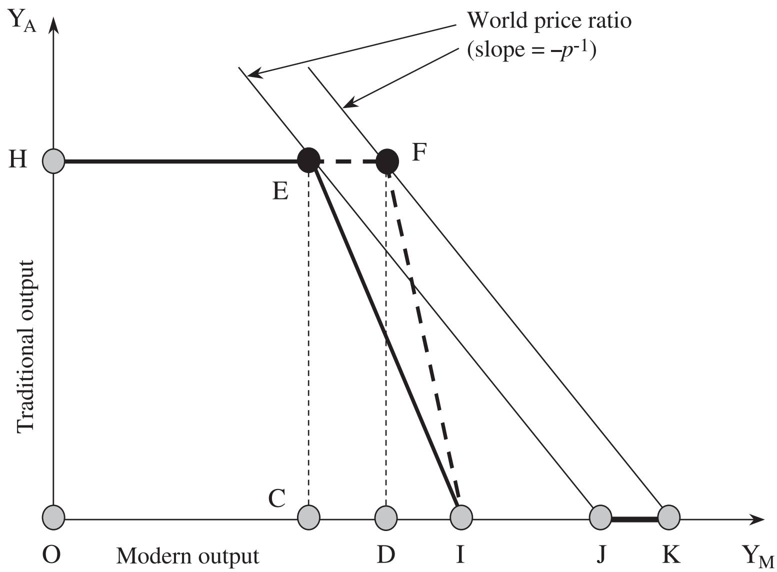


Figure 9.2
Technological change.

the distance OJ. Maximum possible output in manufacturing *BN* is shown by the distance OI and is assumed constant. Labor-saving technological progress in agriculture increases *A* and shifts the production frontier outward from HEI to HFI, increasing modern output and total output by CD = JK.

We assume that farmers differ in their ability to understand and adopt leading-edge technology.[8] The cost function *h* is rising in the rate of technology adoption, *a*, but falling in the ability to take on new technology, *b*:

$$h(a, b),\ h_a > 0,\ h_b < 0,\ h_{ab} < 0,\ h_{aa} > 0. \tag{3}$$

We assume that the cross-derivative is negative, which means that the marginal cost of learning is falling in the ability to learn.

9.1.5 Profits and the Value of Land

A farm generates a stream of revenues. The farmer pays wages *w* to his workers and retains all profits. We assume for simplicity that farmers do not work in the field so that their utility is simply linear in profits.

Farmers continue to farm their land using paid labor until it becomes optimal for them to abandon the farm and become agricultural workers elsewhere. This happens when the expected lifetime utility of working at a different farm (perhaps a bigger and more fertile one) exceeds the expected utility of continuing to farm one's own land.

Farmers maximize the present discounted value of future utility (profits) from time zero to infinity. It follows from our assumed utility function that this amounts to the maximization of the value of land. Profits for a given farmer i in real terms are defined as follows in terms of traditional output:

$$\pi_i = F_i L_i (1 - w/A_i) - h(a_i), \tag{4}$$

where w/A is the cost of producing one unit of output and the cost of technology adoption a is denoted by $h(a)$. The value of a given farm i is then given by

$$V_i = \int_0^\infty [F_i L_i (1 - w/A_{it}^*) - h(a_{it}^*)] e^{-rt}\, dt, \tag{5}$$

which is the present discounted value of expected profits (utility) along the optimal, value-maximizing path per unit of land. In steady state where $a = 0$ and $h(0) = 0$, equation (5) simplifies to

$$V_i = [F_i L_i (1 - w/A_i^*)]/r, \tag{5'}$$

where A^* is the profit-maximizing level of technology—which, as we show later, does not have to equal the state of frontier technology—and r is the exogenous rate of time preference.[9]

The farm will stay in business as long as V_i is greater than the discounted expected value of agricultural wages.[10] If farm wages were to rise dramatically, or if the fertility of land were to fall due to adverse climatic conditions, the farmer might be better off closing down and working for someone else. Clearly, any adverse climatic change or increase in the level of wages will first push those farming the smallest and least fertile plots into abandoning their land.

9.1.6 The Labor Market

We have assumed that labor is heterogeneous when it comes to preferences toward living in rural versus urban areas. Some, but by no means all, workers will decide to migrate to urban areas when rural wages fall below urban wages, and it follows that expected wages are not equalized across the two geographic areas. Labor supply in rural areas N^A is an increasing function of the ratio of agricultural to industrial wages and vice versa for labor supply in urban areas N^I. The sum of labor supplied in the two areas equals the aggregate labor force minus the number of farmers:

$$N^A\left(\frac{w^A}{w^I}\right) + N^I\left(\frac{w^A}{w^I}\right) = \bar{N} - N^F\left(\frac{w^A/r}{V}\right), \tag{6}$$

where $\bar{N}$ denotes the labor force and N^F the number of farmers, which is a negative function of the ratio of the discounted value of future farm wages and the value of owning land.

Labor demand in rural areas is determined by the size of the land, its fertility, and the state of technology and is—at each moment in time—independent of agricultural wages.[11] By equation (1), $N^A = \sum F_i L_i / A_i^*$ where F is the fertility of land and A^* is the optimal level of technology along the optimal path. Labor demand in rural areas is independent of wages—for a given, fixed level of technology A—as long as all farms stay in business. In contrast, the labor demand schedule in urban areas is horizontal at level B. Together, the two labor supply equations and the two labor demand equations determine wages and employment in both sectors.

9.1.7 *Technology Adoption and Closing in on the Frontier*

A farmer maximizes the value of his land V_i. She needs to decide whether to adopt cutting-edge technology or to lag behind, and if so by how much. Backward farms employ low-level technology and compensate by having many workers, while modern farms have cutting-edge technology and fewer workers. We assume that worldwide potential, or leading-edge, technology A^p is constant in the short run but subject to infrequent unanticipated discrete jumps:

$$A_t^p = \bar{A}. \tag{7}$$

The farmer decides on the speed of adoption of state-of-the-art technology—denoted by a—such that his own level of technology evolves according to

$$\dot{A}_{it} = a_i(\bar{A} - A_{it}), \tag{8}$$

where $\dot{A} = dA/dt$. We define a to be a choice variable and assume that the cost of learning depends on the farmer's ability to digest and take on new technologies.[12] In this we follow Schultz (1944), who proposed the idea that the gap between traditional production methods and frontier technology in agriculture creates the conditions necessary for growth.

The essence of the farmer's problem is to choose how many resources to use up today in order to have better technology tomorrow that will allow labor to be shed and wage costs to be cut for a given level of output, which is constrained by the supply and fertility of arable land.[13] There is one control variable, the rate of technology adoption a, and one state variable, the level of technology A. Equation (9) gives the optimal rate of technology adoption:

$$h_{ai} = q_{it}(\bar{A} - A_{it}). \tag{9}$$

The left-hand side shows the marginal cost of learning about new technology and the right-had side shows the marginal benefit, which is equal to the product of the value of new technology at the margin, q, and the marginal effect of increasing the learning intensity on the level of technology. Finally, there is the differential equation for the value of new technology:

$$\dot{q}_{it} = (r + a_i)q_{it} - w\frac{F_i L_i}{A_{it}^2}. \tag{10}$$

Combining equations (9) and (10) gives the rate of change of the intensity of technology adoption:

$$\frac{\dot{a}_{it}}{a_{it}} = \frac{1}{\varepsilon(h_a, a)}\left[r - \frac{w_t F_i L_i(\bar{A} - A_{it})}{A_{it}^2 h_a}\right]. \tag{11}$$

The interest rate reflects the marginal cost of learning about new technology and the second term within the brackets is the marginal benefit of learning, namely, the marginal benefit of increasing a. The marginal benefit consists of the reduction of wage payments made possible by investing in new technology today. The marginal (current) cost of raising a is h_a, and shows up in the denominator in the marginal benefit term, while the absolute fall in wage costs per unit of time is $wF\bar{L}(\bar{A} - A)/A^2$. The ratio of the two is the rate of cost savings per unit of spending on technology adoption—that is, the rate of return to investing in, or learning about, new technology. When the marginal

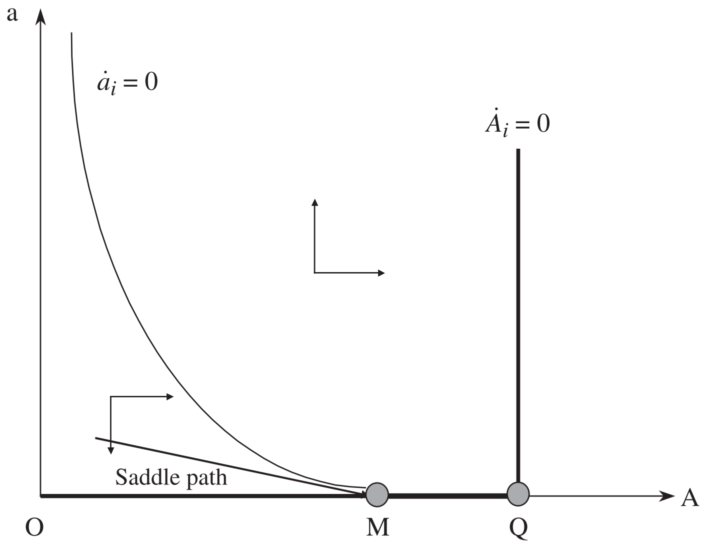


Figure 9.3
The farmer's problem.

benefit term exceeds the marginal cost r, the rate of adoption a is high but falling. When the marginal benefit falls short of the marginal cost, the intensity is low but rising. The term $\varepsilon(h_a, a)$ denotes the elasticity of the marginal adoption cost with respect to adoption a. The higher this elasticity, the more responsive is the farmer to changes in the marginal benefit and marginal cost of learning.

The two differential equations (8) and (11) are solved together in the phase diagram in figure 9.3. The $\dot{A} = 0$ locus starts at the origin, follows the horizontal axis to point Q, and then becomes vertical; the distance OQ equals $\bar{A}$. The $\dot{A} = 0$ locus slopes down throughout and cuts the horizontal axis at M to the left of Q when $r > 0$. Importantly, as long as $r > 0$, the farm will never converge to $\bar{A}$ because the marginal benefit of increasing a is falling, and in the end this is not enough to justify the sacrifice of current profit due to a positive interest rate.

The horizontal segment MQ shows the distance from the technological frontier in steady state. This segment shows the extent to which the representative farm does not adopt leading-edge technology. It is optimal not to converge all the way to the frontier. A country with small agricultural plots lacking in fertility and farmers who find it difficult to adopt new technology (h_a very large) is likely to be located from the frontier.

9.1.8 Optimal Backwardness

It is common nowadays to view economic growth as being driven initially by learning about—that is, imitating—new technologies and converging to a technological frontier. Once the frontier is reached, a process of invention and discovery takes over.[14] In contrast, our simple analysis—as depicted in figure 9.3—shows that it may be optimal for economies to stay away from the technological frontier for reasons having to do with factors exogenous to economic models. Relative backwardness may be the optimal strategy. We can see from equation (11) how the length of the segment MQ—the degree of technological backwardness—is determined within our model, and this gives us several interesting implications.

Optimal backwardness varies directly with the state of frontier technology. The reason is diminishing returns to investing in new technology as the marginal reduction in wage costs is falling in the level of technology A. For this reason the representative farm finds it optimal *not* to keep a constant gap between its own level and the level of leading-edge technology.[15] Instead, the gap is larger the more advanced the frontier technology.[16]

The lower the wages in rural areas, the weaker the incentive to invest in new technologies since farms can make use of cheap rural labor. If a large segment of the population only wants, or is confined by cultural and institutional factors, to live in rural areas, then equilibrium wages will be lower and the incentive to learn about new production methods weaker. Clearly, there is no incentive for technological improvements in a slave economy with abundant labor! A lower level of urban technology B has an effect in the same direction by not creating attractive employment opportunities.

The size of each farm and the fertility of its soil are important for how close to the frontier we come. The bigger the farm, and the more fertile the soil, the greater the incentive to adopt new technologies. Bigger farms using more fertile soil will adopt better technologies than the smaller and less fertile ones. At the aggregate level, the size and fertility distribution will matter for overall agricultural productivity.

Low costs of adopting technology will also speed up the adoption of modern technology and bring us closer to the frontier. This implies that the marginal cost of adoption—the cost of adopting new technology at the margin—is low. One reason could be an educated workforce (see Nelson and Phelps 1966). Again, the distribution of learning abilities among the population of farmers will matter for aggregate outcomes. Also, the higher the rate of time preference r, the farther away from the frontier we find ourselves.

Finally, the speed of adjustment along the saddle path depends on the convexity of the adoption cost function h. When this function is very convex (h_{aa} takes a large value), the speed of adjustment is slower.

9.1.9 The Harris-Todaro Effect: Labor Pulled to the Cities

Technological improvements in the urban manufacturing sector raise urban wages and cause labor supplied to agriculture to fall. Fewer people are now willing to work in agriculture for the prevailing rural wages. There follows an increase in rural wages, and the attendant increase in wage costs encourages farmers to invest in better technology, which lowers labor demand in agriculture. In figure 9.4 the speed of adoption of new technology initially picks up as indicated by the upward shift of the $\dot{a} = 0$ locus, but then falls until a new steady state is reached at point N where technology A is closer to the unchanged frontier at Q.

Rural wages are higher in the new steady state than before because the technological progress and the accompanying fall in labor demand

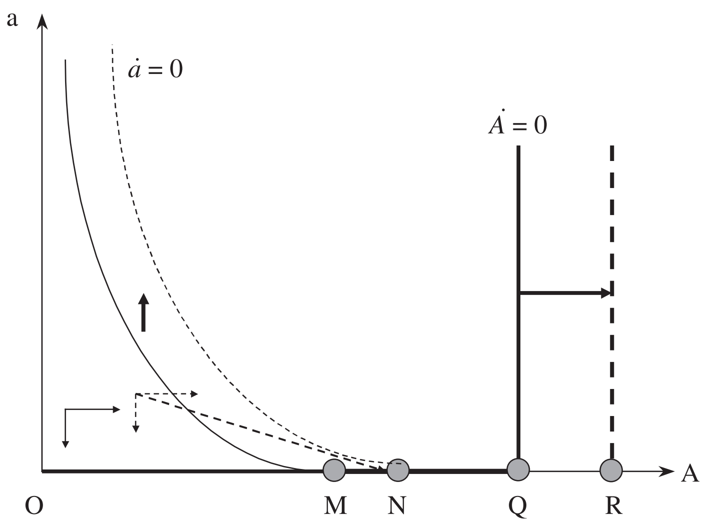


Figure 9.4
Urban "pull" vs. rural "push."

only partially offset the initial fall in labor supply. We are left with the empirical prediction that living standards in rural areas should be rising if the cause of the migration is technological progress in the cities. Notice also that the value of land should be falling. Farmers lose and farmworkers gain.

9.1.10 The Schultz Effect: Labor Pushed to the Cities

From the preceding analysis, we can see that the steady-state level of technology at the farm level is increasing in the level of frontier technology $\bar{A}$. With more and better technology available in the world, each farm ends up more advanced as long as $h_a < \infty$, $w > 0$, and $F_i L_i > 0$. Clearly, a slave economy would not adopt any new technology because labor savings are of no value in this case; the same applies to a farm where the land is useless or the cost of technology adoption is infinite. The increase in $\bar{A}$ shifts both loci to the right in figure 9.4 as well as the saddle path. The level of a jumps to the new saddle path and then gradually falls as we move to the new steady state at N with a higher level of steady state A. The effect on the standard of living in urban areas will now depend on the elasticity of labor supply with respect to wages. If labor supply is very inelastic, that is, if people have a strong preference for living in rural areas, the fall in labor demand will cause the rural wage and hence also the standard of living in rural areas to fall drastically. In contrast, the value of land will increase.[17]

9.1.11 Economic Growth

Our model suggests two possible sources of growth in addition to those familiar from the empirical growth literature. Economic growth can arise from the introduction of new agricultural technology worldwide and its subsequent adoption at the country level (the Schultz effect). In this case, we would expect to see a transfer of resources out of agriculture go along with a modest effect of structural change on economic growth. Growth can also arise from the pull of cities where technological progress in urban areas raises labor demand, pulls labor from agriculture, and raises rural wages, inducing farmers to adopt new technologies (the Harris-Todaro effect). If so, we would expect to see a transfer of resources out of agriculture go along with a stronger effect of structural change on growth than if the Schultz effect were prevalent. Over a period of study—which, in our case, will be 1965–1998—

both the frontier $\bar{A}$ and rural wages w will have increased. In contrast, the fertility of the soil and the size of land will not have changed much. There may have been some change in the level of education among farmers, but the movers and shakers in our model are the productivity frontier and rural wages. In the cross-section of countries under study, the growth rate of technology—our proxy for growth—will turn out to depend on these two shifts, which we represent by the change in the share of agriculture in value added, as well as on various exogenous variables.

Before returning to the empirical relationship between growth and structural change, we want to consider some historical evidence concerning the Schultz effect and the Harris-Todaro effect in Iceland.

9.2 Pushing and Pulling in Iceland

We have found changes in farm technology to be induced either by technical advances and wage hikes in urban areas or by progress in agriculture at the world level, holding fixed the size and fertility of land and the ability of farmers to take on new technologies. One can test which type of process is at work by looking at the evolution of wages per unit output w/A. If labor is pushed to urban areas by technological developments taking place within the agricultural sector, we have the prediction that A goes up on all farms leading to a fall in labor demand and lower wages per unit output. If, in contrast, it is the urban pull that is driving the process, we have rising wages causing farmers to take up labor-saving technology, hence raising A on each farm. In this case, wages per unit output w/A may not fall.

Iceland provides ideal testing grounds for our hypotheses. The economy was based on agriculture and remained stagnant until the end of the nineteenth century. Individual farmlands varied greatly in size and natural yield. The agricultural technology was very basic throughout, and no important improvements occurred before 1900. For example, the use of chemical fertilizers started only after 1920 (Jonsson 1993), which is more than fifty years after their introduction elsewhere. Produce was limited to a small selection of vegetables and hay for feeding livestock over winter. The population remained stagnant for almost a thousand years. It was 50,000 in 1703 and had not grown since the years after the settlement of the island seven hundred years earlier. It remained stagnant for the rest of the eighteenth century and by the late nineteenth century had only grown to 70,000. There was considerable social mobility among servants, tenants, and landowners, which contributed to a less rigid class system than that of European societies (Jonsson 1993).

Icelandic farmers had a larger labor force at their disposal than those of other European countries. This was mainly due to the absence of competing sectors on the island but also helped by legal restrictions on the movement of people from the traditional farm sector to other pursuits. In thirteenth century law a formal permission from local authorities was required for leaving agriculture and local authorities were obliged to provide a form of social insurance for non-farmworkers. A similar law can be found as late as 1887. One rationale for this law was that fishing and commerce were intrinsically more risky or volatile than agriculture. Even so, the law was clearly intended to provide cheap labor to agriculture. The mobility restrictions, which bordered on slavery, affected around 25 percent of the population in the nineteenth century. Workers who did not have farmland were required to reside with an established farmer who "owned" them and was entitled to all their earnings—on the farm as well as outside. In return, the farmers were required to provide food and shelter as well as an annual allowance that amounted to half the value of one cattle. The allowance was generally *not* sufficient to enable a man and a woman employed on the same farm to marry and have children. In fact, workers were not allowed to leave their masters without permission and corporal punishments were common. The mobility restrictions served basically three purposes. First, they created social stability in that a limited number of workers were allowed to rely exclusively on inherently volatile fishing and commerce. Second, and perhaps foremost, the real wage of farmworkers was kept low, which helped sustain farming. Third, population growth was kept down by confining a significant part of the population to slavelike conditions. These laws were abolished in 1893 and all individuals over the age of twenty-one allowed to choose their employment and keep the wages, making farmers face stiffer competition for labor from the expanding fishing villages.

There were some attempts made by Denmark—the colonial ruler until 1918—to promote agricultural reforms. In the eighteenth century, the Danes used laws and regulations, financial incentives, and the publication of books and pamphlets to encourage farmers to adopt new technologies and more efficient farming methods. On one occasion, the Danish authorities sent fourteen Danish and Norwegian farmers to Iceland to train farmers to grow grain and vegetables (Jonsson

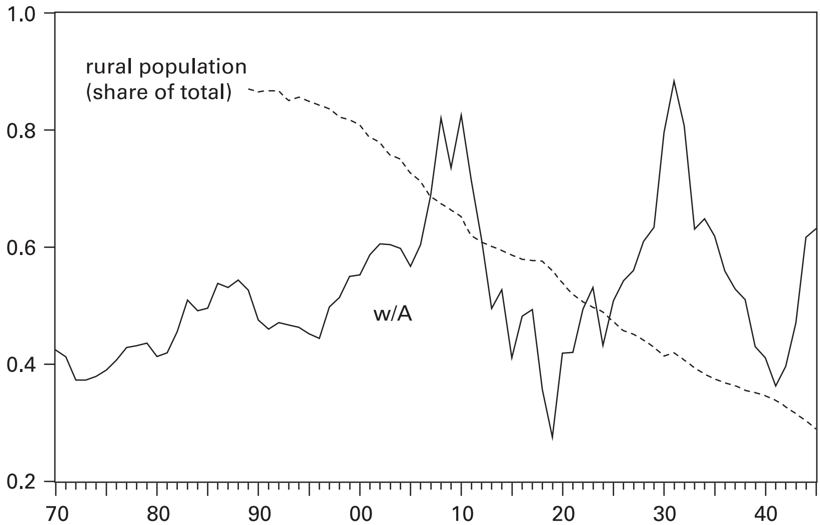


Figure 9.5
The ration w/A for agriculture in Iceland, 1870–1945.
Source: Hagskinna (Statistics Iceland).

1993). Also, a new breed of sheep was introduced with calamitous consequences for the local stock. On the whole, these attempt at promoting better technology proved futile, apart from some increase in vegetable production. The abundance of cheap labor made any productivity improvements a low priority.

The economic growth that started around 1870 coincided with a structural transformation from agriculture to fishing, and later to a service economy. Figure 9.5 shows the ratio of total wage payments on farms (wN^A) and the value of agricultural production (AN^A) in Iceland over the period 1870–1945. This series measures average wages per unit output w/A, which could rise if labor was being pulled to the cities but would fall if indigenous productivity improvements pushed labor out of agriculture. Notice the absence of a downward trend. The period 1870–1905 has rising wage costs. Cyclical behavior follows. The figure also shows the share of the population living in rural areas. The trend is downward throughout, starting around 0.84 and ending at 0.29.[18] These numbers indicate that labor was pulled away from agriculture by an expanding urban sector.

If driven by the pull of emerging towns—mostly fishing villages—we would expect farms using the smallest and least fertile soil to be abandoned. During this period the number of farms starts at 5,652 in 1861, but by 1942 there are 652 farms that have been vacated. Another piece of evidence for the pull theory is the evolution of the ratio of the average farm prices to average wages of farm workers, which fell from 17.8 in 1922 to 6.34 in 1942.[19] Based on the evidence of rising wage costs, falling land values, and infertile farmlands being abandoned, we conclude that labor was pulled from rural areas to the cities by the expansion of new industries in the urban areas.

The story told here fits well within our model. Prior to 1870, agriculture did not take advantage of foreign technology because of the abundance of cheap labor—due to social legislation and a lack of outside opportunities—the small plots of land, the general inhospitable terrain, and the isolation of the country due to its geographical location and also a lack of familiarity with foreign languages (apart from Latin). When progress came, it was not due to any changes on these fronts but caused solely by expanding opportunities in the growing fishing sector, which initially faced constant returns to scale because of the abundance of fish stocks around the island. By pushing up urban as well as rural wages, the agricultural sector was made to modernize. This is the Harris-Todaro effect.

Let us now leave Iceland and return to the global setting.

9.3 Structural Change and Growth around the World

In this section, we want to see, first, whether structural change has played an independent role as a determinant of economic growth across countries and, second, whether structural change responds to some of the same factors as economic growth. To this end, we test whether structural change can help explain the divergent growth experiences of eighty-six countries when we place a measure of structural change—that is, the change in the share of agriculture in value added—side by side with other variables suggested by and used in the growth literature.

Unfortunately, despite the great effort of Madisson (2004) and others to compile historical statistics for empirical use, regression analysis of economic growth across countries does not, for dearth of data, reach farther back in time than to the 1960s. So this is where we have to start our statistical analysis. We study eighty-six countries, some rich and some poor, in the period from 1965 to 1998, using mostly data from the World Bank's *World Development Indicators* (2002), with two exceptions:

the data on natural capital are taken from World Bank 1997 and the data on democracy, our measure here of the quality of institutions, are obtained from the Polity IV Project (Marshall and Jaggers 2001).

Recall figure 9.1, which depicts the raw relationship between per capita economic growth and structural change in 1965–1998. The figure shows that a decrease in the share of agriculture by 13 percentage points from one country to another is associated with an increase in annual per capita growth by 1 percentage point. This is not much different from the results of Temin (1999), based on fifteen European countries. Temin finds that a 20 percent decrease in the share of agriculture in the labor force goes along with a 0.8 percent increase in the rate of growth over the period 1955–1975. True, figure 9.1 shows a mere correlation: the causation can run both ways. Slow growth may hinder structural change just as change may spur growth.

We now proceed to estimate a series of growth regressions for the same eighty-six countries as before, again for the period 1965–1998. The strategy here is to regress the rate of growth of GNP per capita on structural change, as defined in figure 9.1, and then to add other potential determinants of growth to the regression one after another in order to observe the robustness of the initial result—that is, to see whether the structural change variable survives the introduction of additional explanatory variables that are more commonly used in empirical growth research.

Table 9.1 presents the results of this exercise. Model 1 shows the regression behind the bivariate relationship between structural change and growth in figure 9.1. The negative coefficient on structural change does not, however, enable us to determine the relative importance of the Harris-Todaro effect and the Schultz effect—that is, whether labor tends to be pulled rather than pushed out of agriculture. In model 2, we add the share of gross domestic investment in GDP. In model 3, we proceed to add education, represented by the secondary school enrollment rate for both genders; this is the measure of education most commonly used in empirical growth work.[20] Again, education stimulates growth even if no attempt has been made to adjust the school enrollment figures for quality. The effect of investment on growth is now smaller than in model 2 because there investment was presumably picking up some of the effect of education on growth.

In model 4, we add the logarithm of initial income (i.e., in 1965) to capture conditional convergence—the idea that rich countries grow less rapidly than poor ones because the rich have already exploited

Table 9.1
OLS results on economic growth

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Structural change	−0.079	−0.051	−0.068	−0.052	−0.050	−0.038	−0.034	−0.032	−0.032
	(4.52)	(3.31)	(4.37)	(3.97)	(3.89)	(3.13)	(2.92)	(2.77)	(2.89)
Investment		0.168	0.119	0.076	0.088	0.073	0.071	0.058	0.056
		(5.79)	(3.83)	(2.80)	(3.26)	(2.93)	(3.00)	(2.40)	(2.37)
Secondary education			0.019	0.061	0.047	0.038	0.033	0.034	0.032
			(3.31)	(7.19)	(4.56)	(2.88)	(3.48)	(3.68)	(3.59)
Initial income				−1.363	−1.340	−1.412	−1.592	−1.608	−1.723
				(6.01)	(6.04)	(6.94)	(7.80)	(8.02)	(8.56)
Population growth					−0.465	−0.503	−0.435	−0.503	−0.574
					(2.19)	(2.59)	(2.32)	(2.70)	(3.12)
Natural capital						−0.055	−0.056	−0.055	−0.063
						(4.12)	(4.34)	(4.33)	(4.92)
Democracy							0.075	0.081	0.085
							(2.89)	(3.16)	(3.40)
Migration to cities								0.025	0.027
								(2.02)	(2.25)
Hectares per capita									0.674
									(2.32)
Adjusted R^2	0.19	0.41	0.48	0.63	0.65	0.71	0.73	0.74	0.76

Note: t-values are shown within parentheses. Estimation method: Ordinary least squares. Number of countries: 86. Saudi Arabia is not included because of difficulties with its economic growth statistics.

more of the growth opportunities open to them—by sending more young people to school, for instance.[21] Here we see that the coefficient on initial income is significantly negative as expected; the preceding coefficients survive. The coefficient on education is now larger than in model 3 because there the effect of education was being held back by the absence of initial income from model 3. This also helps explain why conditional convergence need not entail absolute convergence: a high initial income impedes growth through the conditional convergence mechanism but encourages growth by enabling parents to send more of their children to school.

In model 5, we enter population growth into the regression to see if it matters for growth as suggested by the Solow model. We see that increased population growth impedes economic growth as expected, without reducing the statistical significance of the explanatory variables already included. Specifically, it takes an increase in annual population growth by about 2 percentage points to reduce per capita growth by 1 percentage point per year. In model 6, we add natural resource dependence, measured by the share of natural capital in total capital, which comprises physical, human, and natural capital (but not social capital; see World Bank 1997); we do this in order to test the resource curse hypothesis (Sachs and Warner 1995). The results show that natural resource dependence hurts growth as hypothesized without knocking out any of the other coefficients.

In model 7, we enter democracy. The democracy index is defined as the difference between an index of democracy that runs from 0 in hard-boiled dictatorships (e.g., Saudi Arabia) to 10 in full-fledged democracies and an index of autocracy that similarly runs from 0 in democracies to 10 in dictatorships. Hence, the democracy index spans the range from −10 in Riyadh to 10 in Reykjavík. The coefficient is significant statistically as well as economically, and it does not displace any of the variables inherited from the preceding models. In particular, we find room for independent contributions to growth from the efficiency gains from democracy as well as from structural change. The magnitude of the democracy coefficient means that an increase in democracy by a bit more than 13 points—for example, from −7 (as in Tunisia) to almost 7 (as in Turkey, with 6.4)—goes along with an increase in growth by 1 percentage point. We find no sign of nonlinearity in the relationship between democracy and growth, reported by Barro (1996), as democracy squared has an insignificant coefficient when added to model 7.

In model 8, we now proceed to add the change in the share of the urban population in total population as an indicator of migration from rural areas to cities. We do this because our structural change variable refers to the decrease of the share of farm output in total output, whereas the migration variable refers to the corresponding change in farm input, that is to say, labor, from 1965 to 1998, and thus reflects a different aspect of the structural transformation from agriculture and industry under study. The results show that migration from farm to city exerts an independent positive influence on economic growth without dislodging any of the earlier explanatory variables.

At last, model 9 shows what happens when we use the number of hectares of arable land per person as a proxy for the fertility of agricultural land. The coefficient in the southeast corner of table 9.1

suggests that more arable hectares per person—that is, more favorable conditions for farming—increase economic growth as suggested by our model in section 9.1 where we argued that more fertile land encourages farmers to adopt new technology more quickly and shed labor. We get the same result if we replace hectares of arable land by natural capital per inhabitant: natural resource abundance tends to stimulate growth even if natural resource dependence, as measured by the share of natural capital in national wealth, hurts growth. The rest of the story remains intact.[22]

The bottom line of table 9.1 shows how the adjusted R^2 rises gradually from 0.19 to 0.76 as more independent variables are added to the growth regression, indicating that model 9 explains around three fourths of the cross-country variations in the growth of output per capita. The results in model 9 accord reasonably well with a number of recent empirical growth studies. The coefficient on the investment rate suggests that an increase in investment by 18 percent of GDP increases annual per capita growth by 1 percentage point, a common result in those growth studies that report a statistically significant effect of investment on growth. The coefficient on the education variable in model 9 means that an increase in secondary school enrollment by 30 percent of each cohort (e.g., from 40 percent to 70 percent) increases per capita growth by 1 percentage point per year. The coefficient on initial income suggests a convergence speed of 1.7 percent per year, which is not far below the 2–3 percent range typically reported in econometric growth research. The coefficient on population growth is consistent with the coefficient on the fertility rate reported by Barro (1999). The coefficient on natural resource dependence suggests that an increase in the share of natural capital in national wealth by about 16 percentage points reduces per capita growth by 1 percentage point. Beginning with Sachs and Warner 1995, several recent studies have reported a similar conclusion, based on a variety of different measures of natural resource dependence. Thus far, however, few studies have reported a significant positive effect of democracy on growth, as we do here. Our results on the effects of migration and acreage of arable land on growth are the first of their kind, as far as we know.

The coefficient on structural change in the northeast corner of table 9.1 means that a decrease of 30 percentage points in the share of agriculture in value added goes along with an increase in per capita growth by one percentage point per year. This means that bringing down the share of agriculture from 50 percent to 20 percent would increase per capita growth by 1 percentage point, other things being the same. Is this a little or a lot? The average rate of growth of output per capita was 1.3 percent on average in the sample as a whole. This suggests that the effects of structural change on economic growth shown in table 9.1 are economically as well as statistically significant. Further, structural change also encourages growth through migration.

We do not have access to data that allow us to study the relationship between the fertility of the soil and farm size, on the one hand, and the pace of structural change, on the other hand. Engerman and Sokoloff 1994 provides indirect evidence, however. They argue that the superior growth performance of Canada and the United States, when compared to other New World economies, was due to less inequality in the distribution of income, which in our model translates into higher relative wages of farm workers. Elsewhere, in Latin America and the American South, the suitability of land for the cultivation of sugar and other crops—which generated economies of scale in the use of slave labor—in addition to a very large supply of Native Americans created great inequalities that excluded large segments of the population from participation in economic life. The result was lower rates of economic growth. This evidence may at first glance appear to go against our model in that large-scale farming was not conducive to growth. But notice the link among scale, institutions, and wages (slave labor!). With farmers facing close to zero wages for their workers, it is clear from our model that the incentives to adopt better technologies are minimal. In this the Latin American countries resembled our previous account of Iceland. Our model implies that rural areas in the North should have shed labor earlier and more rapidly than the South and Latin America. This was the case.

9.4 Concluding Thoughts

We have tried to shed new light on the determinants of the rate of technology adoption at the farm level, which underlies the transformation of societies from an agrarian base to an industrial one. We motivated our study by showing how economic growth in a sample of eighty-six countries is directly related to the devolution of agriculture around the world—that is, to the ongoing transfer of resources from agriculture to industry. We then presented a model showing how productivity gains in agriculture depend on external factors such as geography, the fertility of the soil, and the receptivity of farmers to new ideas and technologies. We found that a certain level of backwardness in agricultural technology could be optimal and that its extent depended on the same set of variables. Moreover, we divided technological progress in agriculture into two types, labor pull and labor push, and found that in Iceland—a country that suffers from a harsh climate and poor soil—it was mainly the pull of rising wages in the fishing sector that made farmers adopt better technology and ended a thousand-year-long period of stagnation. Also, we found that in our sample of eighty-six countries, the pace of structural change and urbanization increases economic growth.

The existence of abundant labor is an important obstacle to productivity growth in agriculture. When the fertility and size of the land is limited, compared with the number of workers living in rural areas, wages will be low and so will be living standards for the majority of the population. Farmers, as owners of land, will in contrast enjoy high standards of living. It is true that their land is not very productive, but they face an abundance of cheap labor and can enjoy the profits. Attempts at educating them and giving them information on how to improve productivity may make some adopt better technologies, although Iceland's experience in the late eighteenth century is not promising in that regard. But the pull of an expanding industry—which makes labor increasingly costly in rural areas—is the magic bullet that induces landowners to expend resources to learn about and adopt more modern technologies. This raises productivity, wages, and living standards for most people. But profits fall, and landowners may then use their influence to fight the emergence and expansion of other sectors of the economy. The road from agriculture is cleared through social conflict.

Appendix: Countries and Data

Table 9A.1
Country data

	Growth per capita	Structural change	Invest-ment rate	Secondary enrollment	Initial income	Popula-tion growth	Natural capital	Democ-racy	Migration to cities	Hectares per capita
Argentina	0.400000	−7.300000	22.80952	56.10345	9.237998	1.470422	6.696826	0.323529	11.80000	1.052941
Australia	1.700000	−4.000000	23.72727	85.75758	9.433151	1.527437	11.88886	10.00000	6.500000	2.982353
Austria	2.600000	−3.900000	23.78788	92.03125	9.202498	0.315617	2.642099	10.00000	−0.100000	0.200000
Bangladesh	1.400000	−28.30000	20.00000	17.71429	6.790419	2.353015	14.06030	−1.037037	17.70000	0.102941
Belgium	2.300000	−2.800000	19.54545	96.87097	9.319531	0.247816	0.003157	10.00000	3.800000	0.100000
Benin	0.100000	−6.400000	15.17647	12.00000	6.720454	2.895813	7.678190	−3.441176	28.20000	0.379412
Botswana	7.700000	−28.60000	26.85294	25.68750	6.217003	3.587553	6.302406	8.696970	44.40000	0.438235
Brazil	2.200000	−10.30000	20.68966	33.09677	8.055255	2.069047	7.894385	−0.382353	28.80000	0.364706
Burk Faso	0.900000	−1.700000	21.00000	3.448276	6.468213	2.270865	16.91058	−4.882353	10.70000	0.385294
Burundi	0.900000	−16.40000	11.50000	3.300000	6.034049	2.170652	19.85842	−5.911765	6.100000	0.208824
Cameroon	1.300000	8.500000	21.45833	17.46667	6.814414	2.719059	21.07668	−6.911765	30.80000	0.641176
Canada	1.800000	−1.800000	21.54545	86.90000	9.446412	1.313181	11.06857	10.00000	5.400000	1.820588
Central African Republic	−1.200000	7.000000	10.40909	9.555556	7.399641	2.212406	30.16046	−4.647059	13.60000	0.782353
Chad	−0.600000	1.700000	7.470588	4.923077	6.935563	2.435090	37.13253	−5.500000	14.20000	0.667647
Chile	1.900000	−0.200000	19.00000	54.80645	8.427527	1.658651	9.781933	0.794118	13.50000	0.297059
China	6.800000	−19.30000	30.61905	44.50000	5.852229	1.678090	7.229295	−7.411765	16.30000	0.100000
Colombia	2.000000	−15.10000	18.97059	39.51613	8.022589	2.246537	7.183313	7.735294	21.10000	0.123529
Congo, Republic	1.400000	−8.100000	31.72000	50.06452	6.281723	2.871701	14.46550	−5.352941	31.40000	0.100000

Costa Rica	1.200000	−13.30000	20.61765	40.24242	8.274037	2.815546	8.205335	10.00000	22.10000	0.126471
Cote d'Ivoire	−0.800000	−15.50000	17.32353	16.42424	7.567558	3.609279	18.00927	−8.323529	19.70000	0.247059
Denmark	1.900000	−3.400000	22.93939	101.4194	9.458631	0.300727	3.752806	10.00000	8.100000	0.500000
Dom Republic	2.300000	−11.70000	20.79412	35.33333	7.624535	2.395684	12.40725	2.764706	28.90000	0.164706
Ecuador	1.800000	−15.20000	19.23529	43.44828	7.418650	2.678254	17.01109	4.323529	24.70000	0.211765
Egypt	3.500000	−11.20000	20.76471	52.45455	6.918640	2.256647	4.549943	−5.264706	2.200000	0.058824
El Salvador	−0.400000	−29.40000	15.50000	24.41935	8.428312	2.173825	2.845800	2.382353	18.80000	0.105882
Finland	2.400000	−8.400000	23.96970	101.9375	9.152389	0.372213	6.601666	10.00000	17.10000	0.505882
France	2.100000	−4.000000	21.75758	86.03125	9.276593	0.566482	2.734652	8.029412	8.000000	0.311765
Gambia	0.400000	−5.300000	19.50000	13.09677	7.132293	3.385164	11.84402	5.676471	15.80000	0.250000
Ghana	−0.800000	−7.500000	11.87500	31.33333	7.723824	2.651659	7.220696	−3.588235	9.500000	0.200000
Greece	2.400000	−6.600000	25.39394	80.12500	8.763739	0.606721	3.656979	5.235294	12.30000	0.300000
Guatemala	0.700000	−5.300000	14.32353	16.37931	7.922867	2.620069	3.309386	0.352941	5.200000	0.173529
Guin-Bissau	−0.100000	12.70000	29.15000	6.416667	6.383902	2.688439	44.20389	−4.880000	15.60000	0.329412
Haiti	−0.800000	0.000000	10.87500	13.30000	7.494176	1.887783	6.682531	−5.970588	16.90000	0.141176
Honduras	0.600000	−20.90000	19.76471	22.08333	7.559643	3.189480	9.940350	2.617647	24.90000	0.405882
India	2.700000	−16.00000	18.55882	33.96875	6.751278	2.138814	19.78769	8.352941	8.400000	0.238235
Indonesia	4.700000	−37.90000	25.50000	31.46875	6.270482	2.040246	12.37799	−6.852941	22.90000	0.117647
Ireland	3.000000	−11.40000	21.03030	89.87500	8.822186	0.740981	8.117382	10.00000	9.900000	0.358824
Italy	2.500000	−5.400000	21.60606	72.31250	9.106717	0.304577	1.320394	10.00000	5.000000	0.182353
Jamaica	−0.400000	−2.400000	24.85294	58.29630	8.247188	1.120549	6.775690	9.823529	17.50000	0.085294
Japan	3.500000	−3.600000	30.81818	92.38710	8.933416	0.746232	0.757908	10.00000	11.50000	0.014706
Jordan	−0.400000	−12.40000	29.39130	47.96970	8.001284	4.430954	1.588902	−7.411765	24.20000	0.138235
Kenya	1.300000	−8.800000	17.38235	17.09677	6.444856	3.406624	9.439440	−5.382353	22.70000	0.241176
Korea, Republic	6.600000	−32.90000	29.35294	71.62500	7.385326	1.487917	1.749621	−0.382353	48.00000	0.055882
Lesotho	3.100000	−47.40000	42.78125	17.78125	6.686018	2.273760	3.332342	−2.757576	20.00000	0.252941

Table 9A.1
(continued)

	Growth per capita	Structural change	Investment rate	Secondary enrollment	Initial income	Population growth	Natural capital	Democracy	Migration to cities	Hectares per capita
Madagascar	−1.800000	5.600000	10.50000	14.83333	7.207412	2.679926	41.87061	−1.470588	15.80000	0.267647
Malawi	0.500000	−16.40000	17.38462	5.870968	6.147146	2.967671	11.78242	−6.617647	9.100000	0.232353
Malaysia	4.100000	−15.50000	28.41176	47.93939	7.622847	2.605487	8.618225	4.411765	26.00000	0.100000
Mali	−0.100000	−22.60000	17.50000	7.121212	6.544762	2.429771	41.04095	−3.970588	16.20000	0.305882
Mauritania	−0.100000	−7.800000	20.35714	9.225806	7.346237	2.519024	21.56988	−6.764706	45.90000	0.173529
Mauritius	3.800000	−13.80000	21.82353	44.81250	7.785509	1.236261	1.245193	9.548387	4.000000	0.100000
Mexico	1.500000	−8.500000	19.58824	43.03226	8.424645	2.449953	5.885405	−2.382353	19.10000	0.352941
Morocco	1.800000	−6.200000	20.44118	25.87500	7.478432	2.259302	4.075259	−8.117647	22.10000	0.408824
Mozambique	0.500000	−4.700000	12.73684	5.000000	6.442061	2.178103	12.68131	−4.625000	24.90000	0.241176
Namibia	0.700000	2.100000	19.05263	51.66667	8.341486	2.725272	10.07103	6.000000	13.20000	0.638235
Nepal	1.100000	−25.60000	17.50000	21.50000	6.713099	2.481033	17.69754	−3.441176	7.700000	0.152941
Netherlands	1.900000	−2.400000	21.84848	99.37500	9.392345	0.742322	1.523718	10.00000	3.700000	0.100000
New Zealand	0.700000	−11.70000	22.24242	86.51515	9.455385	1.156606	18.47280	10.00000	6.700000	0.808824
Nicaragua	−3.300000	7.300000	19.97059	33.51515	8.654876	3.016821	13.87773	−2.088235	12.80000	0.405882
Niger	−2.500000	−25.10000	11.42105	4.093750	7.427161	3.089812	54.24115	−4.911765	12.80000	0.552941
Norway	3.000000	−3.500000	26.69697	93.81250	9.197922	0.526447	10.01640	10.00000	16.60000	0.200000
Pakistan	2.700000	−12.90000	16.26471	15.62963	6.530558	2.817915	5.551898	1.735294	9.100000	0.247059
Panama	0.700000	−1.600000	19.57895	54.90323	8.271884	2.352255	6.472876	−1.205882	11.30000	0.232353
Papua New Guinea	0.500000	−11.40000	23.38235	10.90323	7.533894	2.404573	19.32447	10.00000	11.70000	0.032353
Paraguay	2.300000	−12.40000	20.76471	26.46875	7.618754	2.785736	11.53870	−4.117647	18.30000	0.455882
Peru	−0.300000	−9.200000	20.97059	52.75000	8.437215	2.356105	7.783727	1.029412	20.10000	0.191176

Philippines	0.900000	−9.000000	21.64706	61.66667	7.927151	2.622947	6.174310	0.264706	25.10000	0.102941
Portugal	3.200000	−23.70000	27.00000	59.00000	8.547195	0.286199	2.312559	4.205882	37.20000	0.244118
Rwanda	0.000000	−29.30000	12.64706	4.518519	6.476972	2.854258	21.70829	−6.352941	3.200000	0.126471
Senegal	−0.400000	−7.500000	12.44118	12.33333	7.300074	2.815546	16.78515	−3.264706	13.20000	0.420588
Sierra Leone	−1.600000	11.30000	7.357143	13.15385	6.630344	2.186490	28.00917	−4.558824	21.30000	0.117647
South Africa	0.100000	−5.600000	22.20588	62.23077	8.990545	2.260315	5.042582	4.970588	7.900000	0.464706
Spain	2.300000	−7.400000	23.00000	84.00000	8.927438	0.622871	2.856518	4.088235	15.90000	0.426471
Sri Lanka	3.000000	−7.100000	22.10345	57.03226	7.012424	1.581906	7.420707	6.147059	2.500000	0.091176
Sweden	1.400000	−4.000000	19.93939	90.90625	9.437063	0.439846	5.607987	10.00000	6.100000	0.358824
Switzerland	1.200000	−4.800000	25.18182	85.38710	9.805346	0.562614	0.868441	10.00000	14.70000	0.100000
Thailand	5.000000	−19.70000	28.70588	29.18182	7.006782	2.122661	6.485881	1.314286	6.700000	0.341176
Togo	−0.600000	−5.500000	17.31579	19.81250	7.407937	3.183170	15.18405	−5.882353	21.00000	0.750000
Trinidad and Tobago	2.600000	−5.700000	21.26471	62.33333	8.035911	1.120549	9.487274	8.500000	9.400000	0.100000
Tunisia	2.700000	−8.300000	26.14706	34.12121	7.671251	2.156122	7.908014	−7.029412	24.60000	0.482353
Turkey	2.100000	−23.80000	18.61290	36.75000	8.108092	2.176752	5.018668	6.352941	30.80000	0.555882
United Kingdom	1.900000	−1.600000	17.96970	87.65625	9.297948	0.251427	1.858975	10.00000	2.300000	0.100000
United States	1.600000	−1.900000	18.27273	90.60000	9.759472	1.005412	4.112407	10.00000	4.900000	0.805882
Uruguay	1.200000	−11.50000	14.44118	66.74194	8.658991	0.609946	11.64545	2.794118	10.30000	0.452941
Venezuela	−0.800000	−0.200000	21.94118	32.87097	8.914335	2.876591	18.92948	8.470588	19.70000	0.197059
Zambia	−2.000000	5.500000	17.82759	17.13333	7.185837	3.049176	37.77010	−3.882353	16.10000	0.888235
Zimbabwe	0.500000	1.700000	17.02941	25.69697	7.655047	2.937816	8.483433	−0.413793	19.50000	0.364706

Sources: See text.

Notes

We are grateful to our discussant, Piergiuseppe Fortunato, and to other workshop participants as well as an anonymous referee for helpful comments and suggestions.

1. Data are taken from the World Bank's *World Development Indicators* (2002).

2. The Spearman rank correlation is 0.31 and statistically significant. The Spearman correlation is less sensitive to outlying observations than the standard Pearson correlation.

3. According to his thesis, the preceding thirty years of depression and wars slowed down the rate of industrialization in many European countries—Britain being the most notable exception—and, therefore, the share of agriculture in the labor force was excessive at the end of World War II. This set the stage for the postwar period of high economic growth when the pent-up energy of underutilized ideas and education were harnessed by expanding industries that needed the workers supplied by rural areas.

4. In a recent paper, Temple (2001) conducts a growth-accounting exercise in order to measure the effect of a structural transformation away from agriculture on postwar growth in some large OECD economies. He shows that this factor helps explain differences in the rate of postwar growth across countries, as well as the growth slowdown that occurred after 1970—when the transformation was completed. He uses differences in the marginal product of labor to assess the importance of this transformation by calculating its share of the measured Solow residual. In particular, he derives bounds on the intersectoral wage (productivity) differential to derive upper and lower bounds on the magnitude of the reallocation effect. He finds that labor reallocation typically accounted for a twentieth to a seventh of growth in output per employed worker during the period 1950–1979. The effect was greatest in Italy, Spain, and West Germany. However, he does not try to identify the forces that drive this structural transformation.

5. By *allocative* efficiency gains, we mean those gains that involve the reallocation of resources along the economy's production frontier from less efficient lines of employment of labor, capital, and other inputs to more efficient ones, thereby increasing national economic output at full employment. By *organizational* efficiency gains, we mean those gains that stem from outward shifts of the production frontier as a result of the reorganization of production, for instance, through the adoption of new production methods or better management.

6. In this we are close to Schultz (1944), who worried about the ability of an urban sector to absorb labor shed by agriculture due to labor-saving technical progress.

7. This is as argued by Becker, Glaeser, and Murphy (1999), who claim that increased population in urban areas fosters the division of labor creating constant or even increasing returns to scale while increased population in rural areas that rely on traditional industries is bound to hit diminishing returns.

Note that we could have derived all the results that follow by changing the production technology in the farming sector to a pure diminishing-returns-to-labor technology—without an explicit capacity constraint FL—and a perfectly elastic supply of labor. However, we think the setup in the assumption is equally realistic with the added benefit of making the mathematics that follow a bit easier.

8. Nelson and Phelps (1966) argue that education gives people the ability to learn. In our context, it helps workers learn about new technology.

9. The value of land is increasing in the size of land and its fertility as first proposed by Ricardo (1819).

10. The implied threshold is $w = F/(1 + F/A)$. If wages are higher than this threshold for a farmer with fertility of land F, he will decide to abandon the farming and become a farmworker elsewhere. If not, he continues farming his land.

11. In the long run higher wages will make the farmer adopt new technology, which will reduce labor demand.

12. Equations (3) and (8) imply that technological adoption becomes increasingly costly the closer we get to the technological frontier. Thus, closing half the gap between current and frontier technology always incurs cost $h(1/2)$, but the absolute productivity gain is larger the farther we find ourselves from the frontier. We can imagine the farmer first taking on board the most important ingredients of modern technology, then increasingly focusing on less relevant refinements.

13. The representative farm′s maximization problem can be written as follows:

$$\max_{a,f} \int_0^\infty [FL - w_t N_t - h(a_t, b)]e^{-rt}\, dt,$$

subject to $A_t N_t = FL$ and also that equation (8) holds.

14. A recent paper by Acemoglu, Aghion, and Zilibotti (2003) makes the point that a different set of institutions may be desirable in the transition to steady state than on the frontier.

15. The first derivative of the marginal benefit of improved technology in equation (11) is negative:

$$\frac{d[-wf\bar{L}(\bar{A} - A)/A^2]}{dA} = -wF\bar{L}/A^2[2\bar{A}/A - 1] < 0,$$

while the second derivative is positive:

$$\frac{d[-wF\bar{L}/A^2[2\bar{A}/A - 1]]}{dA} = 2wF\bar{L}/A^3[3\bar{A}/A - 1] > 0.$$

16. Taking the total differential of the terms in the square bracket of equation (11), we find that

$$\frac{dA}{d\bar{A}} = \frac{1}{2\dfrac{\bar{A}}{A} - 1} < 1$$

for the $\dot{a} = 0$ locus while we have $dA/d\bar{A} = 1$ for the $\dot{A} = 0$ locus. Hence, the latter shifts more than the former, and the gap between the level of frontier technology $\bar{A}$ and the steady-state technology A is increased.

17. An improvement in the ability of farmers to adopt new technology, an increase in the size and fertility of the land, and a fall in the rate of time preference would all make productivity in agriculture improve and so push workers to the cities.

18. The transformation has continued until recent years; the share is currently below 5 percent of the total population.

19. See *Hagskinna* (Statistics Iceland) and *Hagvöxtur og iðnvæðing* (National Economic Institute).

20. It makes little difference for the results whether we use the logarithm of the enrollment rate to reflect decreasing returns to education or not.

21. Initial income is defined as purchasing power parity adjusted GNP per capita in 1998 divided by an appropriate growth factor to ensure consistency between our income measures in 1965 and 1998 and our measures of economic growth between those two years.

22. Adding a dummy for Africa to model 9 makes virtually no difference to the results: the dummy coefficient is -0.577 and marginally insignificant ($t = 1.97$). Removing Botswana from our sample—in order to counter claims that Botswana's impressive growth record drives the results reported in some empirical growth studies—makes no material difference either, except the urban migration coefficient drops in size and significance ($t = 1.32$) and the hectares variable becomes marginally insignificant ($t = 1.93$); the rest of the regression remains essentially intact (not shown).

References

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti. 2003. "Distance to Frontier, Selection and Economic Growth." Unpublished manuscript.

Barro, Robert J. 1996. "Democracy and Growth." *Journal of Economic Growth* 1: 1–27.

Barro, Robert J. 1999. *Determinants of Economic Growth*. Cambridge, MA: MIT Press.

Becker, Gary S., Edward L. Glaeser, and Kevin M. Murphy. 1999. "Population and Economic Growth." *American Economic Review, Papers and Proceedings* 89, no. 2 (May): 145–149.

Engerman, Stanley L., and Kenneth L. Sokoloff. 1994. "Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies: A View from Economic Historians of the United States." NBER Working Paper Series on Historical Factors in Long-run Growth, historical paper no. 66.

Galor, Oded, and Omer Moav. 2003. "Das Human Capital." Brown University working paper no. 2000-17.

Galor, Oded, and David N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90, no. 4: 806–828.

Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-sector Analysis." *American Economic Review* 60 (March): 126–142.

Jones, Charles I. 1999. "Was the Industrial Revolution Inevitable? Economic Growth over the Very Long Run." Unpublished manuscript, Stanford University.

Jonsson, Gudmundur. 1993. "Institutional Change in Icelandic Agriculture." *Scandinavian Economic History Review* 41, no. 2: 101–128.

Jonsson, Gudmundur. 1999. *Hagvöxtur og iðnvæðing. Þjóðarframleiðsla á Íslandi 1870–1945* (Economic Growth and Industrialization. Gross National Product in Iceland 1870–1945). Reykjavík: National Economic Institute.

Jonsson, Gudmundur, and Magnus S. Magnusson. 1997. *Hagskinna. Iceland Historical Statistics.* Reykjavík: Statistics Iceland.

Kaldor, Nicolas. 1966. *Causes of the Slow Rate of Growth in the United Kingdom*. Cambridge, UK: Cambridge University Press.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *Manchester School* 22 (May): 139–191.

Maddison, Angus. 2004. *The World Economy: Historical Statistics*. Paris: OECD.

Marshall, Monty G., and Keith Jaggers. 2001. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2000." Available at www.cidcm.umd.edu./inscr/polity/.

Nelson, Richard R., and Edmund S. Phelps. 1966. "Investments in Humans, Technological Diffusion and Economic Growth." *American Economic Review* 56 (May): 69–75.

Ricardo, David. 1819. *Principles of Political Economy and Taxation*. London: John Murray.

Sachs, Jeffrey D., and Andrew M. Warner. 1995. "Natural Resource Abundance and Economic Growth." Rev. 1997, 1999. NBER working paper 5398, Cambridge, MA.

Schultz, Theodore. 1944. "Two Conditions Necessary for Economic Progress in Agriculture." *The Canadian Journal of Economics and Political Science* 20, no. 3: 298–311.

Tamura, Robert. 2002. "Human capital and the switch from agriculture to industry." *Journal of Economic Dynamics and Control* 27, no. 2: 207–242.

Temin, Peter. 1999. "The Golden Age of European Growth Reconsidered." *European Review of Economic History* 6: 3–22.

Temple, Jonathan. 2001. "Structural Change and Europe's Golden Age." CEPR discussion paper no. 2861.

World Bank. 1997. "Expanding the Measure of Wealth: Indicators of Environmentally Sustainable Development." Environmentally Sustainable Development Studies and Monographs Series no. 17, World Bank, Washington, DC.

World Bank. 2002. *World Development Indicators*. New York: Oxford University Press.

List of Contributors

Philippe Aghion
Harvard University

Costas Azariadis
University of California
Los Angeles

Elise S. Brezis
Bar-Ilan University
Israel

Matteo Cervellati
Universitat Pompeu Fabra
and Università di Bologna

François Crouzet
Université de Paris IV
Sorbonne

David de la Croix
IRES and CORE
Université catholique de Louvain

Theo S. Eicher
University of Washington
University of Munich
Ifo Institute of Economic
Research

Piergiuseppe Fortunato
Università di Bologna

Cecilia García-Peñalosa
CNRS and GREQAM

Thorvaldur Gylfason
University of Iceland
CEPR and CESifo

Murat Iyigun
University of Colorado

B. Zorina Khan
Bowdoin College and NBER

Giuseppe Nicoletti
OECD

Dani Rodrik
John F. Kennedy School of
Government
Harvard University

Stefano Scarpetta
World Bank

Kenneth L. Sokoloff
University of California
Los Angeles and NBER

Uwe Sunde
IZA
Bonn and University of Bonn

Utku Teksoz
University of Munich

Gylfi Zoega
University of Iceland
Birkbeck College and CEPR

Index

Abramowitz, Moses, 65
Acemoglu, Daron
  institutional growth and, 11, 17–18, 20–21
  politics and, 215–217, 222, 229, 237–238, 241
  productivity and, 69–70
Aggregation, 81
  equilibrium and, 39–43
  factor endowments and, 220–226
  institutional productivity and, 65–78
  optimal policy choice and, 167–175
  perfect markets and, 39–43
  politics and, 217
  regulation effects and, 97–110 (*see also* Regulation)
  state interference and, 108
  unified approach and, 99, 105–106
Aghion, Howitt, and Mayer-Foulkes (AMH) model, 6–8, 27, 28n12
Aghion, Philippe, 3–31, 90–91
Agriculture, 271
  chemical fertilizers and, 262
  education and, 259–262
  efficiency gains in, 252–262
  growth and, 250–251
  Harris-Todaro effect and, 260–261, 266
  Iceland and, 251–252, 262–265, 271
  investment and, 249
  labor market and, 255–256
  land value and, 249, 254–255
  Leontief production function and, 253
  market competition and, 252–253
  migration and, 272–275
  optimal backwardness and, 259–260
  Schultz effect and, 261, 266
  share of workforce in, 249–250
  Solow model and, 267
  structural change and, 265–270
  technology and, 249, 253–254, 256–258
  vs. urban industries, 249–251, 253–254
  utility and, 253–255
Alesina, A., 88, 93, 103–104, 112, 217, 239
Arellano, M., 112
Aristocracy, 194, 198
Aristotle, and recruitment of elite, 193
Army Alpha, 198
Arrow, Kenneth, 52
Australia, 100, 107
Averch, H., 88
Azariadis, Costas, 33–64

Baldwin, Matthias, 139
Bandiera, Oriana, 45
Banerjee, Abhijit, 43
Barro, Robert J., 4, 22, 81, 215, 237, 269
Barro-Lee, 45
Bayh-Dole Act, 23–24
Beck, T., 7–8
Becker, Gary, 34
Belgium, 97
Benhabib, J., 6
Ben-Porath, Yoram, 37
Bentolila, Samuel, 76
Bible, 193
Black market, 181
Blanchard, O., 88
Bloom, R., 14
Blundell, R., 9, 14
Boix, C., 215, 233, 237–239
Bond, S., 112
Boone, J., 91
Borrowing constraints, 33
  equilibrium and, 39–43
  Gini index and, 48–49
  reform response and, 47–50

Bourguignon, F., 218, 241
Boylaud, O., 93
Brazil, 182
Brezis, Elise S., 191–213
Business Environment Risk Intelligence (BERI), 68

Cambridge University, 195–196, 201
Canada, 147, 270
Capital
borrowing constraints and, 39–43
equilibrium and, 39–43
household effects and, 35–39
institutional productivity and, 65–78
perfect markets and, 39–43
retirees and, 42
Capital-labor ratio, 35
borrowing constraints and, 39–43
foreign direct investment (FDI) and, 89
household issues and, 44–52
information and communication technology (ICT) and, 87
Capital spending, 88
Caprio, Gerard, 45
Carmichael, H. Lorne, 23
Catholic Church, 194
Cervellati, Matteo, 215–248
CES technologies, 35
poverty traps and, 52–57
Chandlerian management, 195
Chemistry, 195, 262
Chenery, Hollis, 52
Cheng, L. K., 114
Chile, 45, 182
China, 159, 184–185
Cobb-Douglas production function, 40, 42, 45, 53
Competition
capital spending and, 88
entry barriers and, 88
growth and, 9–16
incumbent firms and, 10
market reform patterns and, 92–97
multifactor productivity and, 89–91, 114–115
neck-and-neck, 91
network industries and, 88
property rights and, 14–16
R&D effects and, 10–16
regulation and, 88–89
Schumpterian effect and, 11, 16
Conant, James, 198
Consolidation phase, 164
Contract enforceability, 68–69
Convergence, 27
academic institutions and, 23–26
AMH model and, 6–8
appropriate institutions and, 16–22
financial institutions and, 4–9
Gerschenkron and, 16–17
income gap and, 4
intellectual property rights and, 147–148 (*see also* Property rights)
perfect markets and, 35
technology transfer model and, 4–5
Convergence club, 4
Cost discovery phase, 161, 163
Credit rationing, 35
Crouzet, François, 191–213
Crystal Palace Exhibition, 126, 146–147
Czech Republic, 185

Dasgupta, Partha, 23
David, Paul, 23
De Gregorio, José, 34
Deininger, Klaus, 45
de la Croix, David, x, 33–64
Democratization. *See also* Politics
from above, 218–219, 232, 239, 242
from below, 218–219, 232, 242
conflict and, 225, 241
development and, 226–235
fixed rules of, 224
future research for, 242–244
guns condition and, 225, 241
historical perspective on, 235–240
income redistribution and, 218–219
oligarchy and, 218–219
public-goods provision and, 219, 224–235
technology and, 240–241
transition regimes to, 229–235
de Mooij, R. S., 105
Denmark, 263–264
Deregulation, 33–35
Development accounting, 67–68
Dewatripont, M., 24
Diaz-Giminez, Javier, 46
Dictionary of National Biography, 134
Divergence, 6
Djankov, 17
Drazen, Allan, 35–36
Durlauf, Steven N., 183

Eads, James, 139
Eastern Europe, 185
Eastman, George, 139
Ecole Centrale, 195–196
Ecole des Ponts et Chaussées, 195
Ecole Nationale d'Administration (ENA), 195, 199–202, 205, 208
Ecole Polytechnique, 195, 199
"Economic Backwardness in Historical Perspective" (Gerschenkron), 16–17
Ederveen, S., 105
Edison, Thomas, 139
Education, 9, 180
 agriculture and, 259–262
 appropriate institutions and, 20–22
 class stratification magnification and, 192
 corporate economy and, 195
 democratization of, 191, 204
 elitism and, 195–209
 factor accumulation and, 70–71
 France and, 194–196
 Gini index and, 48–49
 higher education institutions and, 191–212
 household issues and, 37–38, 45
 increased number of universities and, 191
 innovation and, 192–193
 investment in, 23–26
 meritocracy and, 191–192, 202–207
 military schools and, 194
 quality and, 191
 recruitment methodology and, 193–207
 religion and, 191, 193–194, 197, 199
 Rybczynski theorem and, 21–22
 Scholastic Aptitude Test (SAT) and, 198–199, 202, 205, 208
 Second Industrial Revolution and, 195
 U.S. inventors and, 134–146
 World War II era and, 191–192
Efficiency
 agricultural growth and, 252–262
 politics and, 218
 slack reduction and, 89–90
Egger, 113
Eicher, Theo S., 65–80
Electricity, 195
Elitism, 231
 business leaders and, 195–200
 class stratification magnification and, 192
 conflict and, 223–226
 early patent systems and, 124–134
 France and, 195, 198–202, 205, 208–209
 guns condition and, 225
 heredity and, 194
 higher education and, 191–212
 innovation and, 192–193
 meritocratic effects and, 202–207
 nepotism and, 194
 politics and, 217–218
 recruitment methodology and, 193–202
 ruling class and, 193–194
 Scholastic Aptitude Test (SAT) and, 198–199
 violence and, 194
Endogeneity, 68–69
Engerman, Stanley L., 270
Entrepreneurship, 186–187
 AMH model and, 6–8
 consolidation phase and, 164
 cost discovery phase and, 161, 163
 diversification and, 159–160
 economic environments for, 162–165
 empirical analysis and, 176–183
 growth and, 180
 imitative entry and, 160
 labor allocation and, 165–167
 market equilibrium and, 165–167
 market reform patterns and, 92–97
 optimal policy choice and, 167–175
 policy tinkering and, 160–167, 176–183
 property rights and, 3
 social infrastructure and, 3
 state ownership and, 159, 184–185
Entry costs, 17
Equality
 equilibrium and, 39–43
 optimal policy choice and, 167–175
 reform effects on household, 39–52
Esteban, J., 241
European Union (EU), 100, 111
 income gap and, 81
 liberalization and, 95–97
Evans, P., 4

Factor endowments
 productivity and, 70–77, 220–226
 public-goods provision and, 224–228
Fanelli, Jose, 33
Fidrmuc, Jan, 52
Foreign direct investment (FDI)
 capital-labor ratio and, 89
 empirical analysis of, 97–110

Foreign direct investment (FDI) (cont.)
  information and communication technology (ICT) and, 87, 101, 103, 109–110
  market reform patterns and, 92–97
  product market policies and, 113–114
  regulation and, 89, 105–107, 113–114
  restrictions on, 93, 95
  technology diffusion and, 90
  unified approach and, 99, 105–106
  United Kingdom and, 106–107
Fortunato, Piergiuseppe, 215–248
France, 97
  early patent system of, 127–129, 153nn10,11, 154n12
  elite education and, 194–202, 205, 208–209
  French Revolution effects and, 128–129, 195
  non-manufacturing sector and, 105
Friedman, Eric, 180

Galor, Oded, 34, 43, 219, 251
García-Peñalosa, Cecilia, 65–80
GDP measurement, 17, 57, 236–237
  AMH model and, 7–8
  Cobb-Douglas production function and, 53
  policy tinkering effects and, 181–183
  regulation effects and, 98, 111
Germany, 97, 105, 147, 196
Geroski, P., 9
Gerschenkron, Alexander, 16–17, 147
Ghana, 45
Giavazzi, F., 88
Gil, R., 239
Gini index, 35, 45–46, 48–49
Glaeser, E. L., 215-217, 239
Gollin, D., 76
Golub, S., 93
Gradstein, M., 218, 241
Grandes écoles (GE), 195, 199–202, 208–209, 210n19
Griffith, R., 9, 14
Griliches, Z., 91
Grossman, H., 241
Growth
  agriculture and, 250–251 (*see also* Agriculture)
  AMH model and, 6–8
  borrowing constraints and, 39–43
  compensating losers and, 50–52
  competition and, 9–16
  convergence and, 4–9, 16–27
  deregulation and, 33–34
  education and, 20–22
  entrepreneurship and, 180
  equilibrium and, 39–43
  Europe's "Golden Age" of, 251
  factor endowments and, 220–226
  Gerschenkron and, 16
  Gini index and, 35, 45–46, 48–49
  growth effect and, 34, 39–43, 62
  higher education and, 191–212
  income gap and, 81
  income per capita and, 84–88
  information and communication technology (ICT) and, 87
  innovation and, 5 (*see also* Innovation)
  institutional role in, 3–31 (*see also* Institutions)
  invention and, 146–152
  level effect and, 34
  meritocracy and, 206–207
  politics and, 215–246 (*see also* Politics)
  population, 45
  poverty traps and, 52–57
  product range and, 159–160
  property rights and, 123
  regulation effects in, 81–117 (*see also* Regulation)
  Schumpeterian theory and, 8–9
  social infrastructure and, 3
  stagnation and, 6
  structural change and, 265–271
  technology transfer model and, 4–5
Guns condition, 225, 241
Gust, C., 110
Gylfason, Thorvaldur, 249–279

Hall, Robert E.
  politics and, 216
  productivity and, 65–72, 76–78
Harberger, Al, 159
Harris, John R., 251, 253
Harris, P., 9
Harrison, Ann E., 76
Harris-Todaro effect, 260–261, 266
Harvard University, 197–198
Hausmann, Ricardo, 160, 163
Hautes Etudes Commerciales (HEC), 196
Hellmann, Thomas, 24
Henderson, Rebecca, 23–24
Holmström, B., 90

Honohan, Patrick, 45
Household issues
  borrowing constraints and, 39–43
  compensating losers and, 50–52
  education and, 37–38, 45
  generational effects and, 36
  perfect market and, 39–43, 58–62
  reform effects and, 33–39, 47–50
  retirees and, 42
Howe, Elias, 139
Howitt, P., 6–8, 14, 90
Human capital, 33, 58–62
  abilities index and, 44
  agriculture and, 249–278 (*see also* Agriculture)
  borrowing constraints and, 47–50
  compensating losers and, 50–52
  conflict and, 223–226
  education and, 199–200 (*see also* Education)
  elitism and, 195–209
  equilibrium and, 39–43
  factor accumulation and, 70–73
  household issues and, 33–39
  information and communication technology (ICT) and, 87, 101, 103, 109–110
  intellectual property rights and, 129 (*see also* Property rights)
  politics and, 218
  poverty traps and, 52–57
  productivity and, 44–45, 65–80 (*see also* Productivity)
  reform response and, 47–50
  regulation effects and, 86–87
  retirees and, 42

Iceland, 251–252, 262–265, 271
Imbs, Jean, 159
Income
  factor endowments and, 220–226
  Harris-Todaro effect and, 260–261
  household issues and, 35–39
  institutional productivity and, 65–78
  optimal policy choice and, 167–175
  output growth and, 84–88
  public-goods provision and, 224–228
  Solow residual, 67
  state role and, 222–223
Income gap, 4, 81
India, 159, 184
Indonesia, 45
Innovation, 3
  AMH model and, 6–8
  appropriate institutions and, 16–22
  basic mechanisms for, 91
  competitive markets and, 9–16
  Crystal Palace Exhibition and, 126, 146–147
  early patent systems and, 124–134
  economic development and, 146–152
  education and, 192–193
  empirical analysis of, 97–110
  financial development and, 5–8
  information and communication technology (ICT) and, 87, 109–110
  inverted-U relationship and, 12–14
  Lerner index and, 16
  meritocracy and, 206–207
  Poisson analysis and, 16
  property rights and, 123
  R&D and, 6, 10–16
  regulation and, 91
  Schumpeterian effect and, 11
  skilled workers and, 6
  stagnation and, 6
  U.S. inventors and, 134–146
Institut d'etudes politiques (IEP), 200
Institutions, 27
  academic, 23–26
  appropriate, 16–22
  black box treatment of, 69–70
  convergence and, 4–9
  development accounting and, 67–68
  direct/indirect effects of, 73–75
  early patent system and, 124–134
  endogeneity and, 68–69
  factor accumulation and, 70–77
  Gerschenkron and, 16–17
  higher education, 191–212
  historical perspective on, 235–240
  income gap and, 4
  invention and, 146–152
  labor allocation and, 165–167
  literature on, 81, 84
  measurement of, 68–69
  mortgage, 33
  output levels and, 67–70
  political, 215–246 (*see also* Politics)
  productivity and, 65–80 (*see also* Productivity)
  public sector and, 222–223
  reform and, 160–161, 184 (*see also* Reform)

Institutions (cont.)
social infrastructure and, 3
Solow residual and, 67, 69, 71–73
state role and, 222–223
technology transfer and, 4–5
Intellectual property rights. *See* Property rights
International Country Risk Guide (ICRG), 68
Invention
early patent systems and, 124–134
economic development and, 146–152
great U.S. examples of, 134–146
meritocracy and, 206–207
Inverted-U relationship, 12–14
Investment
academic institutions and, 23–26
agriculture and, 249
borrowing constraints and, 39–43, 47–50
competition and, 88–89
diversification and, 159–160
in education, 20–22
empirical analysis of, 97–110
FDI, 89 (*see also* Foreign direct investment (FDI))
growth effect and, 34
information and communication technology (ICT) and, 87, 101, 103, 109–110
intellectual property rights and, 147
in non-manufacturing, 103–105
perfect markets and, 39–43
policy tinkering and, 160–167, 176–183
productivity and, 92
regulation effects and, 82–83, 101–107 (*see also* Regulation)
retirees and, 42
unified approach and, 99, 105–106
Israel, 182
Italy, 87
Ivy League, 197–199
Iyigun, Murat, 159–188

Jaffe, Adam, 23–24
Jaggers, Keith, 266
Japan, 17, 19, 81, 97, 100, 111, 147
Jappelli, Tulio, 34
Jefferson, Thomas, 198
Johnson, L. L., 88
Johnson, Simon, 69–70, 216, 241
Jones, Charles I.
agriculture and, 251
politics and, 216
productivity and, 65–72, 76–78
Jonsson, Gudmundur, 262–264

Kaelble, H., 195
Kaldor, Nicolas, 251
Keefer, Philip, 68–69
Kehoe, Timothy, 38
Kerr, Clark, 197
Khan, B. Zorina, 123–158
Kim, Se-Jik, 34
Klette, T. J., 91
Knack, Stephen, 68–69
Koedijk, K., 86
Korea, 17, 19, 45
Kremers, J., 86
Kuznets, Simon, 147, 151
Kwan, Y. K., 114

Labor, 43, 81. *See also* Human capital
agriculture and, 249 (*see also* Agriculture)
allocation of, 165–167
borrowing constraints and, 47–50
Harris-Todaro effect and, 260–261, 266
household issues and, 35–39
Iceland and, 262–265
information and communication technology (ICT) and, 87, 101, 103, 109–110
institutional productivity and, 65–78
market equilibrium and, 165–167
migration of, 250–251, 272–275
optimal backwardness and, 259–260
optimal policy choice and, 167–175
poverty traps and, 52–57
productivity and, 65–80 (*see also* Productivity)
regulation and, 81–117 (*see also* Regulation)
Schultz effect and, 261, 266
skilled, 6, 20–22
structural change and, 265–270
Lach, Saul, 24
La Porta, R., 215
Lee, 22
Lerner index, 16
Level effect, 34, 39–43
Levine, David, 38
Levine, R., 7–8
Lévy-Leboyer, M., 199
Lewis, W. Arthur, 249, 251
Liberal arts, 191

Liberalization, 9
borrowing constraints and, 39–43
compensating losers and, 50–52
consensus on, 33
costs of, 33
dynamic simulations for, 44–52
European Union (EU) and, 95–97
growth effect and, 34, 39–43
household issues and, 33–39
level effect and, 34
market reform patterns and, 92–97
non-manufacturing sector and, 103–105
reform response and, 47–50
retirees and, 42
Washington consensus and, 56–57
Licensing. *See* Property rights
Limongi, F., 215, 237, 239
Lizzeri, A., 217, 230, 239, 241
Lo, Shih-Tse, 152
Loan-to-value ratios, 33
Loayza, N., 7–8
London Stock Exchange, 15
Lucas, R., 20

Maddison, A., 4
Madison, James, 130
Madisson, Angus, 265
Malaysia, 45, 182
Mankiw, Gregory N., 4, 65–66, 70–72, 75–76
Markets, 3
agriculture and, 252–253
black, 181
borrowing constraints and, 47–50
competition and, 9–16, 87–88
cost discovery and, 161
deregulation and, 33–34
diversification and, 159–160
empirical analysis of, 97–110
entry barriers and, 88, 103–106
equilibrium and, 39–43
European Union (EU) and, 81, 95–97, 100, 111
factor endowments and, 220–226
FDI and, 113–114 (*see also* Foreign direct investment (FDI))
household effects and, 35–39
imitative entry and, 160
inflexible, 180
information and communication technology (ICT) and, 87, 101, 103, 109–110
intellectual property rights and, 147, 151
inverted-U relationship and, 12–14
liberalization and, 33–34 (*see also* Liberalization)
non-manufacturing sector and, 103–105
perfect, 35, 39–43, 47–50, 58–62
product range and, 159–160
reform effects and, 39–57, 92–97
regulation of, 81–117 (*see also* Regulation)
Rybczynski theorem and, 21–22
state role and, 108, 222–223
tariffs and, 93, 106
trade barriers and, 93
unified approach and, 99, 105–106
U.S. patent system and, 131–132
Markusen, J. R., 99, 105, 113
Marquez, J., 110
Marshall, Monty G., 266
Maskus, K. E., 105, 113
Mauro, P., 216
Mayer-Foulkes, D., 4, 6–8
Medhora, Rohinton, 33
Meghir, C., 21–22
Meritocracy
Catholic Church and, 194
economic growth and, 206–207
higher education and, 191–192, 202–207
Scholastic Aptitude Test (SAT) and, 198–199, 202, 205, 208
social stratification effects of, 202–207
Mexico, 45
Meyer, M., 90
Military schools, 194
Minhas, B. S., 52
Minier, J. A., 216, 238
Moav, Omer, 219, 251
Mokyr, J., 241
Mortgage institutions, 33
Moses, 193
Moulton, B. R., 115
Mulligan, C. B., 239
Murray, Fiona, 23–24

National Science Foundation (NSF), 23
Nelson, Richard R., 20, 23, 259
Network industries, 88
Newman, Andrew, 43
New Zealand, 97, 100, 107, 182
Nickell, S., 9
Nicoletti, Giuseppe, 81–120
North, Douglass C., 3, 68

Norway, 263–264
Noury, Abdul, 52

Oligarchy, 218–219, 224, 230
Openness, 17
Organization for Economic Cooperation and Development (OECD), 15
  deregulation and, 33–34
  education and, 22
  empirical analysis of, 97–101
  liberalization and, 33
  market reform patterns in, 92–97
  regulation effects in, 81–117 (*see also* Regulation)
  tariffs and, 93
  unified approach and, 99, 105–106
Oxford University, 195, 201

Pagano, Marco, 34
Papaioannou, E., 216
Patent Act of 1836, 131
Patents. *See also* Property rights
  anti-commons effect and, 24
  bargain theory of, 129
  Bayh-Dole Act and, 23–24
  British system of, 125–127
  competition effects and, 14–16
  cost of, 125–126, 128–129, 156n39
  early systems of, 124–134
  French system of, 127–129
  registration system and, 127–128
  Statute of Monopolies and, 125–126
  United States and, 129–136
Penn World Tables, 45
Perotti, Enrico, 24
Persico, N., 217, 230, 239, 241
Persson, T., 217, 235
Phelps, Edmund S., 20, 259
Philosophy, 191
Piketty, Thomas, 34
Plato, and recruitment of elite, 193
Poisson technology, 16
Poland, 185
Policy, 159
  Bayh-Dole Act and, 23–24
  cost discovery and, 161
  early patent systems and, 124–134
  elitism and, 195–209
  liberalization, 9, 33 (*see also* Liberalization)
  optimal choice and, 167–175
  property rights and, 123, 131
  regulation and, 81–117 (*see also* Regulation)
  social infrastructure and, 3
  state interference and, 108
  state ownership and, 159, 184–185
  Statute of Monopolies and, 125–126
  Washington consensus and, 56–57
Politics
  conflict and, 217, 223–226, 225, 241
  corruption and, 217
  democratic vs. nondemocratic societies, 215
  economic environment and, 220–222
  efficiency and, 218
  elitism and, 217–218
  empirical analysis of, 215–216, 235–240
  factor endowments and, 220–226
  first-order effect and, 217
  guns condition and, 225, 241
  historical perspective on, 235–240
  neutrality and, 217
  oligarchy and, 218–219
  openness and, 216
  power and, 223–224
  property rights and, 217–218
  public-goods provision and, 224–228
  public sector and, 222–223
  rent-seeking and, 217
  state role and, 222–223
  timing of events and, 223–226
  voting systems and, 216, 234–235
  World War II era and, 239
Polity IV Project, 266
Poverty traps
  overlapping generations model and, 52
  reform and, 52–57
  Washington consensus and, 56–57
Pritchett, L., 4
Privatization, 83, 92–97
Productivity
  Aghion, Dewatripont, and Stein model and, 24–26
  agriculture and, 253–254
  AMH model and, 6–8
  Cobb-Douglas production function and, 40, 42, 45, 53
  competition effects and, 9–16
  consolidation phase and, 164
  contract enforceability and, 68–69
  cost discovery phase and, 161, 163
  development accounting and, 67–68
  diversification and, 159–160

domestic, 5
empirical analysis of, 97–110
factor endowments and, 70–77, 220–226
household issues and, 44–45
incumbent firms and, 10
information and communication technology (ICT) and, 87
institutional effects on, 65–80
labor allocation and, 165–167
Leontief production function and, 253
modern-sector cost and, 162
multifactor, 82–83, 86–87, 89–91, 97–110, 114–115
optimal policy choice and, 167–175
policy tinkering and, 160–167, 176–183
product range and, 159–160
public-goods provision and, 224–228
R&D and, 5
reform effects and, 52–57, 175
regulation and, 81–117 (*see also* Regulation)
Rybczynski theorem and, 21–22
slack reduction and, 89–90
Solow residual and, 67, 69, 71–73
state role and, 222–223
technology and, 4–5, 9–22, 90 (*see also* Technology)
total factor (TFP), 15, 22, 57, 76, 182–183
Property rights, 9
academic institutions and, 23–26
anti-commons effect and, 24
Bayh-Dole Act and, 23–24
competition effects and, 14–16
control rights commitment and, 24
diversification and, 160
early patent systems and, 124–134
economic development and, 146–152
growth and, 123
politics and, 217–218
social infrastructure and, 3
state ownership and, 159, 184–185
U.S. inventors and, 134–146
Przeworski, A. M., 215, 237, 239
Public sector
conflict and, 223–226
institutions and, 222–223
state role and, 222–223
timing of events and, 223–226

Ray, D., 241
Reform, 159
borrowing constraints and, 39–43, 47–50
compensating losers and, 50–52
cost discovery and, 161
dynamic simulations for, 44–52
empirical analysis of, 97–110
equilibrium and, 39–43
household issues and, 33–39
institutional, 160–161, 175
labor allocation and, 165–167
lending institution creation and, 33
liberalization and, 33 (*see also* Liberalization)
non-manufacturing sector and, 103–105
optimal policy choice and, 167–175
patent, 124–134
perfect markets and, 39–43
policy implications for, 57–58
policy tinkering and, 160–167, 176–183
poverty traps and, 52–57
productivity impact and, 175
response to, 47–50
Sachs, Warner, and Wacziarg-Welch, 181
Washington consensus and, 56–57
Regulation, 81
capital spending and, 88
competition and, 87–89
economy-wide indicators and, 93
empirical analysis of, 97–110, 112
entry barriers and, 88
evidence on economic effects of, 100–110
foreign direct investment (FDI) and, 89, 105–107, 113–114
income per capita and, 84–88
information and communication technology (ICT) and, 87, 109–110
innovation and, 91
investment and, 101–107
market reform patterns and, 92–97
microeconomic theories on, 82, 86–87
multifactor productivity and, 86–87, 89–91, 97–110, 114–115
non-manufacturing sector and, 93, 103–105
output growth and, 84–88
productivity and, 82–83
R&D effects and, 91
slack reduction and, 89–90
technology diffusion and, 90
Religion, 191, 193–194, 197, 199
Rent-seeking, 217
Research and development (R&D)
academic institutions and, 23–26

Research and development (cont.)
Aghion, Dewatripont, and Stein model and, 24–26
AMH model and, 6–8
competition effects and, 10–16
control rights commitment and, 24
Gerschenkron and, 16
incentives for, 23–24, 27
innovation and, 6
intellectual property rights and, 143 (*see also* Property rights)
inverted-U relationship and, 12–14
private, 23
productivity and, 5
regulation effects and, 91
Retirement, 42
Rigobon, R., 216, 239
Risk, 20, 69
Robinson, James A., 69–70, 216–217, 222, 229, 238, 241
Rodriguez, Francisco, 181
Rodrik, Dani, 159–188, 216, 239
Romer, David, 4, 65–66, 70–72, 75, 76
Russia, 147
Rybczynski theorem, 21–22

Sachs, Jeffrey D., 181, 268
Saint-Paul, Gilles, 76
Sala-i-Martin, X., 4, 81, 239
Savings
agriculture and, 249
deregulation and, 33–34
equilibrium and, 39–43
household issues and, 35–39
perfect markets and, 39–43
Scale effects, 99, 105–106
Scarpetta, Stefano, 81–120
Schankerman, Mark, 24
Scharfstein, D., 90
Schiantarelli, Fabio, 45
Schmidt, M. K., 90
Scholastic Aptitude Test (SAT), 198–199, 202, 205, 208
Schultz, Theodore, 257
Schultz effect, 261, 266
Schumpeterian theory, 8–9, 11, 14, 15, 16, 20, 27n1
Second Industrial Revolution, 138, 195
Sensitivity analysis, 48, 60–61
Simouraounis, G., 216
Single Market Program (SMP), 15
Skilled workers, 6, 20–22
Slack reduction, 89–90
Smith, Adam, 9
Social infrastructure, 3
Social security, 236
Sokoloff, Kenneth L., 123–158, 270
Solow, Robert, 52
Solow residual, 67, 69, 71–73, 267
South Africa, 182
Soviet Union, 185
Spain, 87
Spiegel, M., 6
Squire, Lyn, 45
Statue of Monopolies, 125–126
Stein, J., 24
Stern, Scott, 24
Stokes, S. C., 215, 233, 237, 239
Subramanian, A., 216, 239
Sunde, Uwe, 215–248
Sweden, 107, 176
Switzerland, 100

Tabellini, G., 217, 235
Taiwan, 152
Tamura, Robert, 251
Tariffs, 93, 106
Tavares, J., 216, 238
Taxes, 160, 224, 230–231, 235
Technology, 3, 240
adoption of, 90
agriculture and, 249, 253–254, 256–258
appropriate institutions and, 16–22
CES, 35, 52–57
competition effects and, 9–16
constant-returns-to-scale, 162
diffusion of, 90
early patent systems and, 124–134
education and, 20–22, 195 (*see also* Education)
factor endowments and, 220–226
financial development and, 5–8
gaps in, 114–115, 249
guns condition and, 225, 241
Harris-Todaro effect and, 260–261, 266
Iceland and, 262–265
incentives for, 107–108
information and communication (ICT), 87, 101, 103, 109–110
innovation and, 123–157 (*see also* Innovation)
labor migration and, 255–256
Leontief production function and, 253
meritocracy and, 206–207

optimal backwardness and, 259–260
property rights and, 123
public-goods provision and, 227
regulation and, 90
scale and, 40
Schultz effect and, 261, 266
skilled workers and, 6, 20–22
spillover and, 90
Technology transfer, 4–5, 114–115
Teksoz, Utku, 65–80
Temin, Peter, 198–199, 202, 250–251, 266
Theology, 191
Thomas, Robert P., 3
Tobago, 182
Todaro, Michael P., 251, 253
Trajtenberg, Manuel, 23–24
Trebbi, F., 216, 239
Tressel, T., 114–115
Trinidad, 182
Tunisia, 182
Turkey, 45

Unified approach, 99, 105–106
United Kingdom, 97, 111, 239
Crystal Palace Exhibition and, 126, 146–147
early patent system of, 125–127
education and, 195–196
foreign direct investment (FDI) and, 106–107
non-manufacturing sector and, 104–105
restrictive approach of, 101, 103
United States, 111, 270
Bayh-Dole Act and, 23–24
business-friendly approach of, 101, 103
Civil War, 134
Constitution and, 130–131
early patent system of, 129–134, 147
education and, 194–202, 205, 208
examination system and, 147
income gap and, 81
inventors of, 134–146
non-manufacturing sector and, 105
Patent Office and, 15
Scholastic Aptitude Test (SAT) and, 198–199, 202, 205, 208
University of California, 197, 198
U.S. Constitution, 130–131
U.S. Patent Office, 15

Vandenbussche, J., 21–22
Van Reenen, J., 9
Verdier, T., 218, 241
Vickers, J., 9, 90
Vietnam, 159, 184–185
Vollrath, D., 219
Voting, 216, 234–235

Wacziarg, Romain, 159, 181, 216, 238
Warner, Andrew M., 181, 268
Washington, George, 130
Washington consensus, 56–57
Weil, David N., 4
agriculture and, 251
politics and, 219
productivity and, 65–66, 70–72, 75–76
Welch, Karen Horn, 181
World Bank, 265–266
World War II era
economic development after, 239
education and, 191–192, 197, 202–202, 206–209

Zeira, Joseph, 34, 43
Zilibotti, F., 17–18, 20–21
Zimbabwe, 45
Zoega, Gylfi, 249–279