

Introduction Boris Pleskovic and Joseph E. Stiglitz

19221
April 1999

Annual

Joseph E. Stiglitz
James Tobin
Stanley Fischer
James D. Wolfensohn

Opening Addresses

World Bank

Is Geography Destiny?
Paul Krugman
John Luke Gallup and Jeffrey Sachs
with Andrew D. Mellinger

Conference

Competition and Regulation Policy
Paul Joskow
Jean-Jacques Laffont

ON Development

Financial Market Liberalization
Bruce Greenwald
Ashi Demingüç-Kunt and Enrica Detragiache

Economics

1998

Ethnic Conflict
Donald L. Horowitz
Paul Collier



Edited by **Boris Pleskovic and Joseph E. Stiglitz**

Annual
World Bank
Conference
on Development
Economics
1998

Edited by **Boris Pleskovic and Joseph E. Stiglitz**

The World Bank Washington, D.C.

© 1999 The International Bank for Reconstruction and Development / THE WORLD BANK
1818 H Street NW, Washington, DC 20433, U.S.A.
All rights reserved. Manufactured in the United States of America.

First printing April 1999

The Annual World Bank Conference on Development Economics is a forum for discussion and debate of important policy issues facing developing countries. The conferences emphasize the contribution that empirical and basic economic research can make to understanding development processes and to formulating sound development policies. Conference papers are written by researchers in and outside the World Bank. The conference series was started in 1989.

Conference papers are reviewed by the editors and are also subject to internal and external peer review. Some papers were revised after the conference, sometimes to reflect the comments by discussants or from the floor. Discussants' comments were not revised. As a result, discussants' comments may refer to elements of the paper that no longer exist in their original form. Summaries are included of floor discussions, which attempt to convey the sense and substance of what was discussed, interventions by participants from the floor, and responses by panelists. These summaries have not been reviewed by the presenters, the discussants, or other participants. Participants' affiliations identified in this volume are as of the time of the conference, April 20–21, 1998.

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and should not be attributed in any manner to the World Bank, to its affiliated organizations, or to members of its Board of Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. The boundaries, colors, denominations, and other information shown on any map in this volume do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

The material in this publication is copyrighted. The World Bank encourages dissemination of its work and will normally grant permission promptly. Permission to photocopy items for internal or personal use, for the internal or personal use of specific clients, or for educational classroom use is granted by the World Bank, provided that the appropriate fee is paid directly to Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, U.S.A., telephone 978-750-8400, fax 978-750-4470. Please contact the Copyright Clearance Center before photocopying items.

For permission to reprint individual articles or chapters, please fax your request with complete information to the Republication Department, Copyright Clearance Center, fax 978-750-4470.

All other queries on rights and licenses should be addressed to the World Bank at the address above or faxed to 202-522-2422.

Meta de Coquereaumont, Paul Holtz, and Barbara Karni edited this volume. Wendy Guyette prepared the cover and layout, and Donna Allen and Daphne Levitas proofread the book. All are with Communications Development Incorporated.

Annual World Bank Conference on Development Economics is indexed in *Human Resources Abstracts*, the *Index of Economic Articles*, the *Index of Social Science and Humanities Proceedings*, the *Index to International Statistics*, the *Public Affairs Information Service*, *Sage Public Administration Abstracts*, the *Standard Periodical Directory*, and *Ulrich's International Periodicals Directory* and online by ABI/INFORM and DIALOG.

ISSN 1020-4407

ISBN 0-8213-4321-1

Contents

Introduction <i>Boris Pleskovic and Joseph E. Stiglitz</i>	1
OPENING ADDRESS Knowledge for Development: Economic Science, Economic Policy, and Economic Advice <i>Joseph E. Stiglitz</i>	9
WELCOMING ADDRESS Rethinking Development—Principles, Approaches, and Projects <i>James D. Wolfensohn</i>	59
KEYNOTE ADDRESS Financial Globalization: Can National Currencies Survive? <i>James Tobin</i>	63
ABCDE: Past Ten Years, Next Ten Years <i>Stanley Fischer</i>	77
Is Geography Destiny?	87
The Role of Geography in Development <i>Paul Krugman</i>	89
COMMENTS	
<i>Paul A. David</i> 108	
<i>James E. Rauch</i> 118	
FLOOR DISCUSSION 124	

Geography and Economic Development	127
<i>John Luke Gallup and Jeffrey D. Sachs with Andrew D. Mellinger</i>	
COMMENTS	
<i>J. Vernon Henderson</i>	179
<i>Anthony J. Venables</i>	183
FLOOR DISCUSSION	187
Competition and Regulation Policy in Developing Countries	189
Regulatory Priorities for Infrastructure Sector Reform in Developing Countries	191
<i>Paul L. Joskow</i>	
COMMENTS	
<i>Isher Judge Ahluwalia</i>	224
<i>Kotaro Suzumura</i>	228
FLOOR DISCUSSION	233
Competition, Information, and Development	237
<i>Jean-Jacques Laffont</i>	
COMMENTS	
<i>Ana Julia Jatar</i>	258
<i>Richard Schmalensee</i>	262
FLOOR DISCUSSION	267
Financial Market Liberalization in Developing Countries: Theory and Practice	271
International Adjustment in the Face of Imperfect Financial Markets	273
<i>Bruce Greenwald</i>	
COMMENTS	
<i>Gerard Caprio Jr.</i>	290
<i>Ronald McKinnon</i>	294
FLOOR DISCUSSION	300
Financial Liberalization and Financial Fragility	303
<i>Ash Demirgüç-Kunt and Enrica Detragiache</i>	
COMMENTS	
<i>Charles A.E. Goodhart</i>	332
<i>Thomas Hellmann</i>	335
FLOOR DISCUSSION	340

Ethnic Conflicts 343

Structure and Strategy in Ethnic Conflict: A Few Steps toward Synthesis 345
Donald L. Horowitz

COMMENTS

Robert H. Bates 371

Michael Chege 377

FLOOR DISCUSSION 384

The Political Economy of Ethnicity 387
Paul Collier

COMMENTS

Pauline E. Peters 400

Jennifer Widner 406

FLOOR DISCUSSION 410

Introduction

Boris Pleskovic and Joseph E. Stiglitz

The Annual World Bank Conference on Development Economics brings together the world's leading development policy scholars and practitioners to discuss state-of-the-art thinking on vital issues facing developing and industrial countries and to revisit matters of long-standing concern. The conference is dedicated to the principle that economic science and the promotion of research, dissemination, and dialogue in economics and related fields can improve the prospects for growth and poverty reduction in developing countries. By providing a forum for discussing and disseminating cutting-edge research, the conference contributes to the global debate on development and offers an opportunity to examine lessons of experience for policy. To advance these goals, conference participants come from a broad range of academic disciplines and policymaking institutions in both developing and industrial countries.

The 10th anniversary conference, held at the World Bank on April 20 and 21, 1998, brought together some of the most prominent researchers and practitioners in development economics. The discussions focused on four areas of inquiry: the role of geography in countries' success, the role of effective competition and regulatory policies, the causes of financial crises and ways to prevent them, and the effects of ethnic diversity on democracy and growth.

New Perspectives for Development

The welcoming address by World Bank President James D. Wolfensohn, the opening remarks by World Bank Chief Economist Joseph Stiglitz, and the tenth anniversary address by International Monetary Fund First Deputy Managing Director Stanley Fischer all focused both on the role of the conference and on the changing perspectives for development.

In his welcoming address President Wolfensohn noted that the 10 years of the Annual Bank Conference on Development Economics have made an important con-

Boris Pleskovic is acting administrator, Research Advisory Staff, at the World Bank. Joseph E. Stiglitz is senior vice president, Development Economics, and chief economist at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

tribution to the change in thinking on development. Wolfensohn went on to reflect on the recent meeting at the Summit of the Americas in Santiago, Chile, noting the general belief that the Washington Consensus needs to be expanded and broadened. The new consensus should go beyond macroeconomic concerns to include issues of equity, social justice, inclusion, environmental sustainability, and participation. He elaborated on the World Bank's development agenda, arguing that education, the key to opportunity, should be at its center. Other important concerns of the broader development agenda are health, infrastructure, and the justice system. Reflecting on the changes that have occurred in the past 10 years, Wolfensohn noted that there are a greater number of democratic governments today, that 5 billion instead of 1 billion people are in the market system, and that such changes create an enormous opportunity to collaborate with civil society. Wolfensohn concluded that the agenda of the conference reflects changes in the world, as well as within the World Bank, by extending the development debate to issues such as geography, political economy, and ethnic conflicts—issues that both incorporate and extend beyond economics. Greater knowledge about these issues is essential to the results that we all want to achieve.

In his opening address Joseph E. Stiglitz echoed one of the themes struck by President Wolfensohn, namely the need for the Bank, and the development community more generally, to employ new and better models as well as more instruments to pursue a broader agenda than has been pursued in the past. That agenda needs to go beyond increasing GDP to raising living standards and ensuring equitable, democratic, and sustainable development.

Stiglitz also emphasized the basic theme underlying the conference: the importance of research—a combination of theory and empirics—as foundations for sound policy. Because the links between diagnosis and policy are so close, Stiglitz argued that we must be especially careful not to confuse ideology with economic science in our interpretation of complex events such as the East Asian crisis. He suggested that all too often views are shaped more by ideology than by solid research, illustrating the proposition by several themes drawn from recent policy debates. As an example, he noted that emerging from the careful study of recent development experiences were increasing doubts over the Washington Consensus or the “neoliberal model of development,” which accorded the government a minimal role (essentially that of ensuring macroeconomic stability, especially price stability), while remaining hands-off in all other areas and allowing markets to work through trade liberalization and privatization. Some countries that have followed the dictates of the neoliberal model are still waiting for growth to improve, while others that have ignored these dictates have experienced some of the highest sustained growth rates ever.

The recent crisis in East Asia also illustrates both the role ideology has played and the need for more research. The experience shows that the elements emphasized by the Washington Consensus did not protect a country from having a crisis. Maintaining strong financial institutions is every bit as important as fighting inflation. Strong financial institutions in turn require strong financial regulation—not the mindless kind of deregulation that has been shown to cause problems around

the world. By the same token, liberalization of short-term capital flows can be an important source of instability. More generally, thinking more deeply about finance, financial institutions, and the links between finance and the macroeconomy leads to a better understanding of, say, the possible effects of raising interest rates in response to a crisis. After questioning from this perspective the theoretical coherence of the traditional arguments that raising interest rates will lead to a strengthened currency, he presents an alternative argument, particularly applicable for countries (such as those in East Asia) with high short-term leverage. Increasing interest rates leads to more bankruptcies, lowering the certainty equivalent expected return, and that, reinforced by the resulting slowdown in the economy, undermines confidence, discourages new capital from entering, and may actually contribute to capital flight.

Stiglitz concluded by presenting an *economic analysis of advisers*—of the incentives that they face and how that affects both the advice that they give and how they give it (including their openness and transparency). He suggested that there are strong incentives for secrecy, which both undermine democratic processes and, by limiting the scope for “second opinions” and open dialogue, may adversely affect the quality of the decisions made. In general, there is not a single Pareto-dominant policy. He argued that policy advisers need to be open about the uncertainties concerning alternative policies—the recognition of, and indeed quantification of, uncertainty was an essential component of the scientific process—and that advisers should return to the role advocated by John Neville Keynes, namely laying out the consequences of alternative policies and leaving the ultimate choices to the political process.

In his 10th anniversary address Stanley Fischer, the founder of the conference series in 1989, reflected on the development consensus at the beginning of the decade. The lessons of the transition process and the recent financial crises would not change the basic message much, he said, although some issues would receive more emphasis. Globalization, however, could produce major changes in the message. As globalization continues to bring profound changes, Fischer argued, actions on a broad front will be required in most developing countries. Even some advanced economies will need to improve institutions and policies to meet international standards.

Nobel Prize Winner James Tobin presented the keynote address, discussing the implications of financial market globalization for national currencies. He lamented that current policy discussions did not fully integrate the important lessons that should have been drawn from past experiences. Acknowledging that liberalizations have opened economies to increased foreign investment, he also noted that countries are urged to further increase cross-border financial liberalization, despite the evident fact that excessive private external short-term debt was, if not the cause of the crisis, a serious aggravator of it. This is especially clear in fixed exchange rate or adjustable peg regimes. Short-term private bank debts in hard currencies were fatal to the Indonesian and Korean currencies in the last months of 1997, for example. These debts entailed severe negative externalities, bringing about currency crises

devastating to entire economies. An obvious precaution, he argued, is to limit net indebtedness (particularly short-term) in hard currency allowed to any private bank.

Must Geography Be Destiny?

The recent surge of interest in the role of economic geography in development has split along two seemingly contradictory paths. One approach emphasizes the role of inherent features of the landscape in shaping development patterns. Paul Krugman surveyed the alternative approach, which considered how more or less symmetric locations can end up playing very different economic roles. He argued that natural features matter mainly insofar as they act as seeds around which cumulative processes crystallize, suggesting that while geography might have been destiny in the past, it need not be so in the future. The tension between “centripetal” forces, which lead to increased population concentration, such as forward and backward linkages in production and increasing returns in transportation, and “centrifugal” forces, such as factor immobility and land rents, can produce a process of self-organization, an “equilibrium” distribution of economic activity. This process can occur at various levels, as Krugman showed through geographically based models that account for the division of the world into industrial and nonindustrial countries, the emergence of regional inequality within developing countries, and the growth of giant urban centers. Krugman also looked at policy issues, asking why it is so hard to draw policy conclusions from these models.

Starting with the grim reality that two centuries after the start of modern economic growth, a large part of the world remains mired in poverty, John Luke Gallup and Jeffrey Sachs took an opposite tack: they stressed the importance of the inherent features of the landscape. Asking difficult questions about the failure of tropical regions to thrive and the high rate of population growth today in countries least equipped for rapid development, they echoed Stiglitz’s call for development assistance to focus on transmitting good policies and on helping create strong political and economic institutions. Although geography is not everything, Gallup and Sachs argued, it is clearly of prime importance for economic development. This presents pessimistic prospects for countries endowed with unfavorable “geographies.” Elaborating on the importance of geography for development, Gallup and Sachs examined formal models of geography and development and their linkages in cross-country empirical models of growth, population density, and per capita income. They explored in detail the sources of population density and its relationship to income levels, raising several key (but neglected) issues: Why does development lag in tropical regions? Why does growth favor coastal regions? Why is population density advantageous for growth along coasts but not in the hinterland? These fundamental questions can be answered only by reincorporating geographic variables into econometric and theoretical studies of cross-country economic growth. In Africa, for example, basic geographic realities have been omitted from cross-country work, although their implications for policy are vital. While policymakers cannot change a country’s geography, they need to consider a wider range of problems informed by geography, including the special problems of landlocked countries, the challenges

of low agricultural productivity and endemic disease in the tropics, and pressures for mass migration. Such refocusing will lead to better-formulated policies and more effective aid in those areas of the world that need it most.

On the Role and Design of Regulation and Competition Policy

Economical and reliable supplies of such basic infrastructure services as water, electricity, telecommunications, and transportation are crucial to economic growth and development. Paul Joskow noted that in many developing countries these services are supplied by state-owned monopolies, whose performance has historically been poor. As awareness of these shortcomings mounted, a number of countries introduced (or planned) major restructuring programs. These programs often began by privatizing—shifting supply responsibility to private enterprises—and introducing competition, relying on market mechanisms to allocate resources to and within potentially competitive segments of infrastructure sectors. Despite these measures, important segments continue to resemble natural monopolies, requiring ongoing regulation aimed at fostering open and nondiscriminatory access to the service network facilities for new competitors. Accordingly, the success of infrastructure reform depends in large part on creating sound regulatory institutions, guided by the principles of independence, transparency, and expertise. These institutions must be created at the outset of reform, argued Joskow, not as an afterthought, and they should include efficient regulatory agencies, appropriate regulatory goals, a clear role for the regulator, and a credible environment for investors. Joskow paid special attention to the problems of developing countries in implementing effective regulatory reform. He noted that, in general, privatizing infrastructure services works best in an economic environment that protects consumers from exploitation, guards against cross-subsidization and discrimination among competitors, and facilitates competition.

While agreeing that the pressure of competition is generally favorable, Jean-Jacques Laffont also discussed some of the limitations of competition policy in less developed countries, including those associated with implementation problems. Although competition generally has a positive effect, care must be taken in implementing it where institutions are weak. In some sectors, particularly the financial sector, regulation is an essential complement to increased competitiveness, and liberalization can proceed safely only if strong regulatory institutions are in place. While little can be expected in countries where political will is lacking, elsewhere policymakers must bear in mind that a strong state and competition go hand in hand. And as important as stimulating effective competition policy may be, said Laffont, so too is building strong democratic institutions that promote the welfare of consumers and avoid interference by interest groups.

What Causes Financial Crises—and Can We Prevent Them?

The sudden crises in Mexico and Asia in recent years revealed the shortcomings of traditional open economy models, which assume efficient global financial markets,

argued Bruce Greenwald. In these models, global financial and product markets tend to mitigate rather than intensify such crises, suggesting that market imperfections must be at work in these real world crises. Understanding these imperfections helps in formulating effective alternative public policies. Greenwald examined one broad form of financial market imperfection, the asymmetry of information between local firms and bank management, which are usually well-informed about a firm's prospect, and outside investors, which may not be. This form of information imperfection can severely limit financial market transactions. Amplifying, rather than attenuating local disturbances, these imperfections harm prospects for long-term development. And, Greenwald demonstrated, they account for many characteristics of the recent crises. As an alternative to the traditional full-information model, Greenwald offered a model assuming imperfect information, in which an initial shock is followed by a sharp downturn then a slow recovery. In this model certain policies work well in both the short and the long term. Development of effective financial intermediation, for example, is shown to accelerate recovery from a downturn and to offer an effective means of deploying capital to its highest return applications in support of long-term growth. Thus, Greenwald argued, it should be a primary aim of development assistance. He argued that other interventions—especially aggressively restrictive monetary policies designed to raise domestic interest rates—are likely to do more harm than good.

Aslı Demirgüç-Kunt and Enrica Detragiache also stressed the importance of a strong institutional setting before undertaking financial market reform. Liberalizing financial markets has been high on the economic policy agenda of many countries over the past 30 years. But during this period the frequency of systemic banking problems has increased markedly, raising the possibility that greater fragility might be a consequence of liberalization. Using a large panel data set covering 53 industrial and developing economies during 1980–95, Demirgüç-Kunt and Detragiache found that financial fragility is affected by a host of factors, including adverse macroeconomic developments, poor macroeconomic policies, and vulnerability to sudden capital outflows. But even when these factors are controlled for, financial liberalization exerts a significant independent negative effect on the stability of the banking sector. But this effect is weaker when the institutional environment is strong. In particular, respect for the rule of law, a low level of corruption, and good contract enforcement can curb the adverse effects of liberalization on the financial system. These findings suggested the need for a gradual approach to financial liberalization where institutions are not sufficiently developed, even after macroeconomic stabilization has been achieved. They recommended that institutional development begin early in the liberalization process, because strong institutions cannot be created overnight.

How Ethnic Diversity Affects Democracy and Growth

Ethnicity has long been recognized as a powerful force in the politics of many countries, profoundly affecting prospects for democracy, economic development, and the

distribution of public goods. But there is little agreement about the causes of ethnic conflict, noted Donald L. Horowitz. Disagreements relate less to the facts surrounding ethnic conflict than to ways of interpreting those facts. “Hard” and “soft” schools of thought advance polarized views—that ethnic groups are firmly bounded communities inclined toward ethnocentrism or that they are malleable entities with solidarity based on the pursuit of member interests. Horowitz sought a common ground for theories of ethnic conflict, arguing that these perspectives may not be as mutually incompatible as they seem. He examined 10 competing explanations for ethnic conflict, ranging from theories of primordial affiliation to evolutionary concepts of kin selection and genetic imprint. He proposed a view of ethnicity grounded in a deep sense of attachment of individuals to ethnic groups and supported by group members’ feelings of similarity among themselves and of differences from others. In assessing ethnic political behavior in democracies, Horowitz found that the tendencies of groups to separate from other groups and of democratic politics to bifurcate make it easy for political leaders to mobilize along ethnic lines. As a result, it will be hard to achieve high levels of inclusiveness.

Paul Collier investigated the effect of diversity on economic performance. According to recent evidence, Collier noted, ethnic diversity appears to have detrimental microeconomic effects because it tends to harm public sector performance, increase patronage, and reduce trust among individuals. But diversity’s effect on overall economic growth is less clear-cut because it depends on the political environment. Analyzing ethnicity from a highly aggregated, cross-country perspective, Collier found that diversity is not damaging to growth in a democracy because democracy tends to eliminate the potentially negative effects of diversity on growth by disciplining governments to deliver reasonable economic policies and by providing a framework in which groups can negotiate mutually beneficial outcomes. And high diversity protects society from violent conflict even more than a high degree of homogeneity does, because of the higher coordination costs of rebellion for a society with many ethnic groups than for a society with only a few ethnic groups. But ethnic diversity can be highly damaging to growth where political rights are limited—and the risk of violent conflict is also greater. The societies most at risk of conflict are those in the middle of the range of ethnic diversity. That is, the most conflict-prone societies are those with intermediate levels of diversity. Collier also argued that the level of income is an important determinant of the likelihood of conflict—conflict is more likely to arise in poorer countries with a given degree of diversity than in richer countries. Once a society has reached full-scale civil war, the balance of influences appears to change, said Collier, and the persistence of conflict and the sustainability of a settlement may depend more on ethnic composition than on income and political rights. Thus, in societies torn by violence, maintaining or creating extreme ethnic fractionalization may be the best hope for peace. Collier concluded that highly diverse societies such as are typical in Africa are potentially a source of strength, so long as countries have a high degree of political rights. Collier’s study showed why World Bank support to developing countries needs to be informed by an understanding of social and political issues. And that, he

reflected, will require research in a broad range of disciplines, from economics to political science to anthropology.

* * * * *

As in previous years, the planning and organization of the 1998 conference was a joint effort. Special thanks are due to Gregory Ingram, the administrator of the Research Advisory Staff, for his continued support of the conference series. We would also like to thank other staff members, in particular several anonymous reviewers, and the conference coordinators, Mantejwinder Jandu and Jean Gray Ponchamni, whose excellent organizational skills kept the conference on track. Finally we thank the editorial staff, especially Meta de Coquereaumont, Paul Holtz, Barbara Karni, and Daphne Levitas, and layout designer Wendy Guyette.

Knowledge for Development: Economic Science, Economic Policy, and Economic Advice

Joseph E. Stiglitz

At this conference last year, and during the following year, I have tried to set out a vision of a development agenda that embraces broader objectives and more instruments than have often been the focus of development efforts of the past (see World Bank 1997a). I stressed that we are concerned not just with increasing GDP, but also with improving living standards, which includes better health and education. We need sustainable development, not only in the standard sense of protecting the environment for future generations, but also in the sense of being rooted in policies that can withstand the vicissitudes of the political process. We need egalitarian development, in which the fruits of the growing prosperity are shared widely within the population. And finally, we need democratic development, in which citizens participate meaningfully in the collective decisions that affect their lives and livelihoods in so many ways.

To accomplish these goals, I argued that we need to go beyond the kinds of policies that were at the center of discussion a decade ago, when this conference was inaugurated.¹ Those policies were based on a model that accords the government a minimal role, essentially one of ensuring macroeconomic stability, with an emphasis on price stability, while getting out of the way by liberalizing trade, privatizing, and allowing markets to set prices. Many of these policies are necessary for markets to work well and contribute to economic success, but they are far from sufficient. Some aspects of the model might not even be necessary conditions for strong growth, and if undertaken without accompanying measures, say to ensure competition in relevant areas of the economy, they may not bring many gains and could even lead to setbacks. Some countries have closely followed the dictates of the model, but have not seen especially strong economic performance. Other countries have ignored many of the dictates—at least with respect to the crucial details of sequencing—and have experienced some of the highest rates of sustained growth the world has ever seen.² By focusing on too narrow a set of objectives—increasing GDP—other objectives,

Joseph E. Stiglitz is senior vice president, Development Economics, and chief economist at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

such as equity, may have been sacrificed (see Kanbur 1987). By attempting to force rapid transformation—often by imposing strong conditionality for the receipt of vitally needed assistance—not only were democratic processes undermined, but political sustainability was frequently weakened (see Collier 1997).

In their zeal to get countries to adopt the policies recommended by that model, some of its more ardent supporters often turned what should have been simply a means into an end in itself. There is, for instance, a growing consensus that a vibrant market economy should be the center of a successful development strategy and that this in turn requires effective incentives, which depend on well-defined property rights and strong competition. Privatization and trade liberalization have often been employed as means toward achieving these ends, but we should not forget that complementary reforms may be needed. Turning a state monopoly into a private monopoly, for instance, is unlikely to help create a more dynamic market economy.

China has shown that an economy might achieve more effective growth by focusing first on competition, leaving privatization until later. Although in 1978 China accounted for only 40 percent of the population of low-income countries and one-quarter of their GDP, since that time it has accounted for two-thirds of the aggregate increase in GDP for low-income economies. If the 30 provinces of China—many of which are larger than all but the largest developing countries—were treated as separate countries, 20 of the fastest growing economies in the past two decades would all have been Chinese provinces.³

By contrast, competition remains thwarted in many of the former socialist economies that pursued privatization first. The improved incentives associated with privatization and the elimination of distortions associated with central planning should have led to increased efficiency of resource allocation and thereby to increased GDP. But according to official statistics, the “successful” reformers in Eastern Europe have the same per capita GDP today that they had in 1989, and Russia has seen its average income, officially measured, cut in half since then.⁴

The lack of emphasis on competition is only one of the lacunae of the Washington Consensus model. A growing consensus holds that there was a broader lacuna: the lack of emphasis on institutional infrastructure, including not only competition policy, but also legal structures that enforce contracts, implement bankruptcy, and ensure sound financial institutions. For instance, the recent experience of the Czech Republic has demonstrated that in the absence of an effective legal structure (both the appropriate laws and their enforcement), private rent-seeking can be every bit as powerful, and perhaps even more distortionary, than public rent-seeking (see, for instance, Weiss and Nikitin 1998). The experience in Russia shows that, without the appropriate institutional infrastructure, privatization provides incentives for asset stripping—and shipping wealth abroad—rather than for wealth creation.⁵ Similarly, the East Asian crisis has taught us that inadequate financial regulation and weak financial institutions can give rise to macroeconomic instability as bad as or worse than that associated with inadequate fiscal and monetary policies.

Indeed, a vast body of research on the microeconomic foundations of macroeconomics over the past 15 years has demonstrated the importance of understanding

the links between financial markets and the real economy, as well as the key role that financial markets play in economic fluctuations and growth.⁶ This research was not motivated by the recent wave of financial crises, and it certainly was not precognition of the impending turmoil in East Asia. Rather, it was based on more fundamental developments in economic theory and empirics over the past 25 years, such as those in the areas of incomplete information and incomplete markets.⁷ This strand of the literature stands in marked contrast to new classical and real business cycle models whose microfoundations are based on assumptions of representative agents, complete markets, perfect competition and market clearing (so there is neither unemployment nor credit rationing), and complete rationality.⁸ As questionable as those models might be for an advanced industrial economy, they are clearly not appropriate for developing countries, which are our focus here.

In my keynote address to this conference last year I emphasized the importance of financial and other institutions, competition policy, and a variety of other areas that were either omitted or underemphasized by the Washington Consensus. In my talk today I would like to reflect on another area that received insufficient attention.

The Role of Knowledge and Scientific Methodology

This year's *World Development Report*, subtitled *Knowledge for Development* (World Bank 1998c), reflects the growing awareness of the importance of knowledge for development. The report has two central themes. First, one of the major ways in which less developed countries are separated from more developed countries is by their level of knowledge. Successful development thus entails not only closing the gap in physical or even human capital, but also closing the gap in knowledge. That is why the World Bank is increasingly thinking of itself as a knowledge bank, not just a bank for facilitating the transfer of capital to developing countries.⁹ It is not just knowledge of production processes that matters; good institutions and economic management are essential ingredients of successful development efforts.¹⁰ Good institutions and effective management will not only facilitate the transfer of knowledge, but will also enhance the likelihood that such knowledge will be used effectively.

The second theme of the *World Development Report* is that imperfections of information impede the workings of markets. Traditional models assuming perfect information may not provide a good approximation for understanding land, labor, product, and capital markets. They fail to explain essential parts of the institutional structure and performance of all economies. These shortcomings are especially important in developing countries, which suffer from an even greater lack of information. Many institutions in those countries can be viewed as responses to these imperfections of information. And while these economies respond to market forces, their behavior is markedly different from what the standard competitive model would have predicted.¹¹ Effective development strategies must take into account these institutions and strive to improve both the flows of information and the economy's capacity to produce, transmit, disseminate, process, absorb, and use informa-

tion. At the same time, these strategies must recognize that information will inevitably remain imperfect, that market failures associated with information failures will continue to be important in developing countries, and that these imperfections in turn have to be taken into account in the design of policies.

Many have described the current state of the world as a knowledge revolution, an acceleration in the pace of generation of new ideas and of the decline in the cost of disseminating these ideas. The recognition of the centrality of knowledge itself is an important development, perhaps a revolution in its own right.¹² The knowledge revolution and the technology revolution that has helped support it have not only changed perspectives on the role of knowledge in development, but have also opened up new opportunities. Remote villages can be interconnected through the Internet and have access to a knowledge base beyond the dreams of anyone living a century ago. Distance education can bring the finest teachers in the world to countries around the globe. But the knowledge revolution has also brought new challenges: those without access to these resources and the knowledge and training to use them may fall even further behind.

This knowledge revolution is based on modern science and technology, and the change in modes of thinking that modern science brings with it constitutes one of the major aspects of the development process. The scientific mentality does not just accept traditions and received wisdom. It questions beliefs, asking, what is the evidence? And how do we know what we believe? Once we begin to question traditional beliefs, change is inevitable. The scientific process itself is one of constant change, with one idea leading to another, often challenging the previous one as new evidence overturns existing hypotheses and beliefs. Science is thus necessarily iconoclastic—and unsettling.

Science also recognizes its limitations. Uncertainty is endemic, and science attempts to be precise about the degree of uncertainty associated with each statement. It does not pretend to assert with confidence what it only believes on the basis of partial information. The recognition of this uncertainty counsels humility, especially when those supposedly with scientific knowledge apply their imperfect knowledge to real world situations affecting millions of individuals.

Even in the most advanced industrial countries the scientific approach has made only limited inroads. Any reader of management textbooks is well aware that a substantial fraction of the nostrums, as plausible and reasonable as they may be, are not based on scientific evidence. One highly popular book touted its magic formula by showing the role it had played in the success of 43 of America's most successful enterprises; yet just a few years later the performance of these highly praised companies was little if any better than that of a randomly selected list.¹³ To people not committed to the scientific approach, these experiences are hardly a setback: the authors were quickly back in the market with a new formula for success—and new speaker's fees—to explain the new approach to managerial success. Those who eschew the scientific methodology are ready with a thousand and one explanations for the failure of their predictions, many of them reminiscent of the explanations used by medieval doctors engaged in bloodletting: the patient did not precisely fol-

low the prescription, she went off the regimen too early, or the disease would have been even worse without the medicine.

Unfortunately, economics provides few opportunities for the controlled experiments that characterize much of modern science. In this sense it is much like astronomy, left to make inferences from the natural experiments that nature from time to time affords. But it is at an even more disadvantaged position: while astronomers observe a universe governed by unchanging laws, much of economics is concerned with the behavior and responses within *today's* economy. And today's economy differs from that of yesterday and certainly from that of a half century or a century ago. To what extent are those differences salient for making inferences and predictions relevant for today? Worse still, the multiplicity of factors that affect outcomes and the few truly exogenous variables mean that economists are almost always dealing with underidentified systems. These problems are compounded when we are dealing with economic crises, which fortunately are relatively infrequent. (I am not recommending that we solve this problem by increasing the frequency of crises, as helpful as that would be to economic researchers!)

But modern science is, fortunately, richer and draws on sounder bases than simply making inferences from, say, mindless cross-country regressions. Newton's inferences about the existence of the force of gravity and the fact that it varies inversely with the square of distance were based not on a statistical regression but on a careful analysis of the implication of that simple hypothesis for the nature of planetary orbits. A single, striking observation can overthrow a theory (though, to be sure, there are always those who strive to make Ptolemaic revisions to existing theories to preserve existing doctrines). The Great Depression served to overthrow classical economics, which was predicated on the *assumption* that all markets, including the labor market, clear. Although some today see the reduced employment that occurred in the Great Depression simply as a decision by large numbers of labor market participants to increase their enjoyment of leisure,¹⁴ there are enough pieces of evidence against that hypothesis that it has won few adherents outside of a few American universities.

Thus, in making inferences about economies, we bring to bear a wealth of information, theory, statistics, econometrics, and historical experience. We weed out hypotheses through natural experiments or by pointing out their intellectual inconsistencies. Reasonable people may, of course, differ in how they aggregate all these disparate pieces of information, a point to which I return later. Worse still, the question is often whether a particular model is "good enough" for the question being addressed. All models represent simplifications. The proper and often difficult question is, when will a particular simplification provide the key insights, and when will it lead to misleading policy positions?

In many ways this process should be easier than ever before. In the past economic arguments were dominated by grand ideological debates such as that between capitalism and socialism. Adam Smith's revolutionary idea that the wealth of nations was measured by its income, not its gold reserves, helped bring an end to mercantilism (although mercantilist notions still seem to underpin popular fears about trade

deficits). And Karl Marx's ideas—an alternative to capitalism—provided the intellectual basis for an economic system that until recently dominated the lives of 1.5 billion people.

Today, however, the grand ideological battles are over. In most of the world, there is almost universal agreement that markets should be at the center of any vital economy. And few doubt that governments should play an important role as a complement to markets, through competition policy, regulation, funding for education, and support for research and development, to name just a few key roles for government. Within this broad agreement, however, there are continuing debates over more technical matters, such as how to respond to economic crises, how to undertake financial reform, and what is the proper scope and sequencing of privatization. Such debates can have a huge effect on the lives of people throughout the world. These disputes reflect the uncertainty to which I referred earlier: knowledge—theory and evidence—typically is simply insufficient to fully resolve these disputes. While it is imperative to recognize this uncertainty, we should not go to the other extreme of absolute agnosticism. Some perspectives can be rejected outright, and some can be shown to be more “plausible” than others.

In practice, there are often large differences in the understanding of our beliefs about economic issues. The purpose of economic science is to narrow these differences by subjecting the positions and beliefs to rigorous analysis, statistical tests, and vigorous debate. Most economists agree that the outcome of this process should not be dependent on the values of the participants in the debate. In practice, however, it is not uncommon to find, for instance, that a researcher of libertarian leanings will uncover evidence that large governments are bad for growth. As long as there is uncertainty, and there will always be uncertainty, it will be impossible to fully separate values from purely scientific discussions.¹⁵ Once we accept this conclusion, we realize that in giving advice we are not just purveying economic science. And that realization requires us to think seriously about how we give advice and what incentives shape the advice we give.

The experience with communism has shown how economic ideas can have a profound effect on the lives of billions of people. This makes it absolutely essential that we do our best to try to understand the scientific basis for our theories and evidence. Today, I will illustrate this by discussing economic ideas in three areas of “applied” economic knowledge: the causes of and responses to the East Asian crisis,¹⁶ financial market liberalization, and privatization. I will also consider economic “advice”: the role of the policy adviser, the incentives facing advisers, and how *knowledge* about the role and incentives of advisers and the responses of recipients can improve our advice and its effectiveness.

Applied Economic Knowledge I: Causes of and Responses to the East Asian Crisis

Most of the governments in East Asia and their advisers, formal and informal, share the same goal: the speedy resumption of stability and growth in the region. But the

recommended means for achieving this goal have sometimes differed. To some degree this mirrors the different beliefs, values, and interests of the participants in these discussions. The solutions proposed by, say, a worker and a banker will often be based on different objectives. The influence of these factors on the analysis of a phenomenon as complex as the East Asian crisis is evidenced by the close link between diagnosis and prescription. Because the links between diagnosis and policy are so close, we should be especially wary not to confuse ideology with economic science. For instance, in explaining the causes of the crisis, those who believe on ideological grounds that government should play a smaller role in the economy put greater emphasis on excessive government interventions than on inadequate government regulation of the financial sector. There is a certain irony in these positions: not long ago these same critics claimed that East Asia's remarkable growth was due largely to the fact that government did not intervene!

The "Asian Model," Economic Distortions, and the Causes of the Crisis

Lip service is now universally paid to the fact that the current crisis differs markedly from earlier crises. It is private sector debt, especially short-term debt, not public sector profligacy that was at the root of the problem. The countries in East Asia had low inflation, high savings rates, and a fiscal stance and public debt to GDP ratios that were the envy of even the most responsible industrial economies. By now, economists have joined journalists in describing key attributes of these economies that they claim contributed to the current problems: misguided investment, weak financial sectors, lack of transparency, inadequate corporate governance, inappropriate exchange rate policies, and so on.

But lists of factors do not constitute a causal analysis. Many of the listed factors are neither necessary nor sufficient for a crisis. Some countries with less transparency, more political cronyism, weaker financial systems, and larger fiscal deficits not only avoided a crisis, but were also relatively unaffected by contagion. In the past we have seen countries with high degrees of transparency go through major financial crises, most notably the Scandinavian countries in the late 1980s and early 1990s.¹⁷

The debate about the causes of the crises have been confused by a number of fallacies. In some cases the evidence of the allegations of particular weaknesses in the economy is, at best, weak. And even when weaknesses have been identified, there is often a confusion between the presence of a weakness and a "cause" of the crisis. We should not fall into the *post hoc ergo propter hoc* fallacy of identifying any problem as a cause of the crisis. After all, would anyone suggest that the distorted U.S. agricultural sector played a significant role in the country's savings and loan crisis of the 1980s?¹⁸

Indeed, it is particularly ironic that in several areas of alleged weaknesses, at least some of the affected countries had been showing signs of "improvement" in the past decade. For instance, at least by standard measures of transparency, several of the countries not only had ratings that were near, or even above, "average," but a cou-

ple showed marked improvements (see Furman and Stiglitz 1998); and the extent of government involvement in the economy had decreased, at least in the Republic of Korea. In particular, many weaknesses may lower an economy's per capita income, but this does not make it more vulnerable to shocks, more crisis prone. A bank that only lent on the basis of 150 percent collateral might not contribute much to an economy's growth, but it might actually lower vulnerability. This failure to distinguish between factors that lowered per capita GDP and factors that contributed to vulnerability is one of the major sources of confusion in the policy debate.

Even if serious distortions contributed to the vulnerability that resulted in the crisis in East Asia, it does not necessarily follow that eliminating them is the correct way to respond to the immediate crisis. If the United States had removed agricultural subsidies in the midst of the savings and loan crisis, the result would have been plummeting land prices, exacerbating the burst of the real estate bubble and thus worsening the financial crisis. By the same token eliminating the huge energy subsidies embedded in the U.S. tax code would be good for both the environment and the economy, but to have eliminated these distortions in the midst of the financial sector problems in Texas in the mid-1980s would have been, to say the least, foolish. The crises in the Scandinavian countries illustrate similar issues. In Sweden, for instance, the elimination of distortions favoring debt in the tax system, after a period of large-scale debt buildup, was one of the important factors contributing to the financial crisis in the early 1990s (see Bäckström 1997).

To be sure, a weaker economy, in any dimension, can contribute to a weakening of the financial sector. Policy analysts must ask two questions. First, to what extent does a *particular* distortion lead to the *particular* problem? Here, the particular problem is vulnerability to shocks—to what extent did the distortion lead to both a higher likelihood of an adverse shock (and in particular, to the adverse shocks currently buffeting the East Asia region) and to a lower ability to absorb the shock? Second, to what extent would eliminating that distortion resolve that particular problem?¹⁹

ALLEGATIONS OF FUNDAMENTAL WEAKNESSES. Many of the allegations of fundamental weaknesses—and some of the particular accusations concerning specific shortfalls—are, at best tenuous. Were the countries of East Asia highly vulnerable? There are two ways to approach this question. First, did they experience more frequent downturns or crises than other countries? The answer to that is unambiguous. There have, of course, been financial crises in the past, like that in Korea in 1980 and Thailand in 1983. What is striking is the fact that since 1965, Indonesia, Korea, and Malaysia have each seen only one year of negative growth; and Thailand has not seen any. During that period GDP fell in five separate years in both the United States and the United Kingdom (see World Bank 1998b). Thus the East Asia system had delivered not only greater growth but less volatility (measured by vulnerability to outside shocks) than the economic regimes found elsewhere in the world. If the countries were “vulnerable,” it was a newly acquired vulnerability—perhaps a deviation from the “Asian model” that had ensured them such high growth and stability over the

preceding three decades. For this, there is some theory and evidence, to which I will turn later: the financial and capital account liberalization that they were induced to undertake during the past decade.

Another approach is to use statistical models. Every country has strengths and weaknesses. Vulnerability means that the “balance of factors” is such as to render the country crisis prone. The unfortunate plethora of crises in the past quarter century provides a way of assessing whether, on balance, the countries of East Asia were more likely to have such problems in 1997. While the evidence thus suggests strongly that the allegations of overall weaknesses in these economies are based more on ideology than evidence, similar questions concern allegations of particular weaknesses. I have time here to consider briefly only four such accusations.

ALLEGATIONS OF SPECIFIC WEAKNESSES. The first is mismanaged exchange rates. Furman and Stiglitz (1998) show not only that in some cases, such as Korea, there is little evidence for an overvalued exchange rate. They also show that in other cases, such as Thailand, greater flexibility of exchange rates would have resulted in an appreciation of the currency and smaller reserves during the period of rapid capital inflow; a combination that might well have contributed to an even greater crisis. (This is not, however, to suggest that, in general, better exchange rate policies do not reduce the likelihood of a crisis; only that even the best managed exchange rate policies do not inoculate fully against that possibility.)

The second concerns the allegations that lack of openness contributed to the crisis. Those who see this as the problem have tried to push the East Asian countries to open their markets further. But the irony here is that there is a consensus that the success of the East Asian economies was based on their openness, in particular on their willingness and ability to compete effectively in international markets. Certainly, by international standards, it is hard to see lack of openness as one of the fundamental weaknesses of these economies. But even if such an allegation could be established, it is not clear how lack of openness would contribute to vulnerability. True, it may have lowered standards of living and growth—though the counterfactual that the countries could have grown even faster had they been more open is one for which there is little evidence or support. But what is at issue is not inadequacies in their rate of growth, but their vulnerability. The link is simply not apparent.

The third concerns the issue of crony capitalism and lack of transparency. While there is widespread agreement that crony capitalism is bad for an economy and transparency is good, the question again is whether these countries were worse in these dimensions than countries that did not experience crises and whether, and in what ways, these factors contributed to vulnerability. Furman and Stiglitz (1998) show that, in terms of the standard measures of transparency, there was no decrease in transparency in these countries that would seem to account for why they were hit so hard in 1997, after being relatively immune from crises for so long. Further, they show that although some of the countries in the region seemed less transparent than average, others—equally adversely affected—appear by these same statistics to be significantly more transparent than average.

Indeed, there are questions about the extent of crony capitalism and its impact on lending practices and growth as well as on vulnerability. For instance, on balance, there is no reason to believe that the system of government-business cooperation—the much heralded Japan Inc. or Malaysia Inc., which has suddenly been recharacterized as crony capitalism—and the system of financial institutions led to massive misallocations of investment. There is a large scholarly literature that, while recognizing the dangers of abuses, concluded that the system worked well overall.²⁰ Certainly, for three decades it delivered the most impressive level of increases in GDP ever attained in such a short span of time. The fruits of that progress have been widely shared, as reflected in impressive poverty reduction, rising literacy rates, and expanding life expectancies. These increases were far more impressive than those delivered in any major country following solely the prescriptions of the so-called Washington Consensus. On balance, while there may well have been resource misallocations, evidently these were small costs that were outweighed by the real gains.

The fourth concerns the broader question of the *cause* of bad loans. There is, for instance, a consensus that the “bad” loans led to weak banks, and that high leverage, at least in the case of Korea, led to a higher likelihood of nonrepayment. But the more difficult question is to identify the source of the problem. The fact that every loan has both a borrower and a lender suggests that one should look at problems on both sides. Interestingly, in the early months of the crisis creditor countries focused their attention on problems within the debtor countries. The explanation that would be most convenient for those who seek to shift the blame for the problem to the borrowing countries is that the fault lies in crony capitalism—governments induced banks within the country to make bad loans—and lack of transparency—how could banks outside the country make good loans, given the lack of information? But neither of these explanations is convincing.

The prevalence of bad loans, even bad loans to a particular sector, is neither necessary nor sufficient evidence that government pressures led to bad lending decisions or that there were misguided resource allocation decisions. The U.S. government did not direct U.S. banks to make bad real estate loans, though inadequate government actions (high interest rates leading to a devastation of the balance sheets of U.S. savings and loan companies, followed by regulatory forbearance, including nontransparent accounting tricks) surely contributed to the problem. Nor did Western governments presumably pressure their banks to make loans to the already highly leveraged Korean firms.

Moreover, the lack of transparency was well known before the loans were made (and, as already noted, several of these countries were not less transparent than other emerging markets). Was it not the responsibility of the lenders to undertake due diligence, and if they thought there was insufficient information, not to make the loans? It is not as though these countries had suddenly become nontransparent! Indeed, it might be argued that it was the lack of transparency that had attracted them to these markets: if there were full transparency, competition would have driven profits down to zero. Investors presumably thought they were taking advantage of their differential information to obtain higher returns.

Furthermore, at a macro level, there should be concern about the returns from these investments: after all, the countries were already investing a huge percentage of GDP based on their own savings. With diminishing marginal returns lenders should clearly have been aware that there was a risk of low returns—returns insufficient to repay the loan. (This observation relates to an issue discussed below: what were the expected returns to capital market liberalization?)

But while one might have supposed that creditors should have been particularly worried about the prospects of low returns—and after the crisis, they focused on this—the evidence that overall investment was misallocated and had low returns is far from clear. To be sure, firms did face cash flow problems, which made repayment of the loans on demand impossible. But this by itself does not imply that the investment projects were bad. Indeed, even the frequently cited investments in chips in Korea is beginning to look far better, as profits have soared with the recovery of that model. Again, anecdotes make for neither good analysis nor good policy. To be sure, there were bad investment projects (both in East Asia and elsewhere). But the accusations go beyond one or two cases (though reference is made repeatedly to chips and automobiles).²¹ The suggestion is that there is an aggregate problem; but at the macro-level, there seems little if any evidence of systemic and systematic poor investment (as evidenced, for example, by large declines in total factor productivity).²²

One way of thinking about parceling blame between creditors and borrowers is to focus on information and incentives. Both the favorable grades given by the credit rating agencies just before the crisis and the huge shifts in risk premia after the crisis reinforce the view that creditors were making investments with limited and imperfect information; because they had not performed “due diligence”, beliefs were highly sensitive to news (see Furman and Stiglitz 1998).

Moral hazard problems—on the part of creditors and debtors—clearly played a role in bad lending. Both creditors and debtors may have felt that there was likely to be a bailout from the governments or international agencies (a belief that turned out to be correct). While there were losses, the social costs of the bailouts (imposed by the highly contractionary monetary and fiscal policies) of the bad loans clearly exceeded the private costs borne by the parties.²³

Beyond this, there are incentive issues on the part of both borrowers and lenders. While the problems that arise from weak banks are well known, weaknesses in both creditor and debtor countries may have contributed to the problem. For instance, weak banks in creditor countries had an incentive to make risky loans, gambling on resurrection, even to offer such loans at below actuarial rates. If that is the case, should the borrowers later be blamed for taking advantage of such good deals, once the risks are realized?

We need to push the analysis back one step: what accounts for the bad lending practices in industrial country banks as well as in developing country banks? Did inadequately designed risk adjustments for deposit insurance and capital adequacy standards distort banks' incentives? Or is the source of the problem bad governance structures within the lending banks, where pay is often related to relative performance? In this case, if everyone else lent to East Asia, a banker is not punished if

such loans go sour—because the relative-performance benchmark gives him a defense—but only for failing to lend and thereby missing a potentially profitable opportunity seized by all the other banks.²⁴

Another source of bad borrowing often cited recently is weak corporate governance. While good corporate governance is important, the role of weak corporate governance in the recent crisis is more problematic. Indeed, standard economic theory argues that governance (agency) problems are typically less important in family controlled firms. To be sure, many Korean firms focused on market share and growth over profitability. Was the cause of this focus faulty governance structures or misguided views about how to maximize market value in the long run? After all, this strategy had proved enormously successful for three decades.

Would the East Asian economies have done even better had they adopted an alternative system? There is little evidence to support this conjecture. There is, however, an intriguing issue: whatever the historical record, circumstances change. Was the current turmoil the result of a shift in the balance between the advantages and disadvantages of the “system?” Or was it that the system was well suited for one environment, but not for the new world of globalization and increasingly integrated capital markets? Or did the East Asian economies in fact change their system, in response to outside pressure for financial market liberalization, for instance, thereby undermining a system that was relatively well tuned and imposing on it stresses to which it could not adapt?

I have set forth my own views on the answers to these questions elsewhere (Stiglitz 1998b); my point here is to emphasize the importance of research and the imperative of separating solid research—deep analysis—from economic journalism.²⁵ I also want to emphasize that, typically, a multiplicity of factors interacts to produce adverse outcomes; the practice of journalists—and some politicians—of singling out one above the others is not only misleading, but is likely to lead to wrong, or at least highly incomplete, prescriptions.

The key issue that I wish to explore is the extent to which beliefs about both the causes of the crises and the appropriate responses are informed by solid economic science, theory, and empirics and the extent to which those beliefs are dictated by ideology.

Interest Rates, Exchange Rates, and the Restoration of Confidence

I will now focus more narrowly on what is generally agreed to be the central issue in East Asia—the financial crisis. The causes and consequences of financial crises have been extensively studied during the past 15 years. That research is based on our understanding of financial markets and the role they play in the economy. I cannot review that vast literature here, but I will list a few of the more salient findings, insights that are potentially of relevance to the current situation:

- Financial markets play a key role in allocating resources and monitoring the use of those resources; thus financial markets have been likened to the brain of the economy. These are quintessential informational roles, so that the

caveats about market imperfections and the limited applicability of the standard welfare theorems are particularly relevant.

- The recognition that there is a probability that loans will not be repaid is essential to understanding credit markets. If there were no concern about loans being repaid, then of course there would have been no hesitancy by foreign banks in rolling over their loans to Indonesia or Korea.
- The probability of being repaid is an endogenous variable, affected by the overall state of the economy and the interest rate charged. Raising interest rates may lower the probability of being repaid, because it attracts riskier borrowers, encourages borrowers to take actions that lower the repayment probability (for example, more risk taking), and weakens the macroeconomy.
- Lenders care not about the nominal interest rate, but about the certainty equivalent expected return. Raising the nominal interest rate may lower the certainty equivalent expected return both by increasing the probability of not being repaid and by increasing the risk premium.
- These effects can be of sufficient importance that lenders would actually choose not to increase the interest rates charged even when there is an excess demand for loans. This may give rise to credit rationing.²⁶
- The extent of credit availability may be adversely affected by bank closures, increases in uncertainty, and decreases in bank net worth as a consequence of rapid increases in interest rates and bankruptcy rates and rapid deterioration of the overall level of economic activity. Bank closures also result in the destruction of “informational capital,” further contributing to the weakening of the economy.
- General equilibrium “credit links” can be every bit as important as general equilibrium interactions through goods and services markets emphasized in traditional general equilibrium theory. In a market economy, most firms are engaged not only in production, but in credit activities, with suppliers as well as customers. Breakdowns anywhere in the system can have large systemic effects.
- Different suppliers of capital may differ markedly both in their beliefs, in their portfolios, and in their willingness to bear risks.²⁷ Accordingly, responses—say, to actions of the government—may differ markedly. There may be a flight of domestic capital, before or even as foreign capital enters.²⁸

These well-accepted principles of modern macroeconomics and finance are essential for understanding the crisis in East Asia. Their policy implications are apparent. Raising interest rates might not lead to an increased flow of capital into a country; it could have the opposite effect. Countries with an exchange rate crisis have sometimes viewed themselves as facing a tradeoff between the adverse effects of exchange rate depreciation and interest rate increases. But if increases in interest rates lead to a decreased capital flow, there is no tradeoff. Higher interest rates weaken the economy directly and can exacerbate the decline in the exchange rate.

The effect of high interest rates on the exchange rate is an empirical issue, although one on which historical and cross-country experience may shed only lim-

ited light. In normal times conventional theories have not fared very well. One of the most basic building blocks of any exchange rate theory, uncovered interest parity, has been rejected by most empirical tests. Studies have found, if anything, that when the home country's interest rate is greater than the foreign interest rate, the currency tends to appreciate.²⁹ Another standard finding in the literature is that it is impossible to reject the hypothesis that exchange rates follow a random walk.³⁰

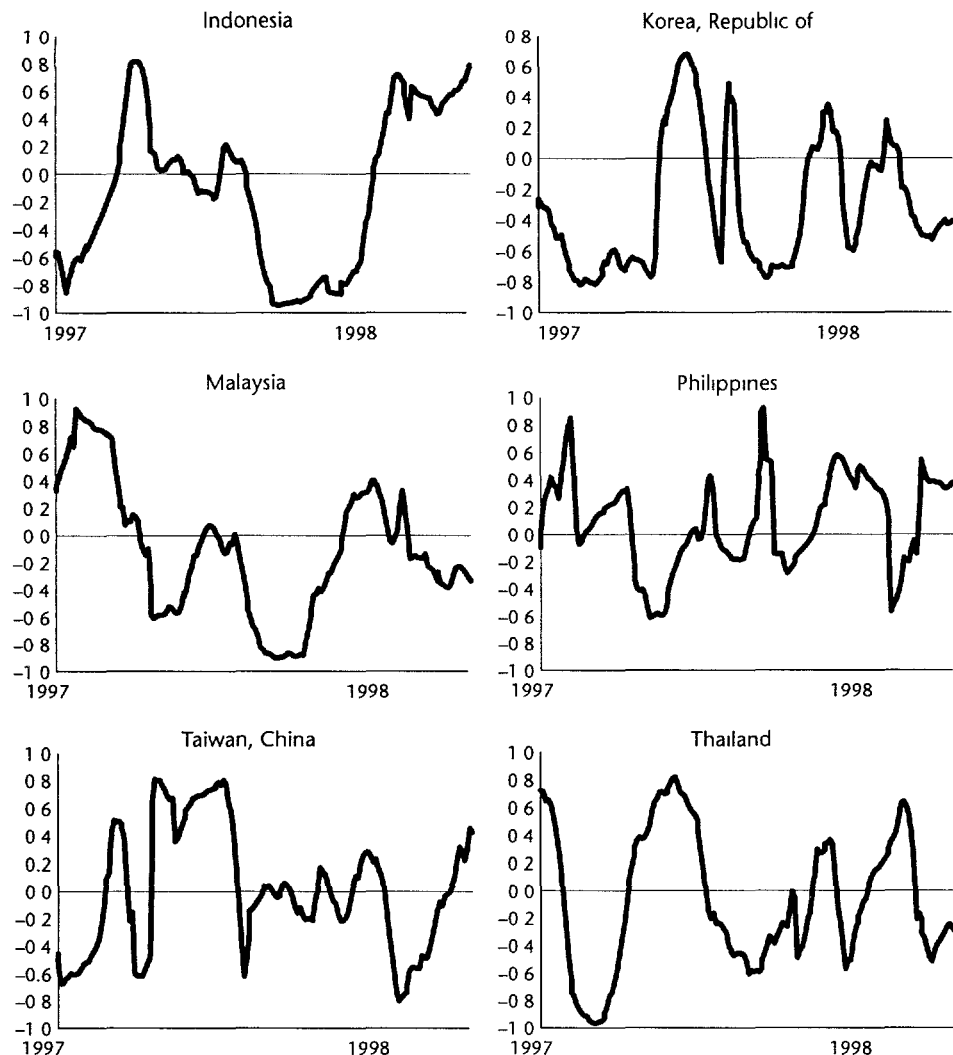
These results, however, concern only "normal" times. What about the relationship between interest rates and exchange rates during crises? One casual observation that this relationship is not more robust than that in normal times is provided by figure 1, which shows the rolling correlations between interest rates and exchange rates in different crisis episodes.³¹ The observation that there are periods in which interest rates and exchange rates are negatively correlated, as well as others in which they are positively correlated, is far from conclusive. Clearly, exchange rates and interest rates are both endogenous variables, and thus any correlation could be attributed to the effects of changes related to a third variable. What the evidence does make clear, however, is that there is no scientific basis to assertions that exchange rate recoveries have always been accompanied by or require increases in interest rates.

Reasoning and evidence may help in resolving the seeming impasse. Advocates of a high interest rate policy typically assert that there will be only minor damage to the economy, because effectively defending the exchange rate requires only temporarily high interest rates. The question then is, why should a temporary increase in the interest rate lead to a permanent shift in the supply curve for capital to a country? After all, in order to defend a currency against the expectation of even a 1 percent fall in the currency the following day, an overnight interest rate of 1 percent per day (which is 3,678 percent per year) is required, assuming risk neutrality. For risk-averse investors the interest rate would have to be considerably higher. (To put this in perspective, between July 1, 1997, and April 1, 1998, the Indonesian rupiah declined at a daily rate of 0.6 percent, some days declining as much as 18 percent. An investor expecting this depreciation would have kept his or her money in Indonesia only in exchange for an almost 500 percent return, at an annual rate, sustained over nine months.)

SIGNALING. The direct effect of higher interest rates on returns would thus seem often not enough to defend a currency; and when these higher interest rates are only believed to be temporary, why should they have any but a temporary effect? And shouldn't the knowledge that the effects will be temporary even dampen the immediate impact? Instead, the best case for a high interest rate policy is that it leads to a change in a state variable, namely beliefs about, say, the resolve of the monetary authorities to pursue low inflation.³² And that beliefs are changed in such a way that when the interventions leading to high interest rates are withdrawn, not only will interest rates fall, but exchange rates will be stabilized at a level higher than they would have been without the intervention.

There is by now a well-developed theory that explains why government interventions might change beliefs. For instance, if the government (or international

Figure 1. Rolling Correlations between Exchange Rates and Interest Rates in East Asia, July 1, 1997 to March 4, 1998



Note. Exchange rate is defined as U.S. dollars per unit of local currency. A positive correlation indicates that higher interest rates are associated with a stronger currency.
Source. Calculations by Sergio Schmukler, World Bank, based on Bloomberg data.

agencies) has information that is not publicly available, agents in the economy may make inferences about the underlying state of the economy from the nature of the interventions.³³ Similarly, observers might well infer that a patient who has been subjected to a strong dose of chemotherapy has been diagnosed not only to have cancer, but also to suffer from a form of cancer that cannot be treated in a more benign way. So too, strong and painful economic actions, especially preemptive actions before signs of a crisis are visible, may be interpreted as signs of serious economic maladies. The actions themselves, of course, *given a particular appraisal of*

the economy, may lead to greater optimism; but whether the combined effect—the change in the appraisal plus the change in actions—is positive or negative is in general ambiguous.³⁴ This is especially the case where the efficacy of the remedies will be established only in the long run, while the impact of the reappraisals will be felt immediately. (Much of this literature is based on rational expectations. But when irrationalities of the kind discussed below are taken into account, the possible adverse effects are magnified.)

There is little if any research that explicitly models or tests the *positive* effects of temporary high interest rates as a signal of the resolve of a central banker to maintain or strengthen the value of the currency.³⁵ But this hypothesis can be subjected to theoretical and empirical tests. At a theoretical level the key question is one of internal consistency: a basic tenet of the theory of signaling is that to be effective, signals must be costly. If it were costless to signal that one were a responsible monetary authority by imposing high interest rates, everyone would do so, rendering high interest rates ineffective as a signaling mechanism. One cannot hold simultaneously to the beliefs that there will not be real, adverse consequences (real costs, at least to some groups within the population) and that high interest rates are an effective signaling mechanism. Thus for high interest rates to be an effective signaling mechanism, there must be significant costs. But costs to whom? Presumably to those making the decision to raise interest rates. But the costs of high interest rates, and the recession they may induce, are borne heavily by workers, small businesses, and debtors more generally. Creditors—so long as they are repaid—are actually better off.

Note the curious irony: as central banks have become more independent and less politically accountable, they, the decisionmakers in raising interest rates, bear less and less of the costs. Indeed, in the many countries where central banks are controlled by financial interests, central bankers may actually benefit from the high rates, at least to the point at which the bankruptcy effect overwhelms the direct interest rate effect. Interest rate increases then become an effective signal only after the point at which there is a large contractionary effect, one so large that even lenders are worse off with the higher interest rates. More generally, the more independent, less politically accountable, and less representative the monetary authorities, the higher are the costs that high interest rates must impose on the economy before they become an effective signaling device.³⁶

There is another aspect of internal consistency of costly signaling mechanisms involving political processes: If the cost is too high, the signal is not credible, because no one will believe that the policy will be sustained (Drazen and Masson 1994).³⁷ Even if the current government establishes its credibility, there is overwhelming evidence that economic downturns lead to an increased likelihood of a change in government (see Lewis-Beck 1988 and Paldam 1991). Although in principle an independent monetary authority might insulate monetary policy from such political pressures, at least for a time, even monetary authorities with a long history of independence recognize their vulnerability. If they push too hard, their independence can be taken away. Governors of central banks certainly appreciate this not-so-subtle point. Paul Volcker put this well when, as chairman of the U.S. Federal

Reserve, he told a Congressional committee that “the Congress created us and the Congress can uncreate us” (Orszag 1991, pp. 1), citing Greider (1987).

Of course, even those who believe that temporarily high interest rates have an effect on beliefs, and therefore on the supply curve of capital, should ask whether there are less costly signaling mechanisms. Or mechanisms that are less costly to innocent bystanders, but perhaps more costly to those engaged in risky behavior, and therefore provide better incentives to prevent future crises while restoring confidence in the midst of a current crisis.³⁸ This is an especially relevant question for international institutions, which may be in a position to help the market coordinate a signaling system.

The hypothesis that high interest rates are an effective (if costly) signaling mechanism can be subjected to empirical testing. The information contained in the signal would presumably be related to prior information and beliefs. Thus a high interest rate policy would be expected to be less effective in conveying information about a monetary authority that had a long reputation for responsible monetary policy (as evidenced by low inflation) because the revision in beliefs would presumably be smaller. This strongly suggests that even if one believed in the signaling theory, a high interest rate policy would be less effective in East Asia than it was in Latin America. The subsequent experience of currencies continuing to depreciate after the initial interest rate hikes seems consistent with this hypothesis. But the hypothesis needs further testing, for example, by comparing the magnitude of responses across countries. It is even possible that higher interest rates could convey negative information, as in cases where investors are looking for evidence not so much for commitment in controlling inflation but rather for competence in managing macroeconomic policy under current circumstances. In that case, raising interest rates to high levels in a very highly leveraged economy with short-term indebtedness, where such a policy can throw the economy into a deep recession, could signal a high level of incompetence. The supply curve of capital could be shifted by the action, but in the wrong direction.

Moreover, what typically conveys information is not just the level of the interest rate, but how long it is maintained. Indeed, the standard “explanation” for the failure of high interest rates to do their work is that countries have not stayed the course. This too is an empirically testable proposition.

MULTIPLE EQUILIBRIA. The other major set of models providing a consistent rationale for higher interest rates as a means of stabilizing the exchange rate are those generating multiple equilibria. Although the results are not in, I am doubtful that these justifications for high interest rates will stand up under close scrutiny. In these models government intervention can sometimes help coordinate the economy to “choose” the good equilibrium. Having intervened to convince market participants about which equilibrium would prevail, the government can step aside. In principle, anything can serve to coordinate such “sunspot” equilibria (Azariadis 1981; Grandmont 1985), and governments should, presumably, choose coordinating signals that do not themselves have adverse effects or change the equilibria.

Those who argue for the use of high interest rates thus have a heavy burden. They need to construct the multiple equilibria models. They need to devise convincing tests that high interest rates are indeed serving as the coordinating mechanism. And they need to show that it would be difficult if not impossible to establish less costly coordinating mechanisms. Although the literature is not replete with convincing models that satisfy these criteria (to say the least), it is easy to construct multiple equilibrium models in which government intervention in the form of sufficiently higher interest rates eliminates the “good” equilibrium, with the remaining stable equilibrium being discretely lower.³⁹

EVALUATING ANY POTENTIAL TRADEOFF. The same mix of theory and evidence can help shed light on another major question: Assume that there is a tradeoff, that is, that higher interest rates do in fact lead to higher exchange rates. The finance-based macroeconomic models make clear that unanticipated increases in interest rates have adverse effects on the economy; but so too can changes in the exchange rate, especially for debtor countries.⁴⁰ The relative importance of the two effects is, of course, an empirical matter. In this case, we do have sound research findings, based on a cross-section of countries: high interest rates have a significant adverse effect—they substantially increase the probability of a financial crisis—while depreciations of currencies have little if any effect (Demirgüç-Kunt and Detragiache 1998a).

Such cross-country regressions always need to be taken with a grain of salt. They often entail regressing two endogenous variables against each other; and they often suffer from spurious correlations.⁴¹ But when such regression results are consistent with theoretical models, they may at least bolster our confidence in the relevance of those models. The empirical finding that banking systems are more sensitive to interest rate increases than to exchange rate depreciations is consistent with theoretical models. Risk-averse firms have an incentive to hedge against foreign exchange risk and have access to instruments that allow them to do so, at least for short-term exposures. By the same token, risk-averse lenders have an incentive to restrict lending to firms with excessive unhedged currency positions. (The argument that the quasi-fixed exchange rate lulled borrowing firms, and banks that lent to them, into a false sense of confidence and that therefore it is the exchange rate regime, not the borrowers or lenders, that is to blame for any excessive exposure, is hardly convincing. First, it seems to suggest a widespread belief in persistent disequilibria. If exchange rates were really fixed, the cost of borrowing in the two currencies would be markedly different, and there would be arbitrage profits. Second, and more important, investors should realize that, without an explicit guarantee, quasi-fixed interest rates are not a guarantee of no exchange rate changes; in every historical case there eventually has been a change. Insurance markets are designed to deal with rare events that have high costs, like fires. If there is a widespread belief that there will not be a devaluation—that the fire will not occur—it will be reflected in a low insurance premium. But rational, risk-averse individuals will still buy insurance. Borrowers could have purchased cover. The only reason that they might not have is an expectation of a bailout—the kinds of expectations that are confirmed by

policies that are justified by worrying about the adverse effects of exchange rates on those who have taken an uncovered position and their creditors.⁴²)

By contrast, maturity mismatches are endemic in all economies. The ability to withdraw credit may have an important role in providing discipline to borrowers, the benefit of which more than offsets the adverse risk effect from the maturity mismatch of assets and liabilities (Rey and Stiglitz 1993).⁴³ Thus the observation that *on average* risks associated with exchange rate changes present less of a threat to financial stability than do interest rate changes is consistent with plausible economic models based on at least a modicum of market rationality.

But even when cross-section results are supported by theoretical models, the results must be applied with care to any particular country. For instance, firms or financial institutions within a particular country may have a very large foreign exchange exposure, suggesting that *if there is a tradeoff between interest rates and exchange rates*, more weight be placed on adverse effects on the exchange rate than would be the case “on average.”⁴⁴ On the other hand, in countries like Colombia or Malaysia, where wise central bankers imposed regulations restricting the foreign exchange exposure of financial institutions (and, in the case of Malaysia, must approve all corporate foreign borrowing on a case-by-case basis), far greater weight should be placed on the disruptive effects of high interest rates.

Contagion and the Response to the Crisis

One of the main justifications for *international* intervention in a crisis has been worry about contagion—a downturn in one country having adverse effects on others. This is a legitimate worry. In any general equilibrium system all the parts are interconnected. But the worry about contagion is more than just the worry that the derivative of one country’s GDP (the “infected” country) with respect to GDP or exchange rates in the crisis country is positive, so that a fall in one country leads to a decline in another. Rather, the worry about contagion is that a large fall in one country will lead to a large decline in another.

Most of the focus has been on contagion of financial variables rather than real variables, though to be sure the two are linked. If one believed that contagion effects were transmitted mainly through real variables, the focus of concern would be on maintaining output in the crisis countries, not on maintaining currency values or even stock market values. Of course, those variables could still be of concern as intermediate variables to the extent that they affected the real economy.

Empirical evidence suggests that although there are systematic co-movements in financial variables, in general the effects are weak. But one cannot refer to general patterns in crisis situations: there clearly can be nonlinearities. And even experience may be a limited guide, given the increased degree of financial market integration. The uncertainty about the significance of contagion effects is thus unlikely to be resolved, at least in the near future.

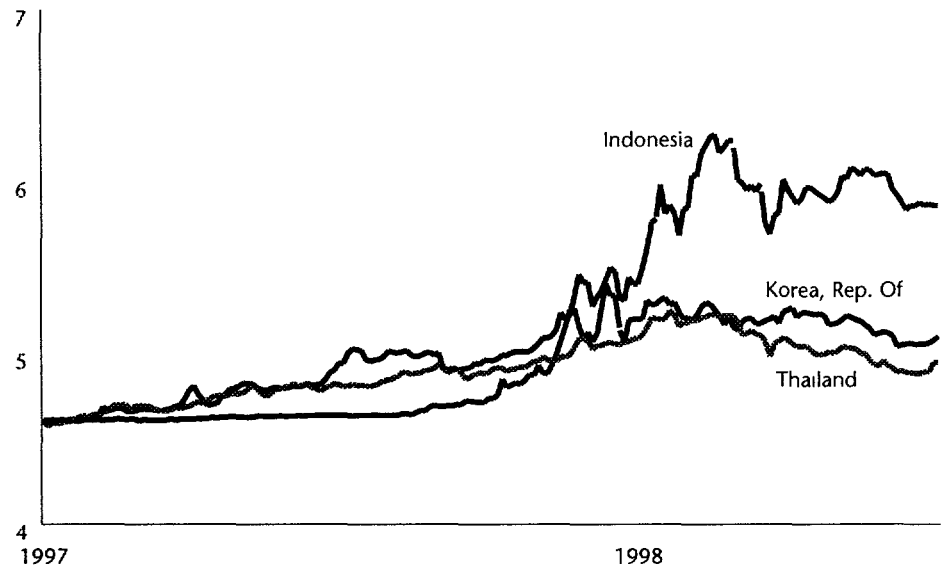
The current situation provides some insights, precisely because the initial remedies failed to resolve the problems in the countries where the crisis originated. Early

interventions in exchange rates, for example, are hardly discernible in the overall patterns of decline (figure 2). The international effects were relatively minimal until the end of October 1997, when the overall risk premia for loans to emerging markets increased significantly (figure 3). Evidently, investors' willingness or ability to bear these risks changed. The fact that emerging markets in general had especially large increases in risk premia suggests that investors perceived greater risks associated with those assets. This is true even though it is hard to see how much new information about Brazil or Russia was revealed by the crisis in East Asia. Instead, what may be at work is Keynes' beauty contest phenomenon: it is not that the investors themselves perceived greater risks, but rather that they believed that others would and that this by itself would suffice to bring down the market. It is worth noting, however, that the correlations seem strongest in the immediate aftermath of the shock. As time has passed, spreads have come back down in the countries outside of East Asia.

Keynes' argument about the irrationality of markets—that changes in expectations are, in many ways, inexplicable, so much so that he referred to them as animal spirits—has been one of the most contested hypotheses in economics. Note that Keynes did not argue that markets would not be affected by fundamentals, but only that a significant part of the movements was inexplicable. Early studies found that it was impossible to reject the hypothesis of the efficiency of markets (see for example Fama 1970). Later work has cast doubt on these results, pointing out that the power of the tests was very low and even finding evidence of systematic deviations

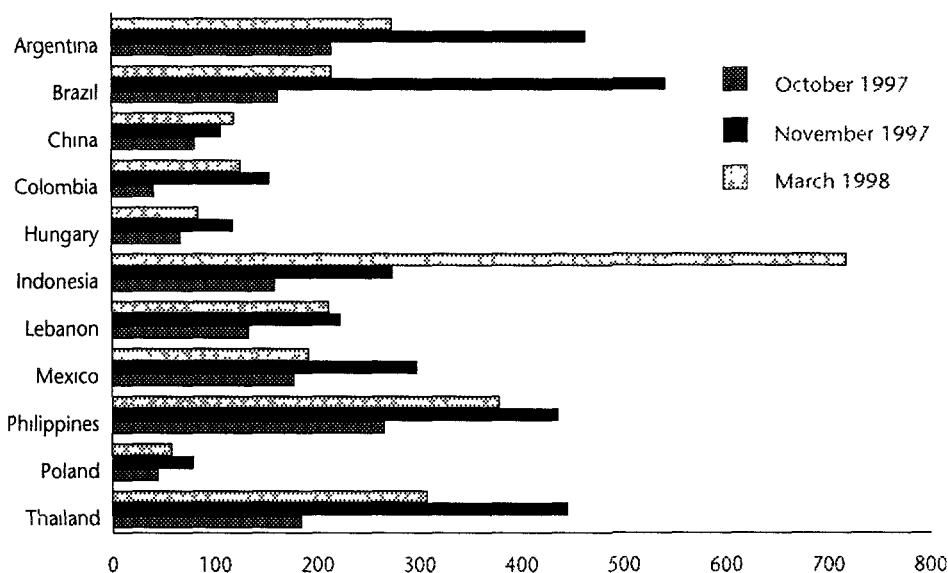
Figure 2. Exchange Rates in East Asia, July 2, 1997 to April 3, 1998

Log exchange rate (July 2, 1997 = 100)



Source: Bloomberg.

Figure 3. Secondary Market Spreads on Sovereign Bond Issues, Various Countries



Source: Bloomberg

from the fundamentals, including mean reversion (Poterba and Summers 1988) and movements in large stocks leading movements in small stocks (Lo and MacKinlay 1990). Further evidence for the deviation of markets from fundamentals comes from econometric studies showing that the variance of asset prices is substantially greater than can be justified by the stochastic properties of earnings streams and discount rates (see Shiller 1989).⁴⁵ Finally, the observation of certain striking events provides convincing support for large deviations from fundamentals, that is, some degree of irrationality. Is there any “news” that could account for the 23 percent decline in the market value of U.S. corporations on October 19, 1987, or the 180 basis point increase in the spread on emerging market debt on October 27, 1997?

If we accept the irrationality hypothesis, then of course “anything is possible.” It is possible that restoring strength to Thailand will help Brazil. But if we cannot explain losses of confidence, how can we be sure about what will restore confidence? And although the irrationality hypothesis argues that there are fluctuations in market values that cannot be explained by “news” of fundamentals, it does not reject all rational analysis. If the reason that Brazil and Russia are suffering is that the East Asian crisis has called attention to the importance of financial fragility, and if addressing financial fragility is viewed as an essential part of the restoration of confidence in Thailand, why should not parallel programs be required by all the countries affected by contagion? Indeed, some have argued that, in principle, a bailout motivated by contagion could actually exacerbate the adverse effects in some other countries. If a large financial package is viewed as essential for addressing the

problems of, say, Mexico, and if other countries face similar problems (the basis of rational contagion) but are not likely to get comparable bailouts, then a Bayesian should view the situation in those other countries with a more jaundiced eye.

There are other dilemmas in thinking about the restoration of confidence. Is it plausible that confidence in an economy can be restored, even as the economy plummets into recession, especially if that economic downturn leads to political and social unrest?

Or consider a more technical puzzle. Standard theories argue that (other things being equal) a reduction in the fiscal deficit depreciates the equilibrium exchange rate.⁴⁶ If this is true, then is it likely to strengthen today's exchange rate?⁴⁷ In a world in which beliefs may be irrational, anything is possible, but assertions need to be backed up by relevant evidence—and there's the rub. Since each crisis differs from previous ones and we have vast problems of underidentification, the relevance of a particular experience becomes a matter of judgment. For instance, some have pointed to the success of the Mexican bailout as proof of all manner of propositions, including the wisdom of staying the course. But Mexico's problems were in many respects more akin to those of Latin America than to those of East Asia. Although Mexico's growth had been strong in the years leading to the crisis, it did not have a decades-long track record of low inflation, fiscal prudence, and strong growth. Moreover, Mexico had a large number of exporters—who had just gained greater access to the lucrative U.S. market—with close ties to U.S. firms (many owned by U.S. companies), and thus the collapse of the domestic credit system would not have the large adverse effects that it would have had in a less fortunate country.⁴⁸

Applied Economic Knowledge II: Financial and Capital Market Liberalization

I do not want to give the impression that ideology is the exclusive province of macroeconomists. I could have drawn equally telling examples from microeconomics. Let me illustrate by a brief discussion of two issues, the liberalization of financial markets and privatization (discussed in the following section).

The ideological basis for the liberalization of financial markets, a process that includes freeing controls on interest rates, banking, and capital flows, is simple and easily summarized. Free and competitive markets are the basis of a capitalist economy, and they have delivered enormous benefits to those that have adopted them. There should be no more question about the virtues of liberalization of financial markets than about the liberalization of trade or any other market within the economy.”

Unfortunately, the *scientific foundations* for this *ideological* position are not very sound. There are important differences between financial markets and other markets, differences that suggest that although there is a presumption that trade liberalization is welfare enhancing, liberalization of financial markets may well not be. Empirical evidence, as well as recent experiences in East Asia and Africa, buttress the theoretical propositions that economies can suffer from too little regulation, just as they can suffer from too much or the wrong kind of regulation.

The underlying rationale for this claim is provided by the general theorem to which I referred earlier, that competitive economies with imperfect information (and incomplete markets) are not constrained Pareto optimal (Greenwald and Stiglitz 1986). A central function of financial markets is providing information—selecting and monitoring projects. Market failures, including those associated with moral hazard and adverse selection, are thus endemic. Elsewhere, I have provided a taxonomy of these market failures and of the appropriate government actions to remedy them (Stiglitz 1994a).

Moreover, there are also general results showing that markets characterized by imperfect information are often imperfectly competitive and that imperfectly competitive markets are typically not Pareto efficient.⁴⁹ Empirical studies show that even though there may be many banks in the economy (and thus effective competition for deposits), there may be submarkets (for example, loans to small businesses in a particular region of a country) in which competition is very limited. These problems are likely to be particularly severe in developing countries, where information markets are less well developed, and recent experiences show that financial market liberalization under such circumstances may lead to larger not smaller spreads between lending and deposit rates.⁵⁰ Economic theory has established that financial markets are different from ordinary markets and more likely to be characterized by market failures requiring government interventions. It has also shown that without such intervention, there may be a tendency toward excessive risk taking, leading to high levels of financial fragility (see, for instance, Hellman, Murdock, and Stiglitz, 1998 and Stiglitz 1996).⁵¹

Experience shows that successful financial markets require strong government regulation, effective laws, and vigilant enforcement. Successful financial systems recognize four goals for financial regulation: maintaining the safety and soundness of the banking system, promoting competition, protecting consumers, and ensuring that all groups have access to capital. The strength of the U.S. economy today is partly due to its strong financial system, a system whose strength is based on many pillars that date back a century or more. Financial regulations have evolved over the past 135 years. The strong protection of minority shareholders rests on common law principles that predate the establishment of the United States. Strong competition laws date back to the turn of the century. And effective securities laws have been spurred by the recognition of the massive abuses in the first part of this century that were undermining confidence in the markets.

By now, we also have ample experience with the consequences of trying to do without a legal and regulatory structure. The experience of some of the former socialist economies shows not only that abuses will arise without such laws (in some cases heightened by the increased sophistication that has accompanied modern capital markets) but also that in these circumstances capital markets fail to perform their fundamental roles (Coffee 1996; Ellerman forthcoming; Nikitin and Weiss 1997).

The history of free banking has been so dismal that today there are few supporters of that ideological position. There is a recognition that good regulation is required for a strong financial system, and a strong financial system is required for

a strong economy. Countries with deeper, more liquid, and better regulated financial markets grow faster (Levine 1997) and are less prone to crises, which in turn have strong adverse effects on the economy. Caprio (1997), for instance, shows that five years after a financial crisis in a developing country, output is 7 percent lower than it would have been had the precrisis growth rates been maintained—and the shortfall is even larger for OECD countries. (He also finds no evidence that precrisis growth rates were higher than normal.) The connection between weak regulation and banking crises had been made clear by the numerous episodes of the past 25 years, including the U.S. savings and loan debacle (which helped spawn much of the recent literature on the topic); and this connection has now been confirmed by cross-country studies (Demirgüç-Kunt and Detragiache 1998a and elsewhere in this volume; Kaminsky and Reinhart 1996).

But while the ideological position of free banking has few adherents today, there is a first-cousin that has gained increasing strength in recent years: regulators should only ensure that banks have adequate capital. So long as banks have adequate capital they will have appropriate incentives to engage in safe lending. Regulators, in this view, should not restrain lending activities (other than, presumably, to restrict insider-lending). There is a clear ideological basis for this position: it entails minimal regulatory intervention. But the general theory of regulation (based on imperfect information) provides no support for this position. (see Stiglitz 1999) For instance, it is not just capital (as conventionally measured) that provides incentives, but also expectations of future profit—what is called franchise value.

We now know a great deal about what affects franchise value and about the interactions between capital requirements and franchise value. Financial restraint—restrictions that, for instance, keep deposit interest rates below market-determined levels—may enhance franchise value and thus actually reduce these problems, leading to more prudent behavior by banks and thus a more efficient financial system (Caprio and Summers 1996). It has also been shown that increases in capital requirements are an inefficient substitute for the franchise value that is lost as a result of full liberalization. Pareto efficiency requires the use of both instruments, even in banking systems without deposit insurance (Hellman, Murdock, and Stiglitz 1997).

These results are confirmed empirically by cross-section regressions which show that financial restraint is associated with higher (or at least not lower) levels of investment and growth.⁵² Had the United States decided, in the midst of the savings and loan crisis, to eliminate the tax distortions associated with preferential treatment of real estate (such as the deductions for home mortgages) and the distortionary agricultural programs that increased land prices, the mild economic downturn that followed that crisis would likely have become a full-blown recession.

Another example may have particular relevance to some recent crises. Many reforms entail reducing rents. But reducing rents affects capital asset values and the financial viability of firms and the financial institutions that have provided them capital. When financial institutions are weak and firms are on the verge of bankruptcy, the costs to the economy, in terms of economic disruption, of undertaking reforms with potentially large and uncertain effects on capital asset values may far outweigh

the short-run efficiency gains. This is especially likely in economies with poor information systems, where the impact on particular firms may be hard to ascertain and so there may be adverse effects even on firms that receive little or no returns from such rents. A slight delay in undertaking such reforms may accordingly be desirable.

Let me be clear: many countries have regulations that serve no useful purpose. They increase transactions costs, decrease efficiency, and do not enhance the stability of the financial system. Governments should eliminate these regulations. There is evidence that financial repression—entailing highly negative real interest rates—is bad for economic performance. But there are many regulations, including restrictions on interest rates or lending to certain sectors (such as speculative real estate), that may enhance the stability of the financial system and thereby increase the efficiency of the economy. Although there *may* be a tradeoff between short-run efficiency and financial system stability, the costs of instability are so great that the long-run gains to the economy may more than offset any short-term losses. (The observation of underregulated financial markets that result in borrowing for huge investments in unoccupied commercial real estate—even when those borrowers were willing to promise to pay higher interest rates than those wishing to invest in productive plant and equipment—raises questions about the existence of the tradeoff.)

The appropriate set of regulations will, of course, depend on the structure of the economy and the capacity of regulators. Developing countries typically face greater risks (because their economies are less diversified and many of them are more dependent on commodities with more variable prices), have less developed information systems, and lower regulatory capacities than more advanced economies. And even more advanced economies, while recognizing the desirability of risk-adjusted capital adequacy standards, have done a poor job of implementing them. (Even seemingly well-respected regulators in advanced economies continue to focus on credit risk, taking inadequate account of risks associated with capital values arising from changes in interest rates and exchange rates. And they continued to do so even after being warned of the distortions in lending policy to which such misguided regulations give rise.⁵³) In such circumstances, not only may higher capital adequacy standards be desirable, but so too may restrictions on interest rates and on certain forms of lending.⁵⁴

I have shown here that elementary considerations—theory and evidence—argue that not only is free banking inappropriate, but so too is financial market liberalization that puts sole, or excessive, reliance on capital adequacy standards.⁵⁵ It is possible that those who have argued for these forms of financial market liberalization and have advised governments to engage in them have done so partly on the basis of an inadequate understanding of financial markets and the principles of regulation. But it is also possible that they were driven to this position by their ideological presuppositions.

The same ambiguity is evidenced in two other related issues. Consider first the issue of the reforms that need to be undertaken in addressing the crisis. Earlier, we noted the confusion between weaknesses in the economy that may have had adverse effects on the overall level of economic performance, and weaknesses in the economy that make it more vulnerable, more crisis prone. Many distortions induce inef-

iciencies, but may not increase vulnerability. Should these be reformed in the midst of a crisis? There are, to be sure, political economy arguments for doing so—the so-called “window of opportunity.” But counterbalancing those arguments is the possibility that such reforms may actually exacerbate the crisis.

The reason is simple. Opening up the capital account may subject the economy to more systemic risks. Whether the gains are worth the risks will presumably depend on the circumstances of the country. Ideology would simply focus on the gains. Economic science would attempt to quantify both the gains and the risks, a balance that may differ markedly for countries in different circumstances. For instance, for a country with a very high savings rate, the marginal return from the additional capital that it might access might fail to compensate for the increased risk.⁵⁶ Interestingly, there is little if any statistical evidence showing that capital market liberalization is associated with faster economic growth or increased investment (Rodrik forthcoming). These studies confront great difficulties, especially in the measurement of capital account liberalization, yet I believe that the following assertion is uncontroversial: Although there is a consensus among economists—based on a wealth of studies—that trade liberalization brings significant economic gains, there is no such consensus about capital account liberalization. There is, however, a consensus that at least one of the consequences of capital account liberalization is to increase the risk facing the economy. One advocate of the *net* benefits of financial liberalization, U.S. Deputy Treasury Secretary Lawrence Summers (1998), put this well when he compared today’s global capital markets to the jet airplane, noting that “the crashes, when they occur, are that much more spectacular.”

Applied Economic Knowledge III: Privatization

No concept is more sacred to the ideology of capitalism than private property, and it is thus not surprising that privatization has become a centerpiece of the modern ideology of reform. The ideology has been buttressed by extensive experiences with inefficient government enterprises. What distinguishes science from ideology is that it seeks to understand the reasons for the differences and therefore the conditions under which privatization is likely to have the desired results, without undue side effects. Again, we can list some of the essential theoretical and empirical propositions:

- The conditions under which a government enterprise can be privatized, capture full rents, and still pursue the same range of social objectives that the public enterprise pursued are highly restrictive, corresponding closely to the conditions under which the fundamental theorems of welfare economics have been established. This result is sometimes called the Fundamental Theorem on Privatization (Sappington and Stiglitz 1987).
- Some public enterprises have operated at a level of efficiency comparable to or greater than that of similarly situated private enterprises. Typically these public enterprises are associated with firms subjected to competition, either in exports (as in Korea’s steel industry) or domestically (as in Canada’s railroads, as described in Caves and Christensen 1980).

- Incentive problems arise in all large organizations and are similar in public and private enterprises engaged in similar kinds of activities (Stiglitz 1989a). Rent seeking activities occur in both private and public organizations (see Shleifer and Vishny 1989; Edlin and Stiglitz 1995). Indeed, much of management activity may be directed to increasing the scope for capturing rents.
- Theorems establishing the efficiency of markets require both private property and competitive markets. Converting a public monopoly into a private monopoly may actually lead to higher, not lower, prices and less, not more, overall economic efficiency.

While hardly a controlled experiment, the contrast between the outcomes of China's emphasis on competition and the emphasis on privatization in the countries of Eastern Europe and the former Soviet Union discussed earlier sheds some light on these issues.

That said, the general principle that institutions should focus on whatever is their comparative advantage suggests that governments should focus on activities unlikely to attract the private sector or in which externalities and public goods effects are likely to be especially important. It makes little sense for governments to produce steel when public responsibilities such as the development and effective implementation of legal structures are inadequately fulfilled. But the hard issues lie in the in-between areas, and here again there is no consensus. Particularly problematic are situations in which there are monopolies—natural or not—and in which governments have not yet developed effective competition and regulatory policies. Recent experiences in Africa in telecommunications show the potential for such measures as introducing competition in procurement (say of lines) as a way to lower costs to consumers and increase availability of services. Elsewhere, we have seen the benefits from competition among new providers, especially of cellular services.

As noted, the sequencing and pacing of reforms matter. Privatizing a monopoly results in a vested interest with the resources to influence the political process, which may successfully prevent the achievement of the second pillar of a successful market economy—competition. Even if a convincing case were made for privatizing social security, doing so prior to the establishment of stable and efficient capital markets, as has been proposed in Russia, raises risks that are only now being fully realized. (Much of the discussion in this area has been tainted by a confusion between the gains from having a fully funded pension system and the gains from privatization itself.⁵⁷)

The Policy Adviser as Interlocutor

Earlier I argued that economic analysis needs to rise above the standards of journalism. Assertions need to be subjected to rigorous analysis—a combination of theory, focusing on internal consistency and the plausibility of the assumptions required to derive the conjectured relationships, and empirics, with that too combining statistical analysis with detailed evidence from particular episodes. Given the ever-changing nature of our economy and institutions, it will never be possible to

completely eliminate uncertainty. Uncertainty will be particularly large in interpreting and responding to those unusual events called crises. And this uncertainty should, at the very least, induce a modicum of humility on the part of advisers.

An essential part of policy advice based on economic science—as opposed to economic ideology—must be an explicit recognition of that uncertainty. Too often, that has not been the case. The presence of uncertainty means, of course, that it will be all the more difficult to assess the quality of the advice being rendered. Successful economic recoveries may occur in spite of the advice, and failures might have been even worse were it not for the advice. It is often difficult to find a counterfactual to use as a benchmark. I shall return to this theme later. Here, I want to stress a somewhat different point.

John Neville Keynes (1919), in his classic book *The Scope and Method of Political Economy*, argued for a well-defined role of economists in describing the consequences of alternative actions—defining the opportunity set, to use today’s language. This role, according to Neville Keynes, should be separated from that of the political process, which entails choosing among the points on the frontier of the opportunity set. Seldom does the frontier consist of a single point, so that a single policy Pareto dominates all others. Accordingly, seldom should the adviser give a single recommendation.

This is particularly the case when there is uncertainty about the consequences of various policies. Advisers, of course, have a responsibility to make sure that the alternative outcomes—and the probabilities associated with them—are understood as precisely as possible. But the way in which risks are weighted and balanced is a political decision that must be made by the country. No outsider should impose risk preferences on those who must live with the consequences.

But here again the analytic and scientific role of the adviser becomes blurred, in a way that Neville Keynes, coming before the great advances in statistics to which his son John Maynard Keynes contributed so much, could not appreciate. Today, whether we know the term or not, we are all Bayesians—or almost all of us. We know that statistical inferences are based on loss functions that evaluate the consequences of different errors—and the appropriate loss function should not be that of the adviser, but that of the party being advised. But at least since the work of Arrow (1963), we also know that there is no general way to aggregate individual preferences in order to derive a social “loss” function. Thus, various groups within a society will have different loss functions, and different Bayesian estimates. It should come as no surprise, for instance, that those in the financial community tend to be more worried than the general public about inflation, for price increases can lead to a decrease in the real value of their financial assets. Those who depend upon financial assets see the tradeoffs between inflation and other variables, such as growth and employment, quite differently from, say, workers, who have more to lose from increased unemployment and more to gain from the erosion in value of their debts. It is thus incumbent on the scientific adviser not only to explain the uncertainties, but also to explain the differences in views, the evidence in favor of the alternative views, and, if possible, how different loss functions weight the evidence to arrive at different Bayesian probabilities.

To be sure, our advice may not be couched in the vocabulary I have just used. But I would argue that our advice becomes more credible when these issues are explained clearly, and the choices and risks well articulated. Whether we like it or not, the alternative views will be on the table anyway—that is the great strength of democracy and modern science—and the country will want to know what is the evidence and reasoning behind the alternative views.

This is becoming all the more important as we engage in capacity building and strengthening democratic institutions within the countries in which we work—a task to which I believe most of us are committed. At the beginning of this paper, I referred to the renewed emphasis on sustainable development—sustainable not only in terms of its impact on the environment and the utilization of natural resources, but also in terms of its durability, its ability to withstand the vicissitudes of political processes. The more widely accepted the premises and appropriateness of any reform, the more sustainable will that reform be. We need to be especially sensitive to reforms and policies for which there may not only be a lack of consensus in the more developed countries, but which are also contrary to long-standing political doctrines in the countries themselves. Changes in political cultures cannot, and should not, be imposed from the top, though leadership—if it is truly convinced of their benefits—can clearly facilitate such changes.

The mark of our success in capacity building over the past couple of decades is that there is a cadre of well-trained professionals who are willing and able to engage in meaningful dialogue on issues of economic policy within most of the developing countries. They bring a local knowledge that even someone who has lived a year or two in a country cannot match—let alone someone flying in for a three-week mission. To be sure, they may lack the cross-country experience that an outside adviser can bring, and it is from the marriage of these two knowledge bases that the most fruitful lessons can be drawn.

Advisers' Incentives

A key part of the advice that any economist gives today concerns incentives. We look at the implicit as well as the explicit incentives facing households, firms, and even government bureaucrats. A central part of the rationale for privatization and liberalization is an improvement in incentives, including those associated with more competition. Often, the most subtle part of an economist's task in analyzing the economic problems facing a country is to understand the complex set of incentives facing various participants as a result of the interaction of institutions, customs, laws, and regulations.

Advisers too are subject to incentive issues, and those who pay attention to their advice, whether the direct recipients or interested third parties, need to be aware of those incentives. Indeed, understanding and reshaping incentives can help us improve the “market for advice.”

The nature of these incentive problems is well illustrated by recent discussions in healthcare. Doctors provide a multitude of services: they diagnose ailments and pro-

vide advice about appropriate prevention and treatment as well as directly providing services intended to address health problems. The market for medical services, however, is distorted by imperfections in information that allow room for the doctors' personal and institutional incentives to shape the kinds of care that they provide.

The most commonly noted incentive effects in today's healthcare system, for example, have to do with the institutional structure through which medical services are provided. The fee-for-service system prevalent in the United States has been widely criticized because doctors have an incentive to overprovide. There is also a widespread view that even though the doctors subscribe to "professional norms," those norms themselves adjust to the incentives facing healthcare providers. Managed care, on the other hand, has been widely criticized because health management organizations (HMOs) have an incentive to underprovide in order to cut costs.

These incentive effects exert a strong influence because of imperfections of information in the field of medical care. Clients rarely know the costs and benefits of the range of treatments available to them and usually cannot serve as a check upon the healthcare providers' preferences. Typically, even information about doctors and their competence is very limited.

The knowledge of particular doctors, and even of particular procedures by those at the cutting edge of medical science, is also imperfect and there is thus often no "scientific" benchmark to balance doctors' or insurers' preferences. Some recent studies have identified examples of procedures, for instance, that are practiced much more frequently in one part of the country than in another, with no statistical evidence showing improved health conditions as a result.⁵⁸ To be sure, most of these procedures have little downside risk, and the patients receiving the treatment typically feel good. Indeed, honest doctors may even report that the procedure does not have benefits for all patients, and that they cannot predict with precision whether a particular patient will benefit. What few doctors will report, of course, is that there is no evidence that the procedure has any effect at all. When confronted with such statistical evidence, they will typically explain the faults in the study—how it left out one critical variable or another—and state that, *given the low risk, in their judgment the procedure is warranted.*

One would think that second opinions could alleviate these information problems and force doctors to justify their decisions both to their patients and to each other.⁵⁹ Typically, however, the second opinion comes from a doctor subscribing to the same local "norms." If it is generally believed in the community that tonsillectomies reduce the incidence of sore throats, then the second opinion is likely to recommend a tonsillectomy for a patient suffering from repeated sore throats—regardless of the evidence questioning the efficacy of the procedure. It is also perhaps worth noting that there is a guild effect (what in other parlance might be described as tacit collusion)—a reluctance by members within a profession to criticize others—which may also undermine the quality of information conveyed in a second opinion.⁶⁰

One of the advances in modern medicine, however, is the recognition of uncertainty and the openness about risks. Modern medicine confers a duty on doctors to perform

controlled studies and statistical tests of the validity of a procedure. It even trusts patients with making judgments when there are various risks associated with alternative procedures.⁶¹ Finally, modern medicine recognizes that an essential part of the successful practice of medicine is doctors' bedside manner, which can determine their ability to match patients' preferences as well as influence the extent to which their orders will be followed.

Pondering the issues faced by doctors, including incentives and uncertainty, has a lot to offer those of us engaged in providing economic advice, or providing economic assistance conditional on certain policies being adopted. To be sure, the problem facing economic advisers in actually implementing prescriptions is far more challenging than that facing doctors. Economists, unlike doctors, do not deal with the entity most affected by their advice. Patients rarely have to be convinced to save their own lives, but politicians often have to be convinced to place the public interests of their country ahead of their own private interests. There is a serious agency problem. At least in the short run (and for politicians, who have an incentive to delay problems until the next person's watch, often the only relevant horizon is the short run), there is a clear conflict of interest, a conflict that could be mitigated (but not eliminated) by a policy of consistent openness and transparency. Second, while doctors deal with decisionmaking by an individual, economists' advice typically has to be implemented by a political process in which the adviser is often an outsider. The economist-politician dialogue is part of a larger game with an audience. Imagine the consequences if others were permitted to listen to both what the doctor says and how the patient responds: a lender might not be so willing to lend money to a patient who has been diagnosed as having incurable cancer or who is refusing to take the medicine that could cure his or her cancer. The adviser's defense strategy of explaining poor results by pointing to the failure of the patient to follow his advice could well exacerbate the difficulties of the afflicted party. The inevitable audience in economic advising creates the added difficulty of such external reactions, particularly when information is released sporadically. The adviser thus is not just providing information, but becomes a player in the political-economic "game."

So far, we have described some of the ways in which the problems he faces are markedly different from (and more complicated than) those facing an individual's medical adviser. By the same token, assessing the incentives of the advisers is perhaps even more important, but also more complicated, than in the case of medical advice.

It should not, however, be concluded that simply because an advisory service is publicly provided it is "objective." Incentives can play a role, even if there are no direct monetary incentives. One has still to look at both individual and organizational incentives. To whom are the advisers accountable? Who pays them? What is the governance structure? Anyone who has dealt with governments, let alone participated in one, recognizes that when an elected president appoints different ministers, all with the mandate to serve the national interest, each soon takes on an advocacy role for his ministry and the special interests that it represents, even if he does not come from those special interests.⁶²

The problem of regulatory capture (Stigler 1981) is equally if not more acute in the so-called independent agencies set up to exercise expertise in a particular domain. Although some countries have tried to combine an independent central bank with expertise and representation, in many countries the central bank is governed largely by those drawn from the financial and business community, with no representation of the concerns of broader segments of society that are affected by monetary policy, such as workers. Finance ministries too, while often entrusted with providing a national overview of budgetary issues, typically have closer ties to the financial community from which they raise funds and draw their top officials, most of whom return to the financial community after government service. Is it a surprise, then, that the advice given by those accountable to these ministries and central banks often differs from that which might have been given by a labor ministry? Or that labor unions in East Asia have been sharply critical of some of the measures taken in the recent crisis? It may not represent a lack of understanding of the economic fundamentals so much as a different reading of the scientific evidence and, more important, a different weighting of the risks.

Like the market for healthcare, the market for economic advice contains many information imperfections and uncertainties that allow these incentive effects to have a strong influence on the “prescription” when there is no universally accepted *right* advice. In some ways the information imperfections are much worse. I have already described one reason for this: the patients’ reluctance to divulge relevant information to the adviser. But there are several others. Governments in many countries have a strong proclivity for secrecy. (This contrasts with the widespread belief in the *presumption* that in competitive markets there will be efficient information disclosure⁶³—a presumption that has seemingly been questioned in recent public policy debates by calls for more disclosure requirements.⁶⁴) Elsewhere (Stiglitz, 1998a, 1999), I have explained the various reasons for this proclivity—institutional and individual incentives. Secrecy provides more scope for the work of special interest groups, greater cover for corruption, and greater opportunities for hiding mistakes. It also creates rents—like any other artificially created scarcity—rents which can be, and are regularly, exchanged for things of value, including better, and often distorted, press coverage. Secrecy can also serve as an entry barrier, reducing the ability of other politicians and bureaucrats to compete effectively.

Most of these same arguments apply to advisers. They have, for instance, incentives to hide mistakes. The adviser, like the doctor, has to worry about his reputation. But in both cases it may be difficult if not impossible to make very accurate inferences about performance. Each heart transplant is different, and the total number is so few: is a 50 percent death rate a mark of a poor doctor or of a willingness to take the riskiest patients? Would the patients who died have died anyway? The only doctor in town may worry about prescribing too tough a medicine, lest sick patients be reluctant to come to him until it is too late; but he may also be tempted to shy away from easy remedies, even when they are just as effective as, or more effective than, tough medicine—the tough regimen shows that the doctor is treating a serious problem and knows what is required! Tough remedies also help preserve the image of “expertise”

versus lesser, “common sense” solutions. And the association of pain with redemption is also very firmly rooted in the religions and cultures of many countries.

At the very least, few of us have an incentive to expose our mistakes. In some cases, there may even be a guild mentality, imposing social sanctions on those who expose the mistakes of other members of the guild.

Secrecy by advisers cannot only serve to hide mistakes, but it can to create rents and reduce competition (partly by product differentiation, partly by creating barriers to entry by, for instance, restricting the flow of information, an essential ingredient in the production process of advice). Secrecy undermines the possibility of effective second opinions: the adviser can always say, “If you only knew what I knew—but can’t tell you—you would think differently about the matter.” There is, however, a marked contrast here in the norms of medical and other advisers: Doctors are expected to share fully in any information. In the case of medicine, it is not secrecy but guild behavior that undermines the effectiveness of second opinions. Just as tacit collusion among doctors may reduce the effectiveness of second opinions, so too in the context of economic advice, though matters here may be worse, as there is even a rationale for collusion: the international community must speak with one voice, and countries will be confused by any dissonance.

Note the contrast between Neville Keynes’ perspectives and those underlying this position. Keynes would have argued that if there are disagreements, if theory or evidence is ambiguous, then the economic adviser should lay out the evidence and arguments. With the insight of modern Bayesian statistics, he would suggest how the evidence might be aggregated under different loss functions (and he would be careful both to point out that the inferences depended on the loss function and that it was the country’s decision as to which was the appropriate loss function). To be sure, different individuals, interest groups within society, might employ different loss functions, and therefore weigh the evidence in different ways. But confidence in democratic political processes necessitates leaving it to the country to aggregate and weigh the information in the appropriate way: it is irresponsible for the adviser to suppress alternative perspectives that meet a certain credibility threshold.

There is another reason that democratic governments themselves may be unwilling to be more transparent and open about international agreements, and that is that those agreements may contain provisions that counter platforms upon which the government was elected or widespread views held within the electorate. But this in turn raises fundamental problems: should governments be encouraged to participate in such agreements, which may serve to undermine democratic processes themselves? To be sure, governments can be encouraged to engage in a national dialogue to reexamine such policies in light of new information.

Thus, there are a variety of reasons why governments may be reluctant to engage in open dialogue, or even to make economic information more readily available. The steps taken in recent years, under the leadership of the International Monetary Fund, to make data more readily available in standardized form are major advances. There is a consensus, however, that better data will not be enough to prevent all financial crises.

Should public agencies that detect a problem make a public statement that could, in fact, exacerbate the problem? The quandary in turn raises numerous questions about institutional arrangements: should statistical agencies, entrusted with gathering and disseminating information, be separated from policy agencies? There are potential conflicts of interest between the collection of data, the analysis of data, and the provision of “treatment” services, analogous in many ways to the conflict of interest in the healthcare market. Recognizing the inherent problems, many countries have chosen to create independent statistical agencies.

The analogy between medicine and economic advice has raised a number of important issues. It reminds us that advisers are every bit as subject to incentive concerns as are other participants in the economy. We should recognize these incentives, and that recognition should color interpretations of that advice—both on the part of recipients and the market more generally. Just as modern medicine recognizes the “patient’s right to know”—patients should be allowed to participate in the decisions that affect them, and should be given the relevant information to make that participation meaningful—so too should modern economic science strive to attain the role that Neville Keynes described so forcefully three-quarters of a century ago. Perhaps the most important question that the analogy raises is this: Have we yet reached the status of modern medicine, where we recognize the risks and uncertainties, or is there still a tendency to blame the patient for not following the advice precisely?

How to Improve Our Advice

As I said at the beginning of this talk, economics as a science can perhaps be best compared to astronomy. Like astronomers, we lack the ability to do controlled experiments. We rely on thought experiments and the natural experiments that history provides. As in astronomy, a theory can be overturned by a single, telling piece of evidence; one does not simply count arguments on one side or the other. And in both sciences, theory—a close examination of the internal consistency of explanations combined with what evidence we do have—can play a pivotal role. There was a high degree of confidence in the existence of black holes, for example, even before the telling evidence of their existence (in the pattern of rotating stars) was discovered, and even today, the number of observations remains very limited (see Gribben 1992).

If we are to move from ideology to science in dispensing policy advice, we need to demand the same attention to internal consistency of our theories and explanations. And our efforts to improve our predictive ability—said to be the hallmark of science—need to take this need for consistency into account. General statements do not seem to be very valuable: there are soothsayers who have predicted nine out of the last two crises, and revel in their success. By the same token, there are weaknesses in dozens of financial markets, and someone who calls attention to these weaknesses might rightly claim that she anticipated the problem; yet she has not distinguished the market that had the crisis from the dozens of others that did not have

a crisis. Growth forecasts should reflect anticipations of crises, and in that sense, none of the major international forecasters seems to have scored well.

But this is perhaps taking too narrow a view of economists' predictions. Those who argued that regulatory forbearance would exacerbate the U.S. banking crisis provided a prediction that was realized. But the prediction was more fine grained: those banks that were the most undercapitalized engaged in the riskiest behavior, just as the theory predicted. By the same token, there are implicit predictions in much of the policy advice: those who advised Kenya to liberalize its financial markets presumably predicted a reduction in spreads. The failure of those reductions to occur should call the underlying theory into question.

Many economic forecasters now focus on making conditional statements, with a ready explanation on hand when reality differs from that predicted. Too often, they engage in ex post rationalizations for these failures: the government did such and such, undermining the anticipated and desirable effects of the policy change. The factors that would undermine the workings of economic policy have to be specified in advance; and if the conditions for success are so onerous that they are unlikely to be met, then that should be taken into account in assessing the desirability of the policy itself.

There is room for far more research on the *modes* in which we give advice as well as on the *issues* on which we give advice. I have alluded to several examples of issues already, for instance, concerning the economic effects of financial market liberalization in highly underdeveloped capital markets.

Consider another issue that I have discussed extensively in this paper—the issue of secrecy. Those who would like to suppress open dialogue and criticism suggest that it may rattle markets, leading to greater uncertainty (and presumably greater risk premiums, or at least greater variance in prices), or that it may weaken a country's resolve to undertake “painful medicine.” Critics of these seemingly antidemocratic views point out, on the contrary, that information eventually comes out anyway, and that more open dialogue is akin to flexible exchange rates: it leads to a smoother flow of information and thereby enhances overall stability. And that open dialogue is likely to result in a consensus behind a policy—presuming that it is a policy that can be justified—and that the only way in which durable, sustainable policies can be implemented is through a process of consensus building. These are, in part at least, testable propositions. What is the evidence, for instance, that prices differ long from fundamentals, that the risk premium is increased by open dialogue?

To be sure, even if it should turn out that there is evidence that the economic costs of open discussion exceed the benefits, that does not end the matter. To return to my earlier theme, the task of the adviser and economist is to lay out the consequences of alternative policies. A country may decide that the advantages, in terms of strengthening democratic institutions, far outweigh the economic costs. As advisers, we should also point out the implicit or explicit incentives of various parties—that is, the incentive of some outside advisers and government officials to prefer secrecy. And the answers to how one weighs these tradeoffs may depend on the strength or fragility of democratic institutions in the country itself.

Similarly, earlier, I emphasized the importance of incentive structures for advisers. In the healthcare market considerable attention has been paid to the question of how those incentives can be altered, in ways that improve the quality of advice. Increasing attention is paid to “report cards” (disclosure of performance records), to increasing the scope for second opinions, to the fact that competition—with improved information—may attenuate some of the perverse incentives, to outside research which can break the strength of guild effects. Many within the health care industry have resisted these advances—for obvious reasons. So too, more attention needs to be paid to altering incentive structures in the context of economic advice. As in healthcare, “second opinions” and greater openness in policy debates could plausibly increase the incentive for economic advisers to act in ways that they can justify to their colleagues and clients. More representation in the governance structures of those proffering economic advice—so that they are less tied to particular interests—or more diversity in the sources of advice would serve to clarify that there is not typically a single Pareto-dominant policy; would provide incentives both to ensure that the impacts on, and risks to, all groups are fully taken into account; and would ensure that the advice is based on strong theory and evidence.

The task of putting our advice on a more scientific basis will be a difficult one, and may result in putting our own institutions and their incentives under closer scrutiny. But it is a quest upon which we must continue. Although it may have become unfashionable to refer to it, the long history of the encounters between many less developed countries and the more advanced countries has left its scars, along with its other, more positive legacies. In the minds of the colonial powers, the three or five decades that have passed may seem an eon, but to many of the afflicted, it is but a short moment in a far longer history of oppression. And many of these early encounters were based on a mixture of self-interest and ideology, with missionaries suppressing local cultures in the name of a higher authority. Although the early efforts at market access were somewhat less delicate—military threats rather than economic sanctions—these efforts were asserted with no less moral authority, even when they were for deeply immoral purposes, such as the opening of Chinese markets to drug traffic in the Opium Wars.⁶⁵

If the advice of outsiders is to be taken seriously, it must be based on reasoned argument—on science, on evidence, with a full recognition of the limitations and uncertainties that are associated with scientific evidence, not the confidence so typically associated with ideology; with the recognition that new evidence may, or indeed is likely to, require a change of policy, even perhaps a change in beliefs, a process of Bayesian sequential decisionmaking that is the foundation of modern decision sciences.

Concluding Remarks

We celebrate today the tenth anniversary of the Annual Bank Conference on Development Economics, a conference dedicated to the principle that economic science—and the promotion of research, dissemination, and dialogue about economics—can improve the chances of growth and the alleviation of poverty in the developing

countries of the world. In the decade since this series was initiated by Stanley Fischer, as the World Bank's chief economist, it has, in my judgment, proved its worth.

I hope that this meeting reflects well the breadth of concerns and the spirit that have marked each of the meetings over the past decade. Our discussions range from the role of geography to that of ethnic conflict, from the implications of financial markets for macroeconomic behavior to the importance of competition policy for microeconomic performance. The willingness of so many distinguished people to share their ideas and time speaks to the significance that this annual conference has attained within the development community.

In this paper, I have tried to identify several of the ways in which science, including economic science, is distinguished from ideology: a willingness to question everything and a recognition of the uncertainties associated with our knowledge, and the concomitant humility that that instills. In closing, I want to emphasize a further attribute of the scientific process: the value it places on openness and democracy. Scientific advances require an open exchange of information: universities are committed to the importance of free speech, and we, and other public agencies that fund research, insist that data be made publicly available, so that any results can be examined for replicability and accuracy. Open debate and discussion is both a natural part of the questioning spirit and the recognition of uncertainty, and a requisite for the successful advance of science.

A second attribute of science is that ideas and arguments are evaluated on their own merit, not on the basis of authority or received wisdom. In this sense, science is very democratic. Science pays no attention to status, social background, age, or any of the other myriad characteristics that form part of our social structures. Even a doctorate from a first-rate university does not give one's opinions or evidence any more weight before the court of scientific opinion.

It is with this spirit—the spirit of questioning, of the recognition of the limitations of our knowledge and our quest to expand the bounds, with the full recognition of the uncertainties, a spirit of open and democratic debate and discussion—that I look forward to the discussion of the next two days.

Notes

1. There is no standard terminology for these sets of doctrines, and various practitioners advocated these doctrines with varying degrees of subtlety and emphasis. The set of views is often summarized as the "Washington Consensus," though to be sure, there never was a consensus even in Washington (let alone outside of Washington) on the appropriateness of these policies. For an excellent and subtle articulation of the Washington consensus and its evolution, see Williamson (1990).

2. See for instance Rodrik (1996), who compares the policies pursued by the Republic of Korea and Taiwan (China) to the 10 elements of the Washington Consensus identified by Williamson (1990). Rodrik finds that judging by the number of prescriptions these countries did or did not follow, we would have to award the Republic of Korea a score of about 5 (out of 10) and Taiwan (China) about 6. He notes that these economies followed most of the macroeconomic prescriptions, but departed from many of the microeconomic recommendations, especially relating to liberalization and deregulation.

3. See World Bank (1997b). Such statistics, while they make a powerful point, need to be interpreted with caution: if China's provinces are to be treated as separate data points, so too should the large provinces of other low- and middle-income countries, such as Brazil and India. I suspect that even then the picture will not be altered much.

4. Some unofficial estimates present a similar picture, although the magnitudes of the declines are less severe. Studies that take into account the shift of production into the unmeasured informal sector (see Kaufmann and Kaliberda 1996) and trends in consumption (see Milanovic 1997) have found that the per capita income decline in the former Soviet Union is closer to one-quarter than one-third of GDP and the declines in Eastern Europe are a few percentage points lower than the official estimates. But there is another set of arguments that suggests that measured GDP may be too high: GDP records transactions as if they occurred at posted prices; but today in Russia and some other countries of the former Soviet Union, a large percentage of transactions are barter. The barter prices are, it is argued, typically lower than posted prices, and if these barter prices were used, GDP would be markedly smaller (see Illarionov 1997). Overall, however, almost no one claims that standards of living in Eastern Europe and the former Soviet Union have improved very much in the last decade, and any improvements have certainly been dwarfed by the enormous progress in China. Moreover, the huge cutback in military expenditures should itself have meant that—even at a given level of GDP—consumption should have risen.

5. From the current perspective, it appears that the declaration of victory in transforming Russia into a market economy by those who had helped design the reform program was a bit premature (see, for example, Aslund 1995 and Fischer 1994). While lauding statistics on the extent of privatization, they failed to note the virtual absence of investment in manufacturing. Privatization is easy to achieve—one can simply give away, or give at below fair market value, national assets to one's cronies and political supporters. Done this way, privatization can serve to undermine confidence in both markets and political processes. Neither should interesting foreign investors in extracting a country's natural resources—at sufficiently favorable terms and with sufficient guarantees—be viewed as a major achievement in the transformation into a modern industrial economy.

6. See, for instance, Greenwald and Stiglitz (1988, 1992, 1993a, 1993b) and the papers on "The Credit Market" collected in Mankiw and Romer (1991).

7. See, for instance, Hirshleifer and Riley (1979) or Stiglitz (1985). For an overview of the implications for macroeconomic behavior, see Stiglitz (1992). This research in turn was motivated in part by observations concerning the fluctuations of economic activity over the 200 or more years of capitalism that could not be reconciled with standard macroeconomic models that ignored finance. In the search to find idiosyncratic explanations for the East Asia crisis, the long history of financial crises seems to have been forgotten.

8. New classical models originate with Lucas (1972) and real business cycle models with Kydland and Prescott (1982) and Prescott (1986). For a trenchant critique of real business cycle models see Mankiw (1989). For a contrast between the alternative approaches, see Greenwald and Stiglitz (1993a).

9. The concept of a knowledge bank was introduced by World Bank President James Wolfensohn (1996) in his address to the 1996 Annual Meetings of the World Bank and the International Monetary Fund

10. See World Bank (1998a) for a good discussion. For particular aspects of institutions and economic management, see Burnside and Dollar (1997), Fischer (1993), Knack and Keefer (1995), and Sachs and Warner (1995). The term *good institutions and economic management* might seem to beg the question: by definition, are good institutions and economic management those that lead to better economic performance? The *content* arises from the *independent* specification of these policies, which include, for example, low public-sector deficits, low inflation, and low levels of corruption. Moreover, these policies not only increase growth, but increase the returns to development aid (World Bank 1999).

11. See, for instance, Braverman, Hoff, and Stiglitz (1993) and Hoff and Stiglitz (1990).
12. Another important aspect of the knowledge revolution is that a central activity of the economy is the production and dissemination of knowledge and information; the innovation process has been systemized. Other implications of the knowledge revolution, for example, for the organization of production, for competition and for public policy more generally, are discussed in Stiglitz (1999).
13. The book is *In Search of Excellence* (Peters and Waterman, 1982). One good critique (Kolodny, Lawrence, and Ghosh, 1989) finds that that *by the authors' criteria* these companies displayed mean reversion in the years following publication of the book. They also find that judged by better motivated criteria, including risk adjustment, the performance of the "excellent" companies is little better than that of a randomly selected control group *even in the authors' sample*.
14. See, for example, Edward Prescott (1986), who writes that economic fluctuations are optimal responses to uncertainty in the rate of technological change.
15. One can even formalize this in terms of Bayesian statistics, where statistical estimates depend on loss functions, that is, on values.
16. It is not my purpose here to review in detail the causes or consequences of the East Asia crisis or to present appropriate responses to it. Those are matters that I have taken up elsewhere. See Stiglitz (1998b) and Furman and Stiglitz (1998).
17. The fiscal costs alone of resolving these crises were substantial. Finland's (1991–93) cost 8.0 percent of GDP, Sweden's (1991) 6.4 percent of GDP, and Norway's (1987–89) 4.0 percent of GDP. By way of contrast the fiscal cost of resolving banking crises ranges from the relatively low 3.2 percent of GDP for the U.S. savings and loan debacle (1984–91) to 55.3 percent of GDP for Argentina's 1980–82 banking crisis (Caprio and Klingebiel 1996).
18. To be sure, distorted agricultural prices may have contributed to high real estate prices in some regions; but these distorted agricultural policies, and in particular changes in these policies, seem to have played no role either in the generation of the real estate bubble or its bursting.
19. For instance, distorted real estate tax policies may well have contributed to the real estate boom in the United States in the 1980s, and the bust after that boom certainly contributed to the savings and loan crisis. Even so, eliminating the preferential tax treatment in the middle of the crisis would have significantly deepened the downturn.
20. See Wade (1990), World Bank (1993), and Stiglitz (1996).
21. The often-discussed massive misallocation of investment funds in the U.S. automobile market is seldom attributed to "crony capitalism"—though agency problems almost surely played an important role.
22. To be sure, this methodology is not sufficiently robust as to give unambiguous results.
23. Of course, as the crisis evolved, many loans that under normal circumstances would have been perfectly good went sour: the huge, and unanticipated, rise in interest rates implied that even investments in projects with very high expected returns might not be able to repay what was borrowed, especially given the high degree of leverage, particularly short-term debt, in East Asia. Thus some of the "bad" performances should be attributed not to bad lending practices but to the severe macroeconomic shock.
24. For a discussion of the consequences of these relative performance reward structures, see Nalebuff and Stiglitz (1983).
25. In particular, I have suggested that at least part of the problem in, say, Thailand lies in excessively rapid liberalization (without putting into place the concomitant regulatory structures): had the government maintained the stricter regulations on lending to real estate, perhaps the real estate bubble would not have occurred. I also have questioned the role that cronyism played in excessive lending: after all, the marginal lending was done by Western and

Japanese banks, which were surely not subject to political pressures to lend; and those banks lent into a situation where Korean firms were already highly leveraged and Korean banks highly vulnerable because of the high debt-equity ratios of their borrowers.

26. See Stiglitz and Weiss (1981). Empirical evidence supporting the importance of this phenomenon includes Drèze, Lanjouw, and Sharma (1997) for India and Biggs and Srivastava (1996) for Sub-Saharan Africa. In the context of capital flows to developing countries, see Eaton and Gersovitz (1981).

27. Frankel and Schmukler (1998) find convincing evidence that country fund holders, who are usually outside investors, have systematically less information about local assets than do the holders of the underlying assets, usually local residents and some large international investors. Greenwald (1998) examines the implications of the asymmetry of information between local and outside investors.

28. It is noteworthy that the International Monetary Fund (IMF) review of the Mexican experience emphasized the role of domestic capital flight, finding that the available data show that the pressure on Mexico's foreign exchange reserves during 1994 and prior to the devaluation came not from the flight of foreign investors or from speculative position-taking by these investors, but from Mexican residents (IMF 1995). Frankel and Schmukler (1996) find systematic evidence supporting this conclusion. By most accounts domestic flight has also played an important role in Indonesia. The reasons for differences in the speed and pattern of domestic and international responses are clear: domestic investors typically are more informed about what is going on within their country. Moreover, the high correlation between returns on human and physical capital within the country mean that domestic investors are less diversified and therefore should act in a more risk-averse manner.

29. Two of the earlier studies are Tyron (1979) and Frankel (1980); and see Hodrick (1988) for a survey. There is much debate about why the joint assumptions of rational expectations plus risk neutrality are rejected in the foreign exchange market, and indeed in most other financial markets. The common explanation that these rejections can be explained by changing risk premia can, from one perspective, be viewed as a tautology: the risk premium is defined to equate the difference between the "expected return" from holding the two different currencies. If there is to be content in this explanation, it must be possible to *predict* changes in risk premia, or at least to *correlate* these changes with contemporaneously observable variables, which can themselves be thought of as indicating a change in risk premia. Viewed from this perspective, the risk premia explanation does not fare well (see Frankel and Froot 1989).

30. The classic paper by Meese and Rogoff (1983) found that none of the major structural models could explain out-of-sample exchange rate movements, and they were all dominated by a simple random walk model.

31. These rolling correlations are based on ongoing work by Sergio Schmukler, World Bank.

32. There is a large literature beginning with Vickers (1986) on signaling in monetary policy.

33. A fascinating study by Romer and Romer (1996) finds that the U.S. Federal Reserve's forecasts of inflation are substantially more accurate than private forecasts. It also finds that monetary policy actions reveal some of this private information, leading commercial forecasters to revise their expectations accordingly.

34. Romer and Romer (1996) show that the information revelation from a monetary tightening (the indication that an economy is in the bad state of high expected inflation) more than offsets its direct economic effects, thus leading commercial forecasters, on average, to revise *up* their expectations for inflation. This finding begs the question of why the U.S. Federal Reserve does not release its contemporaneous forecasts, in contrast to, say, the Administration and the Congressional Budget Office; both of which publicly reveal the forecasts underlying their budgetary policies and proposals.

35. There has, however, been extensive work on the importance of signaling in foreign exchange markets. Agénor (1994) provides a theoretical model in which signaling can help maintain a good equilibrium, although the signals he describes—the removal of capital controls, a drastic cut in the budget deficit, the appointment of a conservative central banker—do not explicitly include high interest rates. Dominguez and Frankel (1990) find that even sterilized interventions can, at least under certain circumstances, affect exchange rates, and ascribe the majority of this effect to the signaling component of the intervention. Wantanabe (1994) finds similar results for Japan. Interestingly, in these analyses foreign exchange interventions are important because they signal the commitment to a stronger currency, implicitly opening up the possibility of further interventions, and possibly even tighter monetary policy, in the future. In contrast, when a country uses high interest rates to defend its currency, usually the authorities' announced intention and the economy's expectation is that these rates will be lowered over time.

36. This thus provides an additional argument to the several already in the literature against excessive independence and lack of representation of central banks. See Stiglitz (1998c)

37. Allan Drazen and Paul Masson provide a nice story illustrating this point: "One afternoon a colleague announces to you that he is serious about losing weight and plans to skip dinner. He adds that he has not eaten for two days. Does this information make it more or less credible that he really will skip dinner?"

38. One possibility is government interventions, which impose Pigouvian corrective taxes on those activities that generated the systemic risks, with revenues being used to provide for improved safety nets, which, in addition to their direct benefits will send a signal that the government's reform efforts will be politically and socially sustainable.

39. The reason for multiple equilibria is easy to see. Normally exchange rate depreciation leads to increased exports and thus a greater demand for local currency. A very large depreciation, however, can result in a substantially higher probability that firms will default on their foreign currency-denominated debts. This, together with weaker economic conditions, reduces capital inflows, resulting in a decrease in the demand for the currency that more than offsets the increased demand for local currency to buy exports. Giffen-like demand curves (as well as backward-bending supply curves) can give rise to multiple equilibria. In an intertemporal model exchange rates today depend on expectations about the equilibrium exchange rate in the future. Higher interest rates leading to lower investment and less competitiveness lead to lower exchange rates in the future. More important, the resulting shifts in the demand and supply curves may lead to the elimination of the "upper" equilibrium exchange rate.

40. Even for countries that are not *net* debtors, particular firms will be debtors while others will be creditors. One of the important lessons to emerge from the new finance-based macroeconomic models is that even if the increase in net worth to creditors from a currency depreciation offsets the losses of debtors (and conversely for currency appreciations), the net macroeconomic effects can be significant and adverse, as the increased investment and employment activity of the gainers will not be sufficient to offset the reductions on the part of the losers. (This is because the relationship between investment, including investments in inventories and new hires, and net worth is concave.) For a discussion of this intuition in the context of changing interest rates see Greenwald and Stiglitz (1993a).

41. This has been an especially virulent criticism of cross-country growth regressions. Examples of this problem abound. Pack (1994), for instance, presents a wide-ranging critique of cross-country growth regressions, pointing out that many of the variables on the right-hand side of the regressions are not really exogenous. For instance, government expenditures are clearly endogenous and could well depend not just on past growth but also on expected future growth (through wealth effects). Most attempts to deal with this kind of simultaneity have been unconvincing. To be sure, this is not a criticism of cross-country regressions per se, only of *bad* cross-country regressions. But the seeming dominance in the field of studies plagued by these and other econometric problems both suggests how difficult it is to make

inferences based on empirics alone and how cautious we should be in applying any so-called results to actual policy situations.

42. Thus while bailouts in general contribute to moral hazard—lack of due diligence by lenders in making loans—bailouts motivated by a worry concerning the adverse effects of further devaluations on borrowers with uncovered foreign exchange exposure and their creditors are even more pernicious, for they contribute to a specific form of moral hazard, excessive lending/borrowing in uncovered foreign exchange. They thus contribute directly to vulnerability to a crisis.

43. Obviously, a country that is borrowing from abroad to finance investment in nontradables will inherently face a foreign exchange exposure risk; but typically, aggregate foreign borrowing does not exceed aggregate investment in tradables, and hence even for longer-term investments, countries can limit their exposure, at least in the aggregate. Many of the cases of excessive foreign exchange exposure reflect misperceptions on the part of borrowers and lenders of the foreign exchange risks, a misperception to which domestic exchange rate policy may have contributed. The repeated instances of bailouts may in fact imply that private lenders may not have faced risks anywhere commensurate with the social risks.

44. After this paper was presented, surveys in Thailand confirmed some of these expectations. The firms with large foreign exchange exposures—besides the commercial real estate firms and the banks that had lent to them, there were already bankrupt, and that would not, in any case, be engaged in further investment in the immediate future, given the high vacancy rates—were the exporting firms, who effectively had cover. The adverse effects of high interest rates on the economy would seem to overwhelm those of the devaluation. (To be sure, the shift of resources, even in a full employment economy, from non-tradeables to tradeables, required if the gap in the current account was to be closed, had adverse effects on those in the non-tradeable sector, but that is part of the inevitable cost of restructuring; and to be sure, those in the export sector complain about the increased cost of inputs resulting from devaluation—they would have preferred simply to see their export prices rise in domestic terms—but this too is an inevitable part of the restructuring from which they gain net.) See Dollar and Dreimeier (1998).

45. For a comprehensive discussion, see Campbell, Lo, and MacKinlay (1997), especially chapter 2.

46. This point was the center of the famous Jackson Hole debate, in which all the economists argued in support of this proposition, while many of the central bankers made assertions to the contrary (without presenting either theory or evidence in support of their beliefs) See Federal Reserve Bank of Kansas City (1995) for a summary of the discussions.

47. There is an important strand of the literature, however, which focuses not on the capital asset equilibrium, but on short-run capital flows. Larger fiscal deficits are associated with larger borrowing from abroad, or decreased holding of reserves. Reduced fiscal deficits lead to reduced need for foreign borrowing and thereby a stronger currency. Although such models provide a rationale for the focus on current account deficits, the evidence that such deficits help predict currency crises is, at best, limited (Frankel and Rose 1996). Part of the reason is that the real issue has to do not with the size of the deficit, but with the ability of the country to finance the deficit. This depends more on stock (state) variables, such as the overall debt or beliefs, than on flow variables (except to the extent that these affect the stock variables, for example, through changing beliefs).

48. Indeed, Mexico *seems* to have recovered, even though the weaknesses in its financial system were not really adequately addressed—precisely because Mexico was not dependent on its own banking system. More recently, questions are even being raised concerning Mexico's recovery: while the export sector is strong, the rest of the economy—dependent on the domestic financial system—remains weak.

Similar words of caution need to be applied in interpreting the *eventual* recovery of, say, Thailand and Korea. There will be those who will be tempted to engage in the *post hoc ergo*

propter hoc fallacy: there was a bailout, and after the bailout the exchange rates stabilized, and the economies eventually recovered—therefore, the bailout, with the accompanying policies, “worked”—they can be credited with the subsequent successes.

49. For an application to financial markets, see Jaffee and Stiglitz (1990).

50. This is the case for Kenya. For further evidence from Pakistan, see Aleem (1990).

51. There is a large literature showing that without government regulation, problems of moral hazard and looting behavior result. See, for example, Akerlof and Romer 1993.

52. Studies have used cross-country regressions to show that growth is positively associated with high real interest rates, a finding that is interpreted as evidence that financial repression (which, according to these studies, is manifested by low real interest rates) is bad for growth (see, for example, Gelb 1989). These studies, however, were shown to be badly flawed for several reasons. The observed results were dominated by the extreme cases of very negative real interest rates. Also, these estimates were biased by simultaneity problems. Implicitly they were based on a model that government policies (in particular not interfering in the setting of interest rates) led to higher real interest rates, which led to faster growth. But it is probably more plausible that high real interest rates and high growth are both simultaneously caused by high productivity. Finally, the regressions upon which these results were based were misspecified. Low real interest rates are correlated with high inflation, a symptom of bad macroeconomic policies, and it is these bad policies that lead to low growth; including inflation as a separate variable eliminates the alleged negative effect of financial repression on growth. See Murdock and Stiglitz (1993) and the discussion in World Bank (1993) and Stiglitz (1994a).

53. On the other hand, the capacity of sophisticated capital markets to circumvent regulations may be greater.

54. Not only on deposit rates (which can help enhance franchise value, as I have already noted), but also on lending rates. When there is deposit insurance (implicit or explicit) with premiums imperfectly adjusted for risk, and without appropriate adjustments in risk adequacy standards, banks have an incentive to undertake excessively risky lending. Ceilings on lending rates or restrictions on certain types of loans can partially offset this proclivity.

55. In this discussion I have not even touched on the inadequacies in the way that capital adequacy standards are actually implemented—the lack of appropriate treatment of risk and inadequate accounting frameworks; these inadequacies imply that excessively rigid enforcement of capital adequacy standards can actually lead to more risk taking and more vulnerability by the banking system. See Stiglitz 1999.

56. To be sure, there are other gains that have to be weighed in the balance—the possibility of greater portfolio diversification actually reducing risk, or the increased efficiency from the discipline provided by open capital accounts.

57. Indeed, privatization typically reduces the scope for intergenerational risk sharing and, with few exceptions, fails to provide full insurance against inflation. Thus any efficiency gains from privatization have to be balanced against the additional risks that individuals typically have to bear. Even for a highly developed capital market such as the United States, it is estimated that privatization will substantially increase transactions costs. See U.S. Council of Economic Advisers (1997).

58. See, for instance, Wennberg, Freeman, and Culp (1987) and Chassin and others (1987).

59. If health maintenance organizations (HMOs) are required to pay for second opinions, for example, this might increase the pressure on doctors to perform a satisfactory analysis the first time. On the other hand, it might create pressure for the doctors to appear more sure of themselves and downplay the uncertainty and cost-benefit analysis involved in choosing a particular treatment.

60. In the United States the guild effect has been somewhat undermined by other incentives, particularly those provided by malpractice suits.

61. Experiments with patient-physician decisionmaking models have shown a divergence between patients' and physicians' preferences: when clients and doctors worked together in two-staff model HMOs, for example, there was a 40 percent decline in surgery rates on non-cancerous enlarged prostates (Robert Wood Johnson Foundation 1997).

62. For my discussion of this issue in the context of the United States, see Stiglitz (1997).

63. Again, there is a conflict between *ideology* and economic science. Work on the economics of information over the past 20 years has made it clear that there may not be efficient incentives for information revelation or information gathering. See, for example, Stiglitz (1975a, 1975b) and Grossman and Stiglitz (1980). Grossman (1981) makes it clear how restrictive the conditions are that are required for markets to provide efficient information disclosure. The widespread recognition of the imperfections in the market for information is today reflected in the numerous disclosure requirements in securities laws, for instance.

64. Most of these calls, however, are for the disclosure of aggregate information, for example, about the total foreign indebtedness of the firms within a country. Note that the call for better information is not consistent with the ideology of a free market economy. To the extent that believers in such ideology attempt to derive its support from the fundamental theorems of welfare economics, they should believe that all relevant information is conveyed in prices. On a more concrete level, the most recent crises have been related to *private* capital flows, and a mark of capital market liberalization is the freedom of agents to engage in such transactions without reporting to the government. It is also interesting to note the curious irony behind those arguing for decentralization, on the one hand, and the need for the kind of aggregate information, which such decentralization typically makes hard to collect.

While we have emphasized the political nature of advice (that there is not a single Pareto-dominant policy), data collection and analysis is itself not necessarily completely free of political judgments and interest considerations. Data (accounting frameworks) focus attention on certain variables; information affects actions—which is precisely why accounting frameworks are so important. For instance, discussions of the bias in the CPI in the United States had strong implications not only for the adjustment of social security payments, but also for wage negotiations and assessments of macropolicy (the extent to which inflation was or was not a problem, the extent to which productivity was or was not stagnating.) Concerns have been raised about whether the Department of Labor, which is entrusted with calculating the CPI, because of institutional mandates, was less receptive to recognizing the nature of the bias than it should have been. Other biases can arise in agencies with close ties to financial markets.

65. For a good discussion of the Opium Wars see Fay (1997). It is hard to escape the irony between the early drug wars—Western powers trying to keep China open to the flow of drugs—and the more recent equally adamant stands trying to stem the flow of drugs into their own countries. Only the lapse of time—and lack of knowledge of these historical experiences—softens what would otherwise seem an intolerable level of hypocrisy.

References

- Agénor, Pierre-Richard. 1994. "Credibility and Exchange Rate Management in Developing Countries." *Journal of Development Economics* 45: 1–16.
- Akerlof, George, and Paul Romer. 1993. "Looting: The Economic Underworld of Bankruptcy for Profit." *Brookings Papers on Economic Activity*, pp. 1–60.
- Aleem, Irfan. 1990. "Imperfect Information, Screening, and the Costs of Informal Lending: A Study of a Rural Credit Market in Pakistan." *World Bank Economic Review* 4 (3). 329–49.
- Arrow, Kenneth. 1963. *Social Choice and Individual Values*. New York: Wiley.
- Aslund, A. 1995. "How Russia Became a Market Economy." Brookings Institution Press 378. Washington, DC.
- Azariadis, Costas. 1981. "Self-Fulfilling Prophecies." *Journal of Economic Theory* 25: 380–96.

- Bäckstrom, Urban. 1997. "What Lessons Can Be Learned from Recent Financial Crises? The Swedish Experience." In *Maintaining Financial Stability in a Global Economy. A Symposium Sponsored by the Federal Reserve Bank of Kansas City*.
- Biggs, Tyler, and Pradeep Srivastava. 1996. *Structural Aspects of Manufacturing in Sub-Saharan Africa: Findings from a Seven Country Enterprise Study*. World Bank Discussion Paper 346. Africa Technical Department Series. Washington, D.C.
- Braverman, Avishay, Karla Hoff, and Joseph E. Stiglitz. 1993. *The Economics of Rural Organization: Theory, Practice, and Policy*. New York: Oxford University Press
- Burnside, Craega, and David Dollar. 1997. "Aid, Policies, and Growth." Policy Research Working Paper 1777. World Bank, Washington, D.C.
- Cambell, John, Andrew Lo, and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
- Caprio, Gerard. 1997. "Safe and Sound Banking in Developing Countries: We're Not in Kansas Anymore." *Research in Financial Services: Private and Public Policy* 9: 79–97.
- Caprio, Gerard, and Daniela Klingebiel. 1996. "Bank Insolvencies: Cross-Country Experience." Policy Research Working Paper 1620. World Bank, Washington, D.C.
- . 1997. "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1996*. Washington, D.C.: World Bank.
- Caprio, Gerard, and Lawrence Summers. 1996. "Financial Reform: Beyond Laissez Faire." In Dimitri Papadimitriou, ed., *Stability of the Financial System*. Macmillan.
- Caves, Douglas, and Laurits Christensen. 1980. "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads." *Journal of Political Economy* 88 (5): 958–76.
- Chassin, Mark, and others. 1987. "Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services?" *Journal of the American Medical Association* 258: 2533–37.
- Clay, K., Bruce Greenwald, and Joseph E. Stiglitz. 1990. "Money Neutrality in a Model of Firm Adjustment." Working Paper. Stanford University, Stanford, Calif.
- Coffee, John. 1996. "Institutional Investors in Transitional Economies: Lessons from the Czech Experience." In R. Frydman, C. W. Gray, and A. Rapaczynski, eds., *Corporate Governance in Central Europe and Russia. Vol. 1, Banks, Funds, and Foreign Investors*. London: Central European University Press.
- Cooper, Russel. 1990. "Predetermined Wages and Prices and the Impact of Expansionary Government Policy." *Review of Economic Studies* 57 (2): 205–14.
- Demirgüç-Kunt, Asli, and Enrica Detragiache. 1998a. "The Determinants of Banking Crises in Developing and Developed Countries." *International Monetary Fund Staff Papers* 45 (1): 81–109.
- . 1998b. "Monitoring Banking Sector Fragility: A Multivariate Logit Approach with an Application to the 1996–97 Banking Crises." World Bank and International Monetary Fund, Washington, D.C.
- Dollar, David, and M. Hallward-Driemeier. 1998. "Crisis, Adjustment, and Reform in Thai Industry." World Bank, Washington, D.C.
- Dominguez, Kathryn, and Jeffrey Frankel. 1990. "Does Foreign Exchange Intervention Matter? Disentangling the Portfolio and Expectations Effects for the Mark." National Bureau of Economic Research Working Paper 3299. Cambridge, Mass.
- Drazen, Allan, and Paul Masson. 1994. "Credibility of Policies Versus Credibility of Policymakers." *Quarterly Journal of Economics* (August): 735–54.
- Drèze, Jean, Peter Lanjouw, and Naresh Sharma. 1997. "Credit in Rural India: A Case Study." LSE STICERD Development Economics Research Paper 6.

- Eaton, Jonathan, and Mark Gersovitz. 1981. "Debt with Potential Repudiation: Theoretical and Empirical Analysis." *Review of Economic Studies* 48 (2): 289–309.
- Edlin, Aaron, and Joseph E. Stiglitz. 1995. "Discouraging Rivals: Managerial Rent-Seeking and Economic Inefficiencies." *American Economic Review* 85 (5): 1301–12.
- Ellerman, David. Forthcoming. "Voucher Privatization with Investment Funds: An Institutional Analysis." Policy Research Working Paper. World Bank, Washington, D.C.
- Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* (May).
- Fay, Peter W. 1997. *The Opium War, 1840–1842*. Chapel Hill: University of North Carolina Press.
- Federal Reserve Bank of Kansas City. 1995. *Budget Deficits and Debts: Issues and Options. A Symposium Sponsored by the Federal Reserve Bank of Kansas City*.
- Fischer, Stanley. 1993. "The Role of Macroeconomic Factors in Growth." *Journal of Monetary Economics* 32: 485–512.
- . 1994. "Russia and the Soviet Union: Then and Now." In O. J. Blanchard, K. A. Froot, and J. D. Sachs, eds., *The Transition of Eastern Europe. Vol 1. Country Studies*. National Bureau of Economic Research Project. University of Chicago Press.
- Frankel, Jeffrey. 1980. "Tests of Rational Expectations in the Forward Exchange Market." *Southern Economic Journal* 46 (4).
- Frankel, Jeffrey, and Ken Froot. 1989. "Forward Discount Bias: Is It an Exchange Risk Premium?" *Quarterly Journal of Economics* 104 (1): 139–61.
- Frankel, Jeffrey, and Andrew Rose. 1996. "Exchange Rate Crashes in Emerging Markets: An Empirical Treatment." *Journal of International Economics* 41: 351–66.
- Frankel, Jeffrey, and Sergio Schmukler. 1996. "Country Fund Discounts and the Mexican Crisis of December 1994: Did Local Residents Turn Pessimistic Before International Investors?" *Open Economies Review* 7.
- . 1998. "Country Funds and Asymmetric Information." Policy Research Working Paper 1886. World Bank, Washington, D.C.
- Furman, Jason, and Joseph E. Stiglitz. 1998. *Economic Crises: Evidence and Insights from East Asia*. Brookings Papers on Economic Activity. Presented September 3–4.
- Gelb, A. 1989. "Financial Policies and Growth." Policy Research Working Paper 202. World Bank, Washington, D.C.
- Gorman, W. 1953. "Community Preference Fields." *Econometrica* 21: 63–80.
- Grandmont, Jean Michel. 1985. "On Endogenous Business Cycles." *Econometrica* 53 (5): 995–1045.
- Greenwald, Bruce. 1998. "International Adjustments in the Face of Imperfect Financial Markets." Paper presented to the Annual Bank Conference on Development Economics, Washington, D.C., April 20–21.
- Greenwald, Bruce, and Joseph E. Stiglitz. 1986. "Externalities in Markets with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics* 101 (May): 229–64.
- . 1988. "Examining Alternative Macroeconomic Theories." *Brookings Papers on Economic Activity* 1: 207–70.
- . 1992. "Information, Finance and Markets: The Architecture of Allocative Mechanisms." *Industrial and Corporate Change* 1 (1): 37–63.
- . 1993a. "Financial Market Imperfections and Business Cycles." *Quarterly Journal of Economics* 108 (1): 77–114.
- . 1993b. "New and Old Keynesians." *Journal of Economic Perspectives* 7 (1): 23–44.

- Greenwald, Bruce, Joseph E. Stiglitz, and Andrew Weiss. 1984. "Informational Imperfections in Capital Markets and Macroeconomic Fluctuations." *American Economic Review* 74 (2): 194–99.
- Greider, William. 1987. *Secrets of the Temple: How the Federal Reserve Runs the Country*. New York: Simon and Schuster.
- Gribben, John. 1992. *Unveiling the Edge of Time: Black Holes, White Holes, Wormholes*. New York: Three Rivers Press.
- Grossman, Sanford. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24: 461–84.
- Hellman, Thomas, Kevin Murdock, and Joseph E. Stiglitz. 1997. "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?"
- Hirshleifer, J., and J. G. Riley. 1979. "The Analytics of Uncertainty and Information—An Expository Survey." *Journal of Economic Literature*. 17 (4): 1375–1421.
- Hodrick, Robert. 1988. "The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets." In *Fundamentals of Pure and Applied Economics*. Harwood Academic Publishers.
- Hoff, Karla, and Joseph E. Stiglitz. 1990. *Imperfect Information and Rural Credit Markets: Puzzles and Policy Perspectives*. University of Maryland, Department of Economics Working Paper 90-25.
- Illarionov, A. 1997. "Rules of Global Inflation." *Problems of Economic Transition* 40 (3): 21–53.
- International Monetary Fund. 1995. *International Capital Markets: Developments, Prospects, and Policy Issues*. Washington, D.C.
- Jaffee, D., and Joseph E. Stiglitz. 1990. "Credit Rationing." In B. Friedman and F. Hahn, eds., *Handbook of Monetary Economics*. Amsterdam: Elsevier Science Publishers.
- Kaminsky, G., and C. Reinhart. 1996. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." Federal Reserve Board, Washington, D.C.
- Kanbur, R. 1987. "Structural Adjustment, Macroeconomic Adjustment and Poverty: A Methodology for Analysis." *World Development* 15: 1515–26.
- Kaufmann, Daniel, and A. Kaliberda. 1996. "Integrating the Unofficial Economy into the Dynamics of Post-Socialist Economies." In B. Kaminsky, ed., *Economic Transition in Russia and the New States in Eurasia*. New York: M. E. Sharpe.
- Keynes, John Neville. 1919. *The Scope and Method of Political Economy*. London: Macmillan.
- Kirman, Alan. 1992. "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives* 6 (2): 117–36.
- Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance. Cross-Country Tests Using Alternative Institutional Measures." *Economics and Politics* 7 (3): 207–27.
- Kolodny, Richard, Martin Lawrence, and Arabinda Ghosh. 1989. "In Search of Excellence .. for Whom?" *Journal of Portfolio Management* 15 (3): 56–60.
- Kydland, Finn, and Edward Prescott. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica* 50 (6): 1345–70.
- Levine, Ross. 1997. "Financial Development and Economic Growth." *Journal of Economic Literature* 35 (2): 688–726.
- Lewbel, A. 1989. "Exact Aggregation and a Representative Consumer." *Quarterly Journal of Economics* 104: 622–33.
- Lewis-Beck, Michael. 1988. *Economics and Election*. Ann Arbor: University of Michigan Press.

- Lo, Andrew, and A. Craig MacKinlay. 1990. "When Are Contrarian Profits Due to Stock Market Overreaction?" *Review of Financial Studies* 3: 175–208.
- Lucas, Robert. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4 (2): 103–24.
- Mankiw, N. Gregory. 1989. "Real Business Cycles: A New Keynesian Perspective." *Journal of Economic Perspectives* 3 (3): 79–90.
- Mankiw, N. Gregory, and David Romer, eds. 1991. *New Keynesian Economics*. Cambridge, Mass.: MIT Press.
- Meese, Richard, and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14: 3–24.
- Milanovic, Branko. 1997. *Income, Inequality, and Poverty during the Transition from Planned to Market Economy*. Washington, D.C.: World Bank.
- Murdock, Kevin, and Joseph E. Stiglitz. 1993. "The Effect of Financial Repression in an Economy with Positive Real Rates." Background paper for *East Asian Miracle*. World Bank, Washington, D.C.
- Nalebuff, Barry, and Joseph E. Stiglitz. 1983. "Information, Competition and Markets." *American Economic Review* 73 (2): 278–84.
- Nikitin, Georgiy, and Andrew Weiss. 1997. "Performance of Czech Companies by Ownership Structure."
- Orszag, Peter. 1991. "Congressional Oversight of the Federal Reserve: Empirical and Theoretical Perspectives." Ph.D. dissertation. Princeton University, Princeton, N.J.
- Pack, Howard. 1994. "Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings." *Journal of Economic Perspectives* 8 (1): 55–72.
- Paldam, Martin. 1991. "How Robust Is the Vote Function? A Study of Seventeen Nations over Four Decades." In Helmut Northrup, Michael Lewis-Beck, and Jean-Dominique Lafay, eds., *Economics and Politics: The Calculus of Support*. Ann Arbor: University of Michigan Press.
- Peters, Thomas, and Robert Waterman. 1982. *In Search of Excellence*. New York: Harper Collins.
- Poterba, James, and Lawrence Summers. 1988. "Mean Reversion in Stock Returns: Evidence and Implications." *Journal of Financial Economics* 22: 27–60.
- Prescott, Edward. 1986. "Theory Ahead of Business Cycle Measurement." *Federal Reserve Bank of Minneapolis Quarterly Review* 10 (4): 9–22.
- Rey, Patrick, and Joseph E. Stiglitz. 1993. "Short-term Contracts as a Monitoring Device." National Bureau of Economic Research Working Paper 4514. Cambridge, Mass.
- Robert Wood Johnson Foundation. 1997. "Health Care Puzzle: Why Health Care Patterns Change With the Territory." *Perspectives*.
- Rodrik, Dani. 1996. "Understanding Economic Policy Reform." *Journal of Economic Literature* 34 (1): 9–41.
- . Forthcoming. "Who Needs Capital-Account Convertibility?" *Princeton Essay in International Finance*. Princeton, N.J.: Princeton University Press.
- Roll, Richard. 1979. "Violations of Purchasing Power Parity and Their Implications for Efficient International Commodity Markets." In M. Sarnat and G. Szego, eds., *International Finance and Trade*. Vol. 1. Cambridge, Mass.: Ballinger.
- Romer, Christina, and David Romer. 1996. "Federal Reserve Private Information and the Behavior of Interest Rates." National Bureau of Economic Research Working Paper 5692. Cambridge, Mass.
- Sachs, Jeffrey, and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–118.

- Salop, Steven. 1976. "Information and Monopolistic Competition." *American Economic Review* 66: 240–45.
- Salop, S., and Joseph E. Stiglitz. 1977. "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersions." *Review of Economic Studies* 44: 493–510.
- Sappington, David, and Joseph E. Stiglitz. 1987. "Privatization, Information and Incentives." *Journal of Policy Analysis and Management* 6 (4): 567–82.
- Shiller, Robert. 1989. *Market Volatility*. Cambridge, Mass.: MIT Press.
- Shleifer, Andrei, and Robert Vishny. 1989. "Management Entrenchment: The Case of Manager-Specific Investments." *Journal of Financial Economics* 25 (1): 123–39.
- Stigler, George. 1981. "The Theory of Regulation." *Bell Journal of Economics* (spring): 3–21.
- Stiglitz, Joseph E. 1985. "Information and Economic Analysis: A Perspective." *Economic Journal* 95: 21–41.
- . 1988. "Why Financial Structure Matters." *Journal of Economic Perspectives* 2: 121–26.
- . 1989. "The Economic Role of the State: Efficiency and Effectiveness." In A. Heertje, ed., *The Economic Role of the State*. London: Basil Blackwell and Bank Insinger de Beaufort NV.
- . 1992. "Capital Markets and Economic Fluctuations in Capitalist Economies." *European Economic Review* 36: 269–306.
- . 1994a. "The Role of the State in Financial Markets." In *Proceedings of the World Bank Conference on Development Economics 1993*. Washington, D.C.: World Bank.
- . 1994b. *Whither Socialism?* Cambridge, Mass.: MIT Press.
- . 1996. "Some Lessons of the East Asian Miracle." *World Bank Research Observer* 11 (2): 151–77.
- . 1997. "Looking Out for the National Interest: The Principles of the Council of Economic Advisers." *American Economic Review* (May): 109–113.
- . 1998a. "The Private Uses of Public Interests: Incentives and Institutions." *Journal of Economic Perspectives* 12 (2): 3–22.
- . 1998b. "Sound Finance and Sustainable Development in Asia." Speech to the Asian Development Forum, Chicago, March 12. Available at <http://www.worldbank.org/html/extdr/extme/jssp031298.htm>.
- . 1998c. "Central Banking in a Democratic Society." *De Economist* 146 (2):199–226.
- . 1999. "Public Policy for a Knowledge Economy." Lecture to U.K. Department for Trade and Industry and Center for Economic Policy Research. London. January 27.
- Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (3).
- Summers, Lawrence. 1998. "Deputy Secretary Summers's Remarks Before the International Monetary Fund." March 19. Available at <http://www.treas.gov/press/releases/pr2286.htm>.
- Tyron, Ralph. 1979. "Testing for Rational Expectations in Foreign Exchange Markets." International Finance Discussion Paper 139. Federal Reserve Board, Washington, D.C.
- U.S. Council of Economic Advisers. 1997. *Economic Report of the President 1997*. Washington, D.C.: Government Printing Office.
- Vickers, John. 1986. "Signaling in a Model of Monetary Policy with Incomplete Information." *Oxford Economic Papers* 38: 443–55.
- Vickers, John, and George Yarrow. 1988. *Privatization: An Economic Analysis*. Cambridge, Mass.: MIT Press.
- Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.

- Watanabe, Tsutomu. 1994. "The Signaling Effect of Foreign Exchange Intervention: The Case of Japan." In Reuven Glick and Michael Hutchison, eds., *Exchange Rate Policy and Interdependence*. Cambridge: Cambridge University Press.
- Weiss, Andrew, and Georgiy Nikitin. 1998. "Performance of Czech Companies by Ownership Structure." Boston University Department of Economics Paper.
- Wennberg, Jack, J. L. Freeman, and W. J. Culp. 1987. "Are Hospital Services Rationed in New Haven or Over-Utilized in Boston?" *Lancet* (1): 1185–88.
- Williamson, John. 1990. "What Washington Means by Policy Reform." In John Williamson, ed., *Latin American Adjustment: How Much Has Happened?* Washington, D.C.: Institute for International Economics.
- Wolfensohn, James. 1996. *Annual Meetings Address*. Washington: World Bank. Available at <http://www.worldbank.org/html/extdr/extme/jdwams96.htm>.
- World Bank. 1993. *The East Asian Miracle*. New York: Oxford University Press.
- . 1997a. *Proceedings of the Annual Bank Conference on Development Economics 1997*. Washington, D.C.
- . 1997b. *China 2020*. Washington, D.C.
- . 1998a. *Rethinking Aid*. Washington, D.C.
- . 1998b. *World Development Indicators 1998*. CD ROM. Washington, D.C.
- . 1998c. *World Development Report 1998: Knowledge and Information for Development*. New York: Oxford University Press.
- . 1999. *Assessing Aid*. New York: Oxford University Press.

WELCOMING ADDRESS

Rethinking Development—Principles, Approaches, and Projects

James D. Wolfensohn

In its 10 years the Annual World Bank Conference on Development Economics (ABCDE) has made extremely important contributions to the work of the World Bank Group. Looking back at the topics that have been discussed over the years, I was particularly struck by Ravi Kanbur's presentation in 1990 on projects and policy reform (Kanbur 1991). One of my greatest concerns in running this institution is how we can develop projects that secure the results we desire. Answering this question may require a lot of new thinking.

Toward a New Consensus

Recently at the Summit of the Americas in Santiago, Chile, heads of government met to discuss where Latin America is headed. What was fascinating to me, in listening to the day and a half of meetings, was that there was very little talk about macroeconomic theory in the traditional sense. At the outset of the meeting there was acceptance of the fact that sound monetary and fiscal policy are important for an open market, as are many of the tools and visions of economics described in the so-called Washington Consensus.

But by the end of the summit there was a general agreement that the Washington Consensus is dated, and that we now need to adopt a new, Santiago Consensus. This new consensus acknowledges that it is crucial to have economic growth and to adhere to tried and true monetary and fiscal policies. But as we go forward, the larger concerns are for equity and social justice. That is, how can we ensure that market-led economic growth benefits all members of society? And how can we deal with poverty in a framework that promotes environmental sustainability and popular participation and that generates significant results? The Santiago Consensus contains few of the items that dominated the ABCDE agenda 10 years ago. The consensus and the agenda have moved on.

James D. Wolfensohn is president of the World Bank Group.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

Education is the first element of the Santiago Consensus, because it is the key to unlocking equal opportunity. Next is health care. Then comes infrastructure—especially rural roads, because without them people cannot get to school. Then a justice system, because you cannot have equity without a legal system that works. The economic distortions created by crime and drugs must be eliminated. And civil safety must be ensured. Other important elements of the new consensus are well-functioning financial systems, urban and rural strategies, and power, sanitation, and communications services. This list is not exhaustive, but it gives a sense of a much more balanced and integrated approach to development than the Washington Consensus gives. But of course, good macroeconomic policy and open trade must be part of these efforts.

Partners in Progress

Even if agreement is reached on the need for these mechanisms, who should implement them? Traditionally the government has been the key player. But governments are no longer just the administration or the presidency. They increasingly include parliaments—as in the United States, where the administration’s program is subject to congressional approval. So when we talk of government, we are not thinking of a single leader, we are thinking of a complex and changing set of relationships within the framework of government. The simple notion that “the government will do something” is no longer a simple notion.

Civil society is another key player. Ten years ago one in four countries was democratic. Now the number is two in three. Ten years ago there were 1 billion people in the market system. Today there are 5 billion. These changes have created opportunities for civil society to take on responsibilities it never had before.

Then there is the private sector. Since the early 1990s there has been a dramatic shift in the flow of funds from official institutions and private investors to developing countries—to the point where private flows are now six times official flows. But the challenge remains of how to integrate the private sector with developing economies so that investors are not only looking at profits, but are also a part of the social sector.

And finally there is us—multilateral and bilateral institutions. How do we make effective policies when we are all treading on each other’s toes, giving competing advice, and taking up our clients’ valuable time? These were the issues raised in Santiago. It was a completely different debate than one we would have had 10 years ago.

Adopting New Approaches

I am pleased to see that the discussions and debates at this ABCDE are extending far beyond those of earlier years. Geography, political economy, and ethnic conflict are being addressed because they are central to the results we want to achieve. Good economic policy alone is no longer enough.

We at the World Bank are also trying to change. We are taking new approaches: working with our clients in a more responsive and less dogmatic fashion, seeking ways to differentiate among economies, building a knowledge bank so that we can leverage knowledge as a key driver of development.

These changes have led us to another issue that requires reconsideration. In the past Bank operations focused on getting projects approved. We are working to transform that approval culture into an implementation culture, where we judge ourselves by our results. The amount of money we lend and the number of projects we support may be useful indicators in some ways. But they are meaningless unless projects are developmentally effective.

But even successful projects may not be enough. The fact is, we are putting a lot of money into developing countries, yet poverty and inequity are increasing. And this raises the question: How can we go beyond projects?

We at the Bank must develop ways to deliver not just projects but also strategies that apply over longer periods and continue through political cycles. We need to think in terms of intersectoral strategies. You cannot deal with education effectively if you do not also consider health, and you cannot think about education and health unless you also consider rural roads. We also need to think in terms of partnerships: How can we bring together the four partners—government, civil society, the private sector, and international organizations—and agree on long-term national strategies that put in place all the components of development?

I wish I were smart enough to figure out how to do all this. As participants at the ABCDE, you are undoubtedly all much smarter. If you can help me with these issues over the next year and address some of them at next year's conference, I will be very happy indeed.

Reference

Kanbur, Ravi. 1991. "Projects versus Policy Reform." In Stanley Fischer, Dennis de Tray, and Shekhar Shah, eds., *Proceedings of the Annual World Bank Conference on Development Economics 1990*. Washington, D.C.: World Bank.

Financial Globalization: Can National Currencies Survive?

James Tobin

The largest private bank in a small country fails. Frightened depositors and creditors desert this country, its banks, and its currency, and its central bank's plea for foreign assistance garners little response. Affected creditors in neighboring countries, banks and central banks alike, scramble for internationally liquid assets. Interest rates zoom up everywhere, loans are called or not renewed, economic activity sinks, and unemployment quickly rises to politically hazardous rates. The managers of the world monetary system, central bankers individually and collectively, strive above all to maintain the credibility of the system and confidence in existing currency rates. But the effects of their deflationary policies on business conditions instead destroy confidence. In the end country after country has to abandon its commitments to redeem its currency at the promised price. In country after country, then and only then does economic recovery begin, and it takes many years.

The place is not East Asia in 1997–98 but Europe and North America in 1931. The bank was the Credit Anstalt in Vienna. The monetary system was the gold standard, as revived after a hiatus due to World War I. Central bankers, finance ministers, prime ministers, and presidents put defense of the gold values of their currencies above all else. Weimar Germany maintained the gold content of the mark but rationed its gold reserves. Its deflationary policies in 1931–32—high interest rates, tax increases, no relief or work for the jobless—paved the way for Adolf Hitler's accession to power in January 1933. In September 1931 Britain was finally forced off gold, after suffering from hard times ever since 1925, when Chancellor Winston Churchill overvalued sterling by returning to gold at the 1914 gold and dollar value of the pound. In the United States, as the recession of 1929–30 became the Great Depression and the banking system collapsed, the Federal Reserve and President Herbert Hoover stubbornly defended the gold value of the dollar. (Hoover actually had some commonsense Keynesian instincts for fiscal and monetary activism until his Treasury warned him that the dollar's gold standard was threatened.) President Franklin Roosevelt devalued the currency in 1933, and recovery began.

James Tobin is Sterling Professor Emeritus of Economics at Yale University.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

I concur with the views of younger scholars that policymakers' *auri sacra fames* (Keynes's and Virgil's term, accursed lust for gold) was responsible for turning a recession into the Great Depression.¹ The international monetary system—the interconnection of national currencies with one another—was then and is now a crucial factor in global economic stability and prosperity. The presumption that currency crises are the fault of the victims is still all too prevalent among the statesmen of world finance and among media pundits. The common view is that good policies and proper institutions will enable a prudent government to keep its currency convertible at an announced parity in gold or in other currencies. The vulnerability of any economy, especially a small country caught in the commodity and financial markets of a big world, is insufficiently appreciated. So is the intrinsic fragility of a fixed exchange rate.

The End of the Bretton Woods Gold-Dollar Standard

It has been a quarter century since the United States ditched the Bretton Woods system and adjustable pegs were abandoned in favor of floating exchange rates (not always clean floating, to be sure) among the major currencies—dollar, yen, and deutsche mark (to which other major Western European currencies have been tied most of the time). Floating among the big three currencies presumably will continue as the euro succeeds the deutsche mark.

A fixed-rate system had failed again. One reason for the failure was that the United States could not devalue the rate of exchange of the key currency (the dollar) against other currencies without the concurrence of the other governments. The United States wanted Germany and Japan, in particular, to appreciate their currencies (by lowering the price of gold in their currencies), and Germany and Japan thought that adjustment was the responsibility of the United States. A second reason for the failure, which complicated the currency rate conflict, was the peculiar role of gold in the system. Dollars held by foreign governments were convertible into gold at a fixed price. Private dollars were not supposed to be convertible, but they became so *de facto* because until 1968 the United States and the United Kingdom fed the private gold market to keep the free-market gold price from getting out of line. U.S. balance of payments deficits increased dollar debt in official hands and depleted the U.S. gold reserves available to redeem those dollars. In the ultimate impasse in 1971–73 the United States abandoned its commitment to pay gold for dollars. Unlike Hoover, President Richard Nixon was not willing to sacrifice U.S. prosperity for the gold standard or for fixed currency exchange rates.

I think he was right. I know that among many wise and experienced observers there is nostalgic longing for a return to fixed rates and talk of a new Bretton Woods. The grass is always greener on the other side of the fence! Floating nominal rates are blamed for excessive variability in real exchange rates. The unanticipated appreciation of the dollar against the yen in the early 1980s looms large in the memories of U.S. businesses affected. I suspect, however, that the shocks that have moved exchange rates significantly since 1973 would have brought irresistible pressures on fixed rates, resulting in reserve crises like those of the 1960s and early 1970s.

Instead of being blamed for the volatility of nominal and real exchange rates, floating rates should perhaps be credited for accomplishing economically desirable revaluations without currency crises. A recent example is the 40 percent decline of the yen against the dollar over two years—a problem never serious enough to be reported on the front pages of U.S. newspapers.

At the opposite extreme, another way to escape currency crises is to adopt permanently and exclusively a common international currency, as is about to occur within the European Union. This approach, of course, has its own problems. Perhaps a worldwide common currency will be adopted sometime in the next century, but not soon.

A Voice from the Past

Given my lack of experience and expertise in the World Bank's world, I am unqualified to speak at this conference. But my old friend and onetime colleague Joseph Stiglitz insisted that I do so anyway. To show that I have at least thought about international monetary problems before and to show that today's issues are not altogether new, consider something I wrote in 1972. This passage occurs in a short book on domestic macroeconomic policy called *The New Economics One Decade Older*:

The most important barrier to flexible monetary policy is the ever-increasing international mobility of liquid capital. The Eurodollar market is unifying the short-term money markets of the major countries on both sides of the Atlantic. European countries have felt keenly, and complained bitterly, that they have lost autonomy in monetary policy. Even the autonomy of the United States Federal Reserve has been diminished. ... The interest sensitivity of short-term funds can be expected to continue to increase and to pose even greater problems for the international monetary system and for national monetary policies. ... [A]s substitution elasticities increase ... the boundless resources of private arbitrageurs will just erase any rate differentials the national monetary authorities try to create and sustain.

There is no more important item on the agenda of the coming negotiations for international monetary reform. On the one hand, some agreed central coordination of national monetary policies is essential. Otherwise the common international interest rate level, from which feasible national deviations are limited, will be left to anarchy and tug-of-war. On the other hand, there is nowhere near enough economic and political unity among Europe, North America, and Japan to support a single international monetary policy for the whole group. The new international arrangements must protect some national autonomy in monetary policy.

Unless the ... world acquiesces permanently in [a] fixed-exchange-rate dollar standard ... we cannot count on a system in which the Federal Reserve makes world monetary policy. Moreover, the Common Market countries will undoubtedly seek greater monetary coordination among

themselves, so that Europe will have more muscle in contest with the Federal Reserve.

[It] is clearly desirable to preserve some possibilities of autonomy in national or continental monetary policies and to defend them against the growing internationalization of money markets. Our economies and governments are not sufficiently unified in other respects—goods, labor, and capital markets, taxes and fiscal policies—to live with a single ... monetary policy. That is where the analogy with the centralization of Federal Reserve policy [in the United States] breaks down. The same forces that unified short-term securities markets throughout the U.S. also produced ... national markets in goods, labor, and capital ... [that] can handle regional differences in ... circumstances in a way that is not possible in today's international economy. And a national government can carry out compensatory fiscal redistributions between regions; there is no comparable international mechanism in prospect.

How can some international monetary autonomy be preserved? Some sand has to be thrown into the well-greased channels of the Eurodollar market. (pp. 84–88)

I went on to advocate increasing exchange risk by making exchange rates more flexible, either through outright floating or widened bands around frequently adjusted parities. Then comes for the first time a proposal for “an internationally agreed uniform tax, say 1 percent, on all spot conversions of one currency into another,” in order to “drive a wedge between short-term interest rates in different national markets”² (p. 89).

My propositions today are similar but applied to a wider universe. Let me summarize them. First, for most countries, fixed exchange rates in their usual form, adjustable pegs, are a bad idea. Developing countries would be well advised to follow the example of the major capitalist countries and let their currencies float like the dollar, yen, and deutsche mark. It is hard to understand why this did not become normal practice long ago. It would have avoided the worst consequences of recent adjustments in exchange rates. Is the reason that it would relax the discipline for “sound” policies exerted by fixed rates?

Second, while globalization of financial markets—the liberalization and deregulation of international financial transactions—has made important contributions to the economic progress of developing and emerging economies and can continue to do so, these trends also threaten the monetary sovereignty of those countries. This is especially the case for a country committed to a fixed exchange rate, an adjustable peg that it has promised not to adjust. The logic of financial globalization is to increase the elasticities of substitution between risk-adjusted rates of return on local assets and debts and those in dollar markets until the local central bank has no margin within which it is free to determine domestic interest rates.

Third, once the central bank cannot make monetary policy, a logical next step is “dollarization”—or “yenization” or “euroization”—allowing one of the hard cur-

rencies to become the smaller country's means of payment and unit of account. This approach does have some advantages. The trouble is that the big central bank has no reason to consider a satellite's conditions and interests. The same problems will arise in the European monetary union, but in that case the central bank is responsible to all the members of the union.

Fourth, to preserve a local currency with residual monetary sovereignty, some friction in international financial institutions and markets needs to be retained. This is also true in a regime of floating rates, though floating itself is some protection. The new global financial system should be able to contribute to development without rendering central banks impotent or whole currencies obsolete. The governments, banks, and businesses of developing countries should eschew short-term demand debt or short-term debt in hard currencies. Flows of capital to developing countries should preferably take the form of direct fixed investment or equity.

Fifth, developing countries need to build institutions of financial reform and regulation supportive of modern national financial systems and of independent currencies. The International Monetary Fund (IMF) should concentrate on its intended function as a principal source of liquidity for its members. The resources of the IMF and of its members are pitifully small and should be augmented.

Adjustable Pegs and Bank Runs

The trouble with adjustable pegs is that they can be adjusted and therefore invite speculation that they will be. They are no less a potential invitation for speculation than are floating rates. Indeed, a discrete change in an official parity is much more traumatic. It is a loss of face and a blow to pride. It is an administrative decision, that is to say a decision of policy and politics. It necessarily requires responsible officials—finance ministers, chancellors, central bank chairpersons—to go back on their solemn word. Moreover, they or their successors have the unenviable task of choosing a new rate in a climate poisoned by distrust, clouded by uncertainties about the fundamentals, and dominated by unpredictable psychology. It is easy to get it wrong, thereby prolonging and aggravating the crisis. For all these reasons, there is great temptation to stick with an overvalued parity too long.

A central bank managing and defending a currency pegged to an external hard currency, or to a basket of hard currencies, is like a conventional commercial bank. The bank's deposit liabilities are fixed in nominal value and payable in cash on demand or quite soon. They are "backed" largely by illiquid or imperfectly liquid assets of uncertain ultimate value. The bank's cash reserves—currency and coin and deposits in the central bank—are only a fraction of its cash liabilities. The bank is nevertheless solvent if patient, informed valuations of these assets equal or exceed the liabilities with high probability. *Patient* means that the valuations allow time for the assets to yield their expected values. Premature liquidations by use or sale are costly or impossible, the more so the greater the proportion of its assets the bank must liquidate.

The discrepancy is the basis for the distinction between insolvency and illiquidity, for the belief that “marking to market” may understate the eventual value of the bank, and for the function assigned to a lender of last resort—namely, to allow the bank time to attract deposits or liquidate assets. The distinction is not absolute. The bank’s world is changed by a bout of illiquidity, and by the events that bring it about, in ways that increase the likelihood of insolvency. Help from a lender of last resort may rescue a bank on the brink of insolvency. On the other hand, the lender may find itself keeping alive a crippled bank that will never survive on its own.

Clearly the expectations and risk assessments of depositors and participants in asset markets are crucial. In a benign equilibrium of the bank these estimates are rational and generate patterns of behavior and valuations that keep a basically sound bank liquid. The “last resort” does not arise, and its lender stays on the sidelines, while its existence supports the equilibrium. But the benign equilibrium is fragile. Every depositor is continuously deciding whether to withdraw cash or not, and her decision depends mainly on what she thinks others are deciding. Adverse events or rumors may tip the scale to runs and panics. Contagion from failing banks nearby can doom intrinsically solvent banks.

Why should a depositor keep funds in a bank if she gains nothing and may lose everything? Bank services are one reason, and nowadays some interest is generally credited. But government deposit insurance, explicit or implicit, is usually essential to keep depositors content. To offset the moral hazard incentive of banks and depositors to seek risky gains while the insurer absorbs the losses, it is necessary to regulate and oversee the balance sheets of insured banks—as the debacle that followed deregulation of the U.S. savings and loan industry in the 1980s dramatically confirmed.

Besides the benign equilibrium, supported by deposit insurance and balance sheet surveillance or not, there is a second, malign equilibrium in which the bank has failed and closed. Depositors have withdrawn all the bank’s reserves in cash. Other assets have been sold at losses in a desperate quest for liquidity. Or better, a regulator has closed the bank promptly enough to conserve some assets for eventual settlements with depositors, insurers, and other creditors.

The analogy of a national currency to a bank is clear. The central bank has promised to buy back its own currency with external currency at an announced price, and for that purpose holds reserves of hard currency. In the benign equilibrium of this case, expectations in currency markets around the world support behaviors that validate the expectations and sustain the pegged exchange rate. In the second, pessimistic equilibrium the central bank defaults on its commitment. Like bank depositors worried about what other depositors will do, holders of a pegged currency fear that they will act too late to save their assets. Potential claims on central bank reserves include not only the external liabilities of the central bank and the government but also those of private banks, businesses, and households (domestic and foreign). All the liquid local currency assets they hold can potentially be tendered to the central bank and government to buy up their holdings of hard currency. If those assets are then spent or exchanged for nonliquid assets, they can again fall into hands that will convert them into hard currency.

The benign equilibrium is fragile because estimates of its viability at home and abroad are interdependent, and panicky rushes to convert local into foreign currency can force the central bank to abandon its commitment and let the currency fall—even though economic fundamentals indicate that the currency is worth, if not its original pegged value, much more than its crisis price.

The analogy between the bank-deposit-and-local-currency choice and the domestic-currency-and-foreign-currency choice is imperfect in one respect. Whereas an uninsured bank deposit will not appreciate above its contracted cash value, the local currency's exchange value can rise as well as fall. In practice, however, the exchange rates of developing country currencies often gravitate to the high ends of their bands or of traders' confidence intervals. When currency speculators see only downside risk, they sell—just as depositors run for cash if they see no chance that their bank's condition will improve.

Advocates of fixed rates regard the benign equilibrium as normal and sustainable and the runs as anomalous and avoidable. It is just a matter of adopting and maintaining policies that engender confidence. In the design of so-called bailout packages to reverse attacks on the currency's exchange rate in times of incipient crisis, the first priority is to promise measures that the "market" will regard as sound. But these measures must overcome the adverse momentum of *sauve qui peut* (get out fast if you can) panic.

Moreover, a suspect currency typically is thought to be overvalued. Perhaps the inflation rate has exceeded the rates of competitors and trading partners. Perhaps export markets are slumping. Perhaps the current account is turning into deficit, reflecting borrowing for domestic consumption rather than productive investment. There may be no credible package of measures that can save the exchange rate or reverse its decline.

An extraordinarily high interest rate is the usual emergency therapy. The idea is to devalue the currency against its own future value, inducing people to hold onto it despite the expected decline in its foreign exchange value. If this works, it raises both the current and the expected market exchange rate. But tomorrow, if the adverse expectation is fulfilled, the rate is again lower. Only when (and if) favorable changes in markets occur can the currency be stabilized without keeping the interest rate extraordinarily high. Meanwhile, the economic damage of that interest rate may be dragging down the exchange rate's prospects. High interest on public debt increases the budget deficit, undermining the recommended therapy of fiscal austerity.

Currency Board, Money Board?

A currency board requires 100 percent reserves in hard currency against the local currency monetary base. If reserve assets can be bought from the central bank with bank checks, this requirement is no guarantee that the central bank will not run out of reserves. If local currency is a ration coupon for external reserves, bank depositors could obtain these tickets by withdrawing cash from their deposits. Distrust of the currency would then be accompanied by a bank panic. After all, the purpose of

a currency board is to freeze the central bank into a permanent commitment to an exchange rate fixed at a particular value, if necessary forcing draconian tactics (most likely astronomical interest rates) to avert defaults on its currency exchange commitment.

Combining a currency board with fractional reserve banking is awkward. If the size of the monetary base is limited to the central bank's holdings of international reserves, the central bank has no way to compensate for increases in the public's demand for local currency at the expense of its willingness to hold local bank deposits. Increased demand could come about for various reasons: some random and innocuous, such as increases in income and consumption spending, some reflecting public concern over the soundness of banks. However caused, the result of shifting a dollar of high-powered money from bank reserves to publicly circulating currency is to substitute one dollar of low-powered money in currency form for, say, five dollars of low-powered money in bank deposit form (assuming a reserve ratio of one-fifth).

Those of us with long memories recall the bank runs in the United States in the early 1930s, triggered by bank failures and, in turn, the cause of further failures. The unwillingness or inability of the Federal Reserve to respond with open market purchases to expand the monetary base was disastrous, dooming both the economy and the banking system. At the time Federal Reserve monetary base liabilities were not constrained by a currency-board-type 100 percent gold reserve requirement, although they were supposed to be backed by some combination of gold, Treasury bonds, and commercial paper eligible for rediscount. The point here is that a currency board makes it impossible for the central bank to perform its normal domestic functions—either that of macroeconomic stabilizer or of lender of last resort.

A 100 percent reserve requirement on bank deposits is a logical extension of the currency board idea to a money board system. It would allow the system, originally used in British colonies dependent mainly on paper money, to catch up with the rise of bank deposits as the main medium of exchange.

A 100 percent reserve banking system would tighten the country's commitment to its exchange parity, but at heavy cost. It would deprive the economy of the intermediary functions performed by fractional reserve banks. Presumably some non-bank intermediaries would take their place. They would seek liabilities to the public as close to bank deposits as the authorities permit. These would require some regulation, although the availability of fully backed deposits in narrow banks would relieve the government of moral compulsion to guarantee the liabilities of other intermediaries.

A currency board, or a more comprehensive money board, sacrifices real macroeconomic performance in all its significant dimensions—employment, production, income, growth, trade, saving, and investment—to the strength of the currency and indirectly to the prevention of inflation. When the successes of the device are touted, it is in these narrow terms. The currency board is an extreme form of the fixed exchange rate as a “real anchor,” a tactic of national self-discipline popular in recent years. However, the true test of successful policy is not conquering inflation

with an open-ended sacrifice of prosperity but conquering inflation while achieving full employment and reasonable growth in economic well-being. Argentina stabilized prices by tying its currency to the dollar, but its unemployment rate is stuck in the double digits.

In any case, it is by no means certain that a currency board or any similar fixed exchange rate commitment will work. Once again, there is a bad outcome as well as a good equilibrium. If the initial stock of external reserves is small, the cost in economic activity of cutting the stock of local currency may be devastating, and it may set off a scramble for hard currency. Those who get their hands on local currency will buy up the central bank's hard currency and force further deflation on banks and the economy. These unstable dynamics will force the country to cut itself loose from the currency board.

A successful commitment to a fixed exchange rate requires an ample initial stock of unborrowed reserves, as well as policies that reinforce the virtuous circle of the good equilibrium, as in Hong Kong (China) and Taiwan (China). In less auspicious circumstances the real anchor strategy has contributed to overvaluation of the currency. This can happen when the small country's inflation exceeds expectations or when the anchor country's inflation declines.

Why Not the Dollar?

The currency board arrangement is a way, albeit somewhat technically flawed, of surrendering independent monetary policy and acknowledging subordination. At that point, why not go all the way, drop the local money, and adopt the hard currency as a medium of exchange and unit of account? Argentina is well on its way to doing this. Federal Reserve policymakers, however, are not going to weigh macroeconomic problems in Patagonia even as much as those in Idaho. But this is the destination to which financial globalization is taking developing countries, whether the IMF, the U.S. Treasury, and the other lords of international finance acknowledge it or not.

In a dollarized regime some of the functions central banks now play might be taken over by private banks. They would accept deposits in dollars, as many do already. The acceptance in New York of checks in dollars drawn on Indonesian or Korean banks might be subject to a discount reflecting the reputation of the bank and its assets, like the discounts on bank notes issued by wildcat banks in the American West in the 19th century. Those discounts could take the place of an exchange rate. Of course, in an integrated global financial system U.S.-chartered banks would be competing in Indonesia and the Republic of Korea for deposits and loans.

The informal use everywhere of dollar bills for hand-to-hand currency and "under the mattress" hoarding is one thing. Adoption of the dollar as legal tender, in place of or in addition to a country's national currency, is quite another. This move would best be negotiated between governments, so that the handling of checks between banks in the smaller country and the United States could be systematized.

Similar problems have been painstakingly resolved in the European Union. Effective internationalization is not unmitigated *laissez faire*.

Dollarization deprives the government of the small country not only of monetary sovereignty but also of seigniorage. Given the costs of borrowing in dollars, inclusive of country risk premium, this could be a substantial fiscal loss.

In Praise of Dirty Floating

Surely the most important lesson of currency crises is the most obvious. Countries should not peg their exchange rates. They should let them float. They should not even confine rates to a broad band, with or without a moving central parity. If an exchange rate hits the bottom of the band, it is pegged and invites speculative attack. Just let it float.

I am not a purist. I think dirty floating is alright. Interventions are sometimes called for and need not be transparent. For occasional dirty floating—defensive or offensive—hard currency reserves are needed. The central bank must husband them in the national interest, whether the exchange rate is fixed or floating. In either case the government needs to limit private external financial transactions that may force the country to lose reserves or to suffer unwelcome currency depreciation.

Globalization and Financial Reform

Developing economies, especially the East Asian tigers, have made great strides in liberalizing and globalizing their financial systems, markets, and institutions. Local non-financial businesses borrow, lend, and sell shares in major international markets. The balance sheets of banks and other financial institutions contain assets and liabilities in various currencies. Gross volumes of currency transactions involving these economies have multiplied, and net flows of private capital into these economies have greatly increased. No doubt these developments reflect liberalizations that have opened these economies to foreign investment and made them increasingly attractive.

At the same time, some aspects of financial globalization are perilous to the health of central banks and economies, as recent currency crises show. This is especially clear in fixed exchange rate or adjustable peg regimes. When private banks and businesses can borrow in whatever amounts, maturities, and currencies they choose, they create future claims on their country's reserves. This may force on the central bank and the government monetary and fiscal policies that sacrifice the country's prosperity and growth in order to protect the reserves on which these debts are potential claims. They might indeed threaten to exhaust the nation's reserves.

Integration and perfection of financial markets will bring money market interest rates in different financial centers closer and closer together. In 1997 U.S., Japanese, and European banks saw loans to Korean banks as great opportunities because the interest rates were higher than those they could earn at home. At the same time, Korean banks seized the chance to borrow at what they regarded as low rates. Arbitrage was chipping away at the risk premium implicit in the rate differentials. The

longer the peg of the Korean won survived, the closer the Korean short rate (whether on won or dollar liabilities) would come to New York or Tokyo rates. As net demands across markets become more elastic with respect to interest rate differentials, the less autonomy the central bank of the smaller country will have over its interest rates and monetary policies. The smaller country loses monetary sovereignty and becomes in effect a monetary province of the large country to whose currency its own is pegged.

Short-term private bank debts in hard currencies were fatal to the Indonesian and Korean currencies in the last months of 1997. These debts, though they seemed to the foreign lenders and domestic borrowers directly involved to be straightforward business deals, visited that were severe negative externalities on their fellow citizens, bringing about currency crises devastating to entire economies.

The central bank, committed to honor the peg and to maintain the country's terms of trade, has to protect its reserves. It cannot be indifferent to the claims on those reserves negotiated by private parties, domestic and foreign, who ignore the social risks. An obvious precaution is to limit even to zero the net indebtedness (particularly the short-term debt) in hard currency permitted any private bank. The device used in Chile and Colombia, an extra reserve requirement, is evidently successful. It is more important to slow down incoming funds than outgoing money and to install such hurdles permanently rather than just in emergencies.

These grains of sand in the wheels are, to be sure, departures from the goal of complete integration, with universally free asset markets blind to currency denominations, geographic locations, political jurisdictions, and nationalities of transactors. But it is hard to see how governments outside the major industrial capitalist democracies can maintain monetary sovereignty without some regulations to protect their international reserves.

It is worth considering why China is evidently immune to the "Asian flu." It is not because China has a currency board, or, like Taiwan (China) and Hong Kong (China), immense nest eggs of hard currency reserves. Rather, it is because China restricts the convertibility of its currency. Only foreigners who have earned renminbi in commercial transactions are guaranteed the right to convert them into foreign currencies. Free capital account convertibility, the essence of financial globalization, does not exist. As a result, China has ample monetary sovereignty. Restricted convertibility has not deprived China of massive infusions of foreign capital and technology.

In the "bailout" packages for East Asian economies further cross-border financial liberalization was one of the conditions imposed by the IMF and the U.S. Treasury for official loans. This was a surprising requirement, given the evident facts that excessive private external short-term debt was, if not a cause of the crisis, a serious aggravation of it, and that banking and financial institutions seemed to need more regulations in several respects as well as fewer in other respects.

U.S. experience suggests the importance of distinguishing among several kinds of public regulations of financial institutions and markets. First are requirements designed to make markets work better, by outlawing fraud and self-dealing and by requiring depositories, investment banks, and sellers of financial instruments to inform the public clearly and completely exactly what it is they are selling. Second

are limitations on balance sheets of intermediaries, in cases where the public cannot be sufficiently protected by information alone or where the state has an implicit or explicit responsibility to compensate losers. Third are protections of competition against concentration and collusion in restraint of trade. Fourth are regulations like reserve requirements and capital ratios, which are essential to make government policies workable. Fifth are orderly legal procedures for handling bankruptcies and defaults. (A useful precedent for handling the rash of insolvencies now afflicting some East Asian economies is the U.S. Depression-era Reconstruction Finance Corporation. For example, this corporation put public money into defunct companies by investing in their preferred stock. Preferred stock was also offered to depositors in failed banks up to the amount of their lost deposits.)

On the other hand, U.S. history is also full of insalubrious regulations designed to protect vested interests against competition—prohibiting entry into particular markets, setting prices and interest rates, distorting market outcomes by taxes and subsidies. Presumably we want developing countries to follow our good examples and not our bad. Let us encourage them to build good national financial systems, not just to open their doors ever wider.

Some critics of the victims of the currency crises and of the bailouts extended to them assert that if governments and international agencies would just get out of the way, free markets would reach ideal solutions to all the problems. We economists should be cautious in applying “invisible hand” propositions—the theorems of optimality of competitive equilibrium that we love so much—to money and finance, especially international money and finance. Those theorems apply strictly to a single closed real economy, without money, presumably one where incredibly efficient multilateral barter determines relative prices and allocates resources over future times and states of nature.

Fiat money does not figure in production or utility functions, so why it has a particular value or any value at all is one of those puzzles that economic theorists pose for themselves. Even more mysterious are the relative values (exchange rates) of various fiat moneys, none of which has intrinsic value. Since these are creatures of governments, it is not surprising that some government regulations are necessary to make them work. I certainly am not saying that we can dispense with governments or money. Quite the opposite. We do not in fact have moneyless efficient multilateral barter. I am simply warning against relying on a priori ideological shortcuts instead of pragmatic architecture.

We Need Lenders of Last Resort

Moral hazard has become almost as fashionable an expression as *transparency*. Many pundits have discovered that moral hazard is intrinsic in “bailouts,” the prejudicial word for lender of last resort operations, and they are generally quite indignant about it. Among economists and financiers looking for better ways of handling future debt crises, minimization of moral hazard seems to be the primary goal. I think this is a misplaced priority. In liquidity crises that necessitate emergency loans,

lenders and borrowers may be spared losses of principal, but they hardly escape unscathed. They are not likely to find the experience one they would wish to repeat. More important, the social costs of unmitigated currency collapse extend far beyond the parties to financial deals, to ordinary people who lose jobs, savings, and income. It is worth putting up with some moral hazard in order to limit these third-party effects.

The IMF needs to take a lender of last resort responsibility more seriously. It was founded to tide members over during temporary liquidity crises, not to shape the permanent economic structures of economies and guide their long-run development strategies. For its fire-department function, the IMF needs more money, not less, and more than is currently under debate. Aggregate quotas of \$150 billion? That is peanuts! Within larger quotas, members should have bigger unconditional drawing rights.

A Plea for Humility

A final remark. One of the more unseemly by-products of the East Asian crisis is the triumphalism of U.S. commentaries on the events. The currency troubles have been interpreted as demonstrating the hollowness of the “Asian model” of capitalism. Not many years ago many Americans admired and feared the performance of Japan and the smaller East Asian economies. We wondered whether their model of capitalism, in particular of corporate governance and employment, was better than ours—and we hated having to consider that possibility. Some silly popular economics books exploited these worries. Then the Japanese slump of the 1990s and now the come-uppance of the miraculous East Asian tigers, in contrast to American and British prosperity, have given us new confidence in the “Anglo-American model.” Some spokespeople for our kind of capitalism have not resisted the temptations of triumphalism—even though the overzealous reach of our practitioners of global finance might bear some responsibility for the crisis.

Notes

1. Keynes's essay of that title was published in his *Essays in Persuasion* (New York: Norton Library, 1963, pp.181–85).
2. The 1 percent tax I suggested then was much too high. A practical tax would be one- or two-tenths of 1 percent.

Reference

- Tobin, James. 1972. *The New Economics One Decade Older*. Princeton, N.J.: Princeton University Press.

ABCDE: Past Ten Years, Next Ten Years

Stanley Fischer

It is a privilege to speak at this 10th Annual World Bank Conference on Development Economics (ABCDE), and I want to thank Joseph Stiglitz for inviting me to help celebrate the 9th anniversary and the 10th conference. I cannot think back to the start of these conferences without reflecting on my brief and happy years at the World Bank, and I hope you will excuse me if I start by straying from my subject to say a few words of thanks to the Bank and the people who serve in it.

Let me say first how grateful and proud I am to have served as chief economist of this remarkable institution and thus to be part of a chain that includes Hollis Chenery, Anne Krueger, Lawrence Summers, Michael Bruno, and Joseph Stiglitz, from all of whom we have all learned so much. I started learning from Joe more than 30 years ago, when I was a graduate student at the Massachusetts Institute of Technology and he was an assistant professor, back from his initial foray into the study of development in Nairobi. I would in particular like to pay tribute to my predecessor, Anne Krueger, whose then-controversial insistence on the centrality of trade liberalization in economic development has been amply borne out by subsequent research.

It was not only an education but also a pleasure to work with so many talented and devoted people in the World Bank—the dedicated leadership and staff in the Development Economics Vice Presidency and other friends and colleagues throughout the institution. Although any list is bound to be invidious, I would particularly like to thank some of my closest associates: my advisers, first Johannes Linn and then Andrew Steer; Dennis de Tray, research director and co-conspirator in setting up the ABCDE; the leaders of the *World Development Report* teams for 1989, 1990, and 1991—Millard Long, Lyn Squire, and Vinod Thomas; and Kate Oram. Although I would like to go on and on in this vein, I shall now turn to the topic of the ABCDE.

Stanley Fischer is first deputy managing director of the International Monetary Fund. The author is grateful to Boris Pleskovic and Vinod Thomas for helpful discussions.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

Goals of the ABCDE

According to the introduction to the first volume of conference proceedings, the ABCDE was created to improve member country and World Bank policymaking by enhancing the knowledge base (Fischer and de Tray 1990). The goals were to open up the Bank to outside ideas and problems, and if possible to help shape the research agendas of those outside the Bank who were also thinking about development.

When the conference series was formally evaluated in 1995, after seven conferences, the goals were described more precisely as being:

- To expose World Bank economists to fresh insights and recent developments in economics that are influencing views outside the Bank and may alter Bank policy advice.
- To draw attention to issues that are of crucial interest to a wide range of development practitioners.
- To induce leading researchers to explore and account for the real-world implications of their work, and to incorporate the Bank's practical knowledge of developing and transition economies in their analyses.
- To improve policymaking in the Bank and its member countries by enhancing our understanding of economic processes.

The title assigned to this address, "ABCDE: Past Ten Years, Next Ten Years," suggests that I should review the record of the past nine conferences and then look ahead. I will look back, but not to review the record. That was done by the 1995 evaluators, who pronounced themselves on the whole satisfied. Based on a selective reading of the conference volumes, I agree with their assessment.

Instead I will reflect on the development consensus when I left the Bank in 1990 and how its focus has changed. Looking forward, the assigned title "ABCDE: Next Ten Years" violates the fundamental rule of forecasting, which is to forecast an event or a date but not both. Rather than attempt to forecast the content of future ABCDEs, I will end by speculating on the implications of globalization for developing countries.

Development Issues at the Start of the 1990s

When the first ABCDE took place in 1989, four important developments stood out among developing countries:

- The debt crisis was on its way to resolution, and the lost Latin American decade was drawing to its end.
- The transition was beginning in Eastern Europe, though even as late as 1989 hardly anyone anticipated that the Soviet Union would soon disintegrate.
- The East Asian miracle was in full swing, with per capita growth over the previous 25 years averaging more than 6 percent, and per capita GDP in China having more than doubled during the 1980s.
- Average per capita GDP in Sub-Saharan Africa had declined during the 1980s, nearly offsetting the gains since independence. Indeed, in some African countries per capita incomes had been falling for 25 years.

In thinking about economic development, there was growing consensus about many of the policies needed to produce growth. At the time of its original presentation in 1989 John Williamson's Washington Consensus summarized views that were held in much of official Washington, including the World Bank (see Williamson 1990). But in the next few years the consensus broadened to include many researchers and policymakers in developing countries in Latin America and elsewhere.

The consensus within the World Bank, no doubt also representative of views within much of the development economics field, was best captured in *World Development Report 1991: The Challenge of Development*, which summarized its preferred approach as market-friendly. That report was preceded by *World Development Report 1990: Poverty*, which is an essential companion to its successor, and whose message—crucial to the Bank's mission—is no less relevant today than it was then. However, I will focus on *World Development Report 1991*, both because I want to examine the overall strategy for development espoused by the Bank at that time and because sustained reductions in poverty are best attained in a growing economy.

World Development Report 1991 argued that the primary responsibility for development rests with developing countries, which should emphasize:

- Investing in people
- Improving the climate for enterprise
- Opening economies to international trade and investment
- Getting macroeconomic policy right.¹

The report also argued for a reappraisal of the roles of the market and the state:

Put simply, governments need to do less in those areas where markets work, or can be made to work, reasonably well. In many countries it would help to privatize many of the state-owned enterprises. Governments need to let domestic and international competition flourish.... [They] need to do more in those areas where markets alone cannot be relied upon. Above all, this means investing in education, health, nutrition, family planning, and poverty alleviation; building social, physical, administrative, regulatory, and legal infrastructure of better quality; mobilizing the resources to finance public expenditures; and providing a stable macroeconomic foundation, without which little can be achieved. (p. 9)

How did the messages of the report relate to the four issues—the end of the debt crisis, the start of transition, the East Asian miracle, and negative growth in Africa—then dominating the development agenda? The message was clearly relevant for all these situations, particularly Latin America. But it needed further development and specificity to provide a practical guide for action for transition economies. And what precisely the role of the state had been in East Asia and how—if at all—that fit into the paradigm so eloquently defined by *World Development Report 1991* remained as major question marks.²

At about the same time that *World Development Report 1991* was published, a mainstream position began to emerge on the strategy for reform in transition

economies (see Lipton and Sachs 1990 and Fischer and Gelb 1991). Despite agreement on the reforms needed in different sectors, the speed and sequencing of reforms provided ample room for debate, especially over whether and in what areas of reform very rapid adjustment—shock treatment—might be preferable. This issue gained particular salience in light of the contrast between the negative growth of reforming Eastern European countries and the stellar performance of China, with its more gradualist strategy, particularly in privatization (see Sachs and Woo 1994 and World Bank 1996).

The World Bank's *The East Asian Miracle*, published in 1993, sought to answer the questions about the development strategies followed in the eight high-performing East Asian economies, a group that included Japan but not China. These economies had succeeded not only in achieving unprecedented growth rates but also in maintaining relatively equal distributions of income. In most—but not all—cases superb growth was accompanied by high rates of saving and investment.

The study concluded that these economies had succeeded first by getting the basics right, particularly in investing in human capital and in ensuring macroeconomic stability. Further, all the economies had kept price distortions within bounds, especially by limiting the bias against agriculture found in many developing countries. They had also encouraged the import and absorption of technology. But this was not the whole story, for in some of the eight economies government had intervened systematically and in many ways to foster development. The most important intervention came through various measures of export promotion. Also important were significant financial sector interventions, in some cases import protection efforts, subsidies to declining domestic industries, and investments in applied research. These interventions were more pervasive in Japan, the Republic of Korea, Singapore, and Taiwan (China) than in the other high-performing Asian economies. Indeed, one of the striking results of *The East Asian Miracle* study is how few generalizations apply to all eight economies.

The answers to questions beget further questions—in this case, what makes for successful interventions, and whether other countries can hope to succeed by pursuing a similar approach. Here the study emphasized the creation of institutions, including capable and reputable bureaucracies, whose procedures were shielded from political interference, and mechanisms of consultation among government, business, and others, including academics and journalists. In most cases these took the form of deliberation councils. Export promotion strategies generally keyed off world prices and, in some cases, used export targets and contests among local firms to provide incentives. Although all the economies except Hong Kong went through a phase of import substitution, these policies were later abandoned. Attitudes toward foreign direct investment varied, but in cases where it was encouraged the focus was on export promotion rather than import substitution. The study concluded that the promotion of specific industries generally had failed, while directed credit and the repression of financial systems had in some cases succeeded. In all cases the study emphasized that subsidies and distortions were limited—and modified or abandoned if they threatened macroeconomic stability.

The East Asian Miracle produced mixed reactions, with some complaining that it did not sufficiently emphasize the positive role of the state. But no reader could come away from the volume without a reinforced belief in the importance of getting the fundamentals right, while at the same time thinking that export promotion strategies had played an important role in development and that the quality of institutions matters a great deal.

Looking Back

If *World Development Report 1991* were to be rewritten today, its basic message probably would not change much. That is not because there is nothing new under the sun, but because most of what is in the report is analytically and empirically well founded. Although we should always emphasize the tenuous nature of our knowledge, we should also acknowledge that the field of development economics—which was born less than 60 years ago—continues to mature as the stock of theory, data, and country experiences on which it can draw increases.³ In the characteristically measured words of Michael Bruno's (1995, p. 17) keynote address to this conference in 1994, "a hard core of knowledge—small but increasing—has been sustained and buttressed through the turbulence."

At the same time, the experiences, reflections, and research of the 1990s—much of it presented in *The East Asian Miracle* and in *World Development Reports* since 1991—should lead to some changes in emphasis and views. What are they? The experience of the transition economies, some of it studied in *World Development Report 1996: From Plan to Market* as well as in the European Bank for Reconstruction and Development's excellent *Transition Reports*, largely supports the consensus view of *World Development Report 1991*. Some controversy might remain over the speed of adjustment. I believe that it should be very fast to achieve macroeconomic stability and price and trade liberalization. Other reforms, which are bound to take longer, should proceed as quickly as possible.

In coming years we will also need to draw lessons for the development consensus from the East Asian crisis. It is far too early to tell what these lessons will be. But some elements are clear—none of them clearer than the need for a robust banking and financial system. The Basle Committee's Core Principles go a long way toward summarizing what is needed, and more details are provided in the IMF's Framework for Financial Stability (Folkerts-Landau and Lindgren 1998). Weak supervision and regulation in this area sow the seeds of future crises, underscoring the point that government needs not only to get out of certain areas of regulation but also to strengthen others. The importance of the financial sector is not a new theme for the World Bank; it was the subject of *World Development Report 1989: Financial Systems and Development*. But the devastation that has been propagated by weak banking systems in the East Asian crisis—including in Japan—would surely strengthen the emphasis on the need for healthy financial systems in any future *World Development Report* on development strategies.

It is interesting to consider whether the current crisis will lead to a modification of the agnostic views on financial repression expressed in *The East Asian Miracle*. The answer should be related to the conclusions drawn from East Asia's experience with capital account liberalization. Recent experience should reinforce the widely held view that capital accounts should not be liberalized until domestic financial systems, including their regulation, are strengthened. It should also reinforce the urgency of strengthening domestic financial systems. Given that, I doubt that there remains a strong case for financial repression. But vulnerability to short-term capital outflows should be controlled through strong prudential regulations for the financial system and close monitoring of corporate borrowing from abroad. There is also a case for using market-based measures to control the pace and volume of short-term capital inflows, as is now done in several countries.

Beyond the financial system, the East Asian crisis will lead to a reexamination of the benefits of the close relations among government, business, and the financial sector that have been practiced in several East Asian economies. We are likely to conclude that the opacity of financial relations within the corporate sector and among these three sectors should not survive. That does not necessarily rule out a continuation of close consultation among the three sectors, and with labor as well.

Let me now list a few other topics that merit more attention in a future *World Development Report* on development strategies: efficient regulation, institutional development, governance, environmental regulation, urbanization, and income distribution.

- Enough has been said about the combination of strong state intervention in some parts of East Asian economies and inadequate financial sector regulation and supervision to make the point that any future *World Development Report* on development strategies will have to identify the types of regulation that are needed to strengthen economic performance (see Stiglitz 1997, pp. 11–23). Such an analysis could draw on some of the excellent material in *World Development Report 1997: The State in a Changing World*.
- By emphasizing the need for efficient bureaucracies and human capital creation, *World Development Report 1991* directed attention to the role of institutions in economic development. The question of how to build the institutions needed for economic development should be taken further in a future *World Development Report*. Among these institutions are the education system and an efficient government. The question of institutional development is one that development economics has struggled with for some time. But there should be further insights from the experience of transition economies, many of which already possessed technically knowledgeable individuals as these economies began their transition to a market system. This human capital base made it possible to develop some institutions, such as central banks, quickly. The development of other institutions, including, in Russia, the tax system, has proved more difficult. We need to find out why and see whether it is possible to do better—especially in countries where overall technical training is much lower.⁴

- The ability to run the government well is one meaning of governance, a topic that has gained increasing attention since the early 1990s. Governance in its other meaning—as it relates to corruption—also deserves more attention in a future *World Development Report* on development strategies.
- *World Development Report 1991* wrestled with the issue of the environment, and *World Development Report 1992: Development and the Environment* was devoted to the topic. The environment deserves more attention, and efforts could certainly draw on work done for *World Development Report 1992* and subsequently in the World Bank and elsewhere.
- Anyone who has recently visited the capitals of developing countries must have been struck by the problems of urbanization. This issue was discussed at some length in both the first and the second *World Development Reports*, for 1978 and 1979. Urbanization interacts with problems of population growth and surely deserves more attention in a future *World Development Report*.
- The distribution of income in many developing countries, especially in Latin America, is very unequal. Social justice, as well as the sustainability of the development effort, requires that we find ways of making development more equitable.

The subjects I have cited have been studied extensively in the Bank and elsewhere in the development economics field, and it is unlikely that their inclusion in an update of *World Development Report 1991* would produce major surprises or a major change in message.

Globalization

Next I turn to a subject that could produce major changes—globalization. Over the past 50 years the volume of world trade has increased more rapidly than GDP, and most economies have become ever more open to international trade. Despite the obstacles to trade against which we regularly and rightly inveigh, this trend is likely to continue—particularly since the ardor for regional trading arrangements has not diminished and as the World Trade Organization becomes more active. This trend is part of the process of globalization, including the globalization of production, which will shape the world economy in the decades to come.

At the same time, the globalization of international capital markets has accelerated at an extraordinary pace in the 1990s. If the Republic of Korea and Thailand stick with current reforms, even the East Asian financial crisis—like the Mexican peso crisis before it—will likely have only a passing effect on the volume of international capital flows. Still, increases in interest rates in industrial countries, which should not be ruled out, will probably cause a slowdown in funds flowing to emerging markets.

The Mexican and East Asian crises have demonstrated the power of international capital markets. Although some countries will conclude that they want to stay out of these markets, most will not. Two of the countries hardest hit in the tequila crisis—Mexico and Argentina—have continued to open their markets to foreign capi-

tal. Similarly, Korea and Thailand will be more, not less, open to international capital flows after this crisis.

Countries that want to participate in international capital markets will have to strengthen their macroeconomic policies and their financial systems. Capital market liberalization should be gradual and should take place only as the domestic financial system is strengthened and prudential and other controls are put in place. But most countries will liberalize, more or less gradually.

To deal with the risks posed by the globalization of capital markets, actions will be needed by industrial and developing country governments, to strengthen not only the international economic system but also their domestic economies. I will focus on one aspect of recent discussions about the strengthening of the international economic system: the notion that a variety of standards should be developed and put in place.

The Basle Committee's Core Principles provide an agreed standard for banking system behavior and supervision. Similarly, the IMF's Special Data Dissemination Standard constitutes an agreed international standard for statistical data. Codes of good practice can be envisaged for accounting standards, corporate governance, securities markets, and other aspects of private sector behavior. The IMF recently produced a code of good practices on fiscal transparency—Australia and New Zealand have already introduced their own such codes—and has been asked to develop a code of good practices with respect to monetary policies.

Such codes will provide a comprehensive set of rules that countries can implement to improve their economic performance and to guide international capital flows.⁵ If bank regulators in creditor countries cooperate, a system of risk weighting for investments in developing countries could emerge based on the extent to which such standards are observed. Of course, this would require that implementation of the standards be monitored and certified. In a rational world observance of the standards would also help determine the terms on which corporations and governments would have access to international capital markets. In this way macroeconomic policy, corporate transparency, and financial market regulation in developing countries could be improved using the discipline of international capital markets.

As globalization proceeds, the question of the optimal exchange rate arrangement and of the desirability of maintaining a national currency will again come to the fore. The arguments on these issues are well known. To cut them short, let me simply predict that if the euro succeeds—and I expect it to—we will likely see the development of several large currency blocs associated with large trading areas. These in turn could eventually—in the Keynesian long run—coalesce into a single currency.

The ongoing globalization of goods and capital markets promises to bring profound changes to the global economy and to individual economies. Although most advanced economies will be well equipped to deal with these changes, even some of them will have to improve institutions and policies to meet international standards. Most developing economies will need to take actions on many fronts to meet international standards. They will need help to strengthen their institutions and build

their human capital. Some that may not have access to international capital markets will need financial assistance. Here one thinks most of the needs of many Sub-Saharan African countries.

These are early days for the international economy that, spurred by the liberalization of trade and capital flows as well as by technological change in the financial sector, is continuing to emerge from the destruction wrought by the Great Depression and World War II. The system is accident-prone and lacks a regulatory authority and lender of last resort. We need to work on the architecture internationally and in individual economies to make it safer. But at the same time we should not forget that this system has brought unprecedented and sustained, though not uniformly shared, economic growth and prosperity—and that it will continue to do so for countries that follow the right policies.

Conclusion

Let me conclude with a confession. For a long time, even after *World Development Report 1991* was published, I believed that there was an elixir of growth, a magic ingredient missing from the set of policies listed in the report, that if included would make a miracle—even an East Asian miracle—possible. I no longer believe that. Or rather, I believe that I know the missing ingredient. It is hard work. For it is a long and arduous task, a matter of many people doing many things right over many years, to make a country grow.

Notes

1. The report also analyzed the contributions to the growth of developing countries that could be made by industrial countries and international financial institutions.
2. Amsden (1989) challenges the views of international financial institutions by arguing that in the Republic of Korea the government contributed to fast growth by systematically distorting market incentives (“getting prices wrong”). See also Wade (1990).
3. Although Adam Smith and other founding fathers of economics can be thought of as fundamentally concerned with economic growth and development, and although development economics draws on other branches of economics, especially trade theory, development economics as a field should probably be dated to the 1940s. Systematic empirical work had to wait for the development of a database, which began in the 1950s.
4. These issues are discussed in *World Development Report 1997*, which suggests focusing on the basics when the capacity of the state is limited.
5. The idea that such rules would improve economic performance is part of the German notion of *ordnungs-politik*, associated with Walter Eucken in the 1940s.

References

- Amsden, Alice. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.
- Bruno, Michael. 1995. “Development Issues in a Changing World: New Lessons, Old Debates, Open Questions.” In Michael Bruno and Boris Pleskovic, eds., *Proceedings of the*

World Bank Annual Conference on Development Economics 1994. Washington, D.C.: World Bank.

EBRD (European Bank for Reconstruction and Development). Various years. *Transition Report*. London.

Fischer, Stanley, and Dennis de Tray. 1990. "Introduction." In Stanley Fischer and Dennis de Tray, eds., *Proceedings of the World Bank Annual Conference on Development Economics 1989*. Washington, D.C.: World Bank.

Fischer, Stanley, and Alan Gelb. 1991. "The Process of Socialist Economic Transformation." *Journal of Economic Perspectives* (fall): 91–106.

Folkerts-Landau, David, and Carl Lindgren. 1998. "Toward a Framework for Financial Stability." *World Economic and Financial Survey*. International Monetary Fund, Washington, D.C.

Lipton, David, and Jeffrey Sachs. 1990. "Creating a Market Economy in Eastern Europe: The Case of Poland." *Brookings Papers on Economic Activity 1*. Washington, D.C.: Brookings Institution.

Sachs, Jeffrey, and Wing Thye Woo. 1994. "Structural Factors in the Economic Reforms of China, Eastern Europe, and the Former Soviet Union." *Economic Policy* 18 (April): 101–45

Stiglitz, Joseph E. 1997. "The Role of Government in Economic Development." In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1996*. Washington, D.C.: World Bank.

Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.

Williamson, John, ed. 1990. *Latin American Adjustment: How Much Has Happened?* Washington, D.C.: Institute of International Economics.

World Bank. 1993. *The East Asian Miracle: Economic Growth and Public Policy*. A Policy Research Report. New York: Oxford University Press.

———. Various years. *World Development Report*. New York: Oxford University Press.

*Is Geography
Destiny?*

The Role of Geography in Development

Paul Krugman

The recent surge of interest in the role of geography in economic development has divided into two seemingly contradictory approaches. One approach emphasizes the role of inherent features of the landscape in shaping development patterns. This article, however, mainly assesses the alternative approach, which stresses how the tension between centripetal forces (such as forward and backward linkages in production and increasing returns in transportation) and centrifugal forces (such as factor immobility and land rents) can produce a process of self-organization in which more or less symmetric locations end up playing very different economic roles. Such processes can occur at several levels. The article discusses geographic models of the division of the world into industrial and developing countries, of the emergence of regional inequality within developing countries, and of the emergence of giant urban centers. It argues that the conflict between “predestination” and “self-organizing” approaches to economic geography may be more apparent than real: natural features matter so much largely because they inspire self-reinforcing agglomerations. Thus geography may have been destiny in the past, but it need not be in the future. Finally, the article briefly discusses policy—mainly in terms of why it is so hard to draw policy conclusions from economic geography models.

In recent years there has been a surge of interest in the geographic aspects of development—that is, in the question of where economic activities take place. There is nothing surprising about this interest—or perhaps the surprise is that it took so long for this interest to become a mainstream concern within economics. After all, even a casual look at a map of the world suggests that differences in economic development are at the very least associated with location: countries close to the equator tend to be poorer than those in temperate zones, and per capita income in Europe seems to follow a downward gradient from the northwest corner of the continent. It is also apparent that there are large regional inequali-

Paul Krugman is professor of economics at the Massachusetts Institute of Technology

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

ties within countries and, often, a powerful tendency for populations to concentrate in a few densely populated regions and cities. But only recently have attempts to explain such patterns become a subject for research by large numbers of economists.

The new interest in economic geography usually takes one of two seemingly contradictory approaches. One approach—exemplified by John Luke Gallup and Jeffrey Sachs’s article in this volume—attempts to explain the differences in economic development between locations in terms of underlying, inherent differences in those locations. That is, it looks for associations such as the tendency of countries with tropical climates to have low per capita income, or of great cities to emerge where there are good harbors.

The other approach typically asks why the economic destinies of locations might diverge even in the absence of such inherent advantages or disadvantages—why small historical accidents can cause one country to become part of the industrial core while another becomes part of the primary-producing periphery, or why some more or less arbitrary location becomes the site of a megacity containing 10 million or more people. These two approaches may well seem contradictory: one seems to be a story of predestination, the other a story of chance. As I argue later in this article, however, the contradiction is more apparent than real. In fact, understanding why small random events can have large consequences for economic geography is also crucial to understanding why underlying differences in natural geography can have such large effects. Thus the two approaches turn out to be complementary rather than contradictory.

In any case, most of this article is devoted to understanding how the geography of the world economy—both between and within nations—can engage in a process of self-organization in which locations with seemingly identical potential end up playing very different economic roles.

Theoretical Principles of the New Economic Geography

Many economic activities are concentrated geographically. Most people in advanced countries, and a growing number in developing countries, live in large, densely populated metropolitan areas. Many industries—including service industries such as banking—are also concentrated geographically, and such clusters are an important source of international specialization and trade. Yet we do not all live in one big city, nor does the world economy concentrate production of each good in a single location. Why?

Centripetal and Centrifugal Forces

Obviously there is a tug of war between forces that promote geographic concentration and those that oppose it—between “centripetal” and “centrifugal” forces. These forces can be represented by the items shown in table 1. This list is not comprehensive; it is merely a selection of some forces that may be important in practice.

Table 1. Forces Affecting Geographic Concentration

<i>Centripetal forces</i>	<i>Centrifugal forces</i>
Market size effects (linkages)	Immobile factors
Thick labor markets	Land rents
Pure external economies	Pure external diseconomies

The centripetal forces listed in the first column of table 1 are the three classic Marshallian sources of external economies. A large local market creates both backward linkages—that is, sites with good access to large markets are preferred locations for the production of goods subject to economies of scale—and forward linkages—a large local market supports the local production of intermediate goods, lowering costs for downstream producers. An industrial concentration supports a thick local labor market, especially for specialized skills, so it is easier for workers to find employers and for employers to find workers. And a local concentration of economic activity may create more or less pure external economies through information spillovers.

The centrifugal forces in the second column of table 1 are less standard but offer a useful breakdown. Immobile factors—certainly land and natural resources, and in an international context people as well—militate against concentration of production, both from the supply side (some production must go to where workers are) and from the demand side (dispersed factors create a dispersed market, and some production will have an incentive to locate close to consumers). Concentrations of economic activity increase the demand for local land, driving up land rents and so discouraging further concentration. And concentrations of activity can generate more or less pure external diseconomies such as congestion.

In the real world agglomeration in general, as well as any example of it, typically reflects all the items in table 1. Why is the financial services industry concentrated in New York City? Partly because the city's size makes it an attractive place to do business and because the concentration of the financial industry means that many clients and ancillary services are located there. Also important are the city's thick market for those with special skills, such as securities lawyers, and the general importance of being in the midst of the buzz. But why isn't all financial business concentrated in New York City? Partly because many clients are not there, partly because office space is expensive, and partly because it is a nuisance to deal with the city's traffic, crime, and other urban realities.

To conduct analytical work on economic geography, however, it is necessary to cut through the complexities of the real world and focus on a more limited set of forces. In fact, the natural thing is to pick one force from the first column of table 1 and one from the second: to focus on the tension between just one centripetal and one centrifugal force. In the line of work on economic geography started by my 1991 article and book (Krugman 1991a, b) most models have chosen the first item in each column—analyzing linkages as the force for concentration and immobile factors as the force opposing concentration.

These choices are dictated less by empirical judgment than by two strategic modeling considerations. First, it is desirable to put some distance between assumptions and conclusions—to avoid an approach that appears to assert that agglomeration takes place because of agglomeration economies. Much of the analysis we want to undertake involves asking how a changing economic environment alters economic geography. This will be an ill-defined task if the forces producing that geography are inside a black box labeled *external effects*. So the pure external economies and diseconomies are put to one side, in favor of forces that are more amenable to analysis.

Second, if location is the issue, it is helpful to be able to deal with models in which distance enters in a natural way. Linkage effects, which are mediated by transportation costs, are naturally tied to distance; so is access to immobile factors. By contrast, the thickness of the labor market must have something to do with distance, but it does not lend itself quite so easily to being placed in a spatial setting. And land rents as a centrifugal force pose conceptual challenges—the “infinite Los Angeles problem”—that are discussed briefly in the section on chance and necessity below.

Modeling Tricks

The idea is hardly new that there may be a circular process in which the decisions of individual producers to choose a location with good access to markets and suppliers improve the market or supply access of other producers in that location. Indeed, that was the central theme of studies by Harris (1954) and Pred (1966), both well known among geographers. Why, then, did this idea not become widely known in economics until the 1990s?

The most likely answer is that underlying the work of Harris and Pred is the implicit assumption that there are substantial economies of scale at the level of the plant. In the absence of such scale economies, producers would have no incentive to concentrate their activity: they would simply supply consumers from many local plants. An expansion of a regional market would not predictably lead to an increase in the range of goods produced in that region. Increasing returns, in other words, are central to the story.

The same may be said of spatial economics in general. Almost all the interesting ideas in location theory rely implicitly or explicitly on the assumption that important economies of scale enforce the geographic concentration of some activities. Thus Weber's (1909) analysis of the location decisions of an individual producer trying to minimize the combined costs of production and delivery assumes that there can be only one production site; Christaller's (1933) suggestion that cities form a hierarchy of central places depends on the assumption that larger cities can support a wider range of activities; and Lösch's (1940) famous demonstration that an efficient pattern of central places would imply hexagonal market areas assumes that some economic activities can be undertaken only at a limited number of sites. (The main example of a location model that does not rely on some form of scale

economies, the land-rent analysis of von Thünen (1826), in effect hides the role of increasing returns by simply assuming the existence of a central city.) But unexhausted economies of scale at the level of the firm necessarily undermine perfect competition.

The reason geography has finally made it into the economic mainstream is therefore obvious: imperfect competition is no longer regarded as impossible to model, so stories that crucially involve unexhausted scale economies are no longer out of bounds. Indeed, the new interest in geography can be viewed as the fourth (and final?) wave of the increasing returns—imperfect competition revolution that has swept through economics over the past 20 years. First came the new industrial organization, which created a toolbox of tractable if not entirely convincing models of imperfect competition. Then the new trade theory, which used that toolbox to build models of international trade in the presence of increasing returns. Then the new growth theory, which did much the same for economic growth. What happened after 1990 was the emergence of the new economic geography, which might best be described as a genre of economic analysis that tries to explain the spatial structure of the economy using technical tricks to produce models in which there are increasing returns and markets characterized by imperfect competition. Fujita, Krugman, and Venables (forthcoming) summarize these tricks as results of “Dixit-Stiglitz, icebergs, evolution, and the computer.” Why, and how?

DIXIT-STIGLITZ. The remarkable model of monopolistic competition developed by Dixit and Stiglitz (1977) has become a workhorse in many areas of economics. In the new economic geography it has one especially appealing feature: because it assumes a continuum of goods, it lets modelers respect the integer nature of many location decisions—no fractional plants allowed—yet analyze their models in terms of the behavior of continuous variables like the share of manufacturing in a particular region. In effect, Dixit-Stiglitz lets us have our cake and cut it into arbitrarily small pieces too.

ICEBERGS. “Icebergs” are a less familiar technical trick. Transportation costs are of the essence in the new economic geography. Yet any attempt to develop a general equilibrium model of economic geography would be substantially complicated by the need to model transportation as well as goods-producing sectors. Worse yet, transportation costs can undermine the constant demand elasticity that is one of the crucial simplifying assumptions of the Dixit-Stiglitz model. Both problems can be sidestepped with an assumption first introduced by Samuelson (1954) in international trade theory: that a fraction of any shipped good simply “melts away” in transit, so that transport costs are in effect incurred in the good shipped. (In new economic geography models melting is usually assumed to take place at a constant rate per distance covered—for example, 1 percent of the cargo melts away per mile.) In terms of modeling convenience there turns out to be a spectacular synergy between the Dixit-Stiglitz market structure and “iceberg” transport costs: not only can one avoid the need to model an additional industry, but because the transport

cost between any two locations is always a constant fraction of the f.o.b. (free on board) price, the constant elasticity of demand is preserved.

EVOLUTION. Interesting stories about economic geography often seem to imply multiple equilibria. Suppose, for example, that producers want to locate where other producers choose to locate; this immediately suggests some arbitrariness about where they actually end up. But which equilibrium does the economy select? New economic geography models typically assume an ad hoc process of adjustment in which factors of production gradually move toward locations that offer higher current real returns. This sort of dynamic process was initially proposed apologetically, since it neglects the role of expectations. But it is possible to regard models of geography as games in which actors choose locations rather than strategies—or rather, in which locations are strategies—in which case one is engaged not in old-fashioned static expectations analysis but rather in state-of-the-art evolutionary game theory! (To middle-brow modelers like myself it sometimes seems that the main contribution of evolutionary game theory has been to relegitimize those little arrows we always want to draw on our diagrams.)

THE COMPUTER. Finally, despite the best efforts of theorists, all but the simplest models of economic geography usually turn out to be beyond the reach of paper-and-pencil analysis. As a result the genre relies to an unusual extent on numerical examples—on the exploration of models using both static calculations and dynamic simulations.

Dynamics of Geographic Change

Suppose that an economic activity has a slightly larger initial concentration in one location than in another. Will that concentration be self-reinforcing, with a growing disparity between the locations, or will there be a tendency back toward a symmetric state? The answer presumably depends on the relative strength of centripetal and centrifugal forces.

Suppose, on the other hand, that a concentration of economic activity already exists, but that some of that activity moves elsewhere. Will the activity move back, or will the concentration unravel? The answer to this question similarly depends on the relative strength of centripetal and centrifugal forces.

As these generic questions suggest, models of economic geography typically exhibit a pattern in which the qualitative behavior of the model changes abruptly when the quantitative balance of forces passes some critical level. That is, the models are characterized by bifurcations. And bifurcation diagrams are therefore a central analytical tool in this literature.

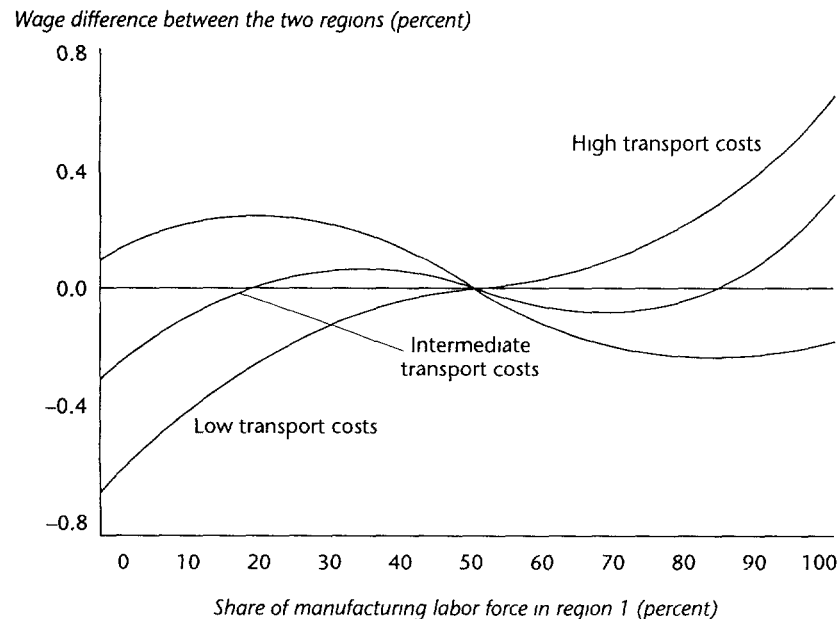
The typical form of these bifurcations are illustrated in figures 1 and 2, which show results from a simulation of the model introduced in Krugman (1991b). That paper was, in effect, an attempt to formalize the story suggested by Harris (1954) and Pred (1966). The model envisaged an economy consisting of two regions, each

with two industries: immobile, perfectly competitive agriculture and mobile, imperfectly competitive (Dixit-Stiglitz) manufacturing. The backward and forward linkages in manufacturing generated centripetal forces; the pull of the immobile farmers generated the centrifugal force.

Figure 1 shows how the difference in real wages between the two regions depends on the allocation of manufacturing between them (a calculation that involves repeatedly solving a small computable general equilibrium model). The horizontal axis shows the share of manufacturing workers living in region 1; the vertical axis shows the difference between real wages in region 1 and region 2. Each curve is calculated for a different level of transport costs.

The rough intuition behind these curves runs as follows. If transport costs are high, there is relatively little interregional trade. So the wages workers can earn depend mainly on the amount of local competition and thus decrease as the number of other workers in the same region increases. When transport costs are low, a typical firm sells extensively in both regions. But since it has better access to markets if it is located in the region with the larger population of workers, it can afford to pay higher wages—and the purchasing power of those wages is also higher because workers have better access to consumer goods. So in that case real wages increase with a region’s population of workers. At intermediate transport costs these two forces are nearly balanced. The particular curve shown, in which centripetal forces are stronger when regions are very unequal, while centrifugal forces are stronger when they are nearly symmetric, is an artifact of the particular functional forms used in this exercise.

Figure 1. Relationship between Regional Manufacturing Populations and Real Wages with Varying Transport Costs



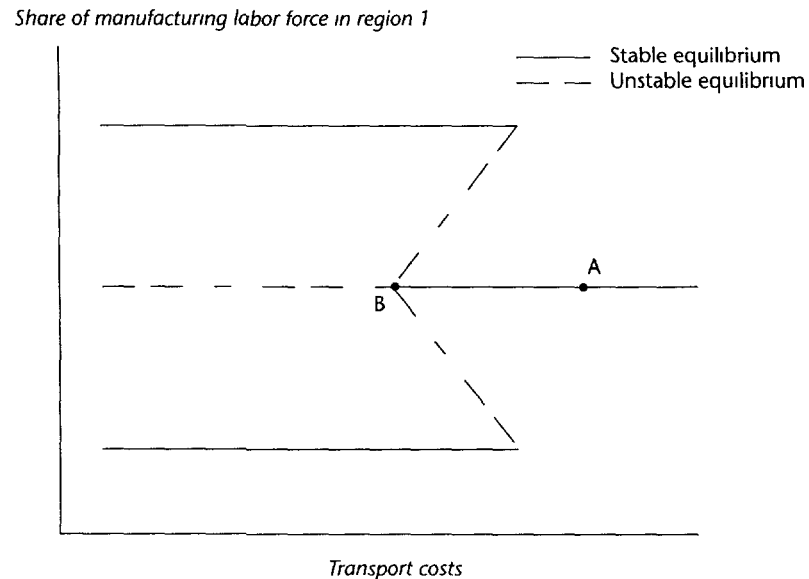
Because workers are assumed to move to whichever region offers the higher real wage, in the case of high transport costs there is a unique equilibrium with workers evenly divided between the regions. In the case of low transport costs there are three equilibria—one with workers evenly divided, two with workers concentrated in either region. And in the intermediate case there are five equilibria.

Figure 2 shows the bifurcation diagram that results from the assumption that workers gradually move toward the region offering the higher real wage. It shows how the set of equilibria—as measured by the share of the manufacturing labor force in region 1—depend on transport costs, with solid lines indicating stable and broken lines unstable equilibria. The figure illustrates nicely one of the appealing features of the new economic geography: it easily allows one to work through interesting “imaginary histories.” Suppose, for example, that we imagine an economy that starts with high transport costs and therefore with an even division of manufacturing between regions, a situation illustrated by point A in figure 2. Then suppose that transport costs were to fall. When the economy reached point B, it would begin a cumulative process in which a growing concentration of manufacturing in one region would lead to an ever-larger concentration of manufacturing in that region. That is, the economy would spontaneously organize itself into a core-periphery geography.

Geographic Theories of the World Economy

A generation ago it was common for critics of the economic system to argue that developing countries were not simply economies on the same road as industrial economies, though less advanced. Rather, they argued, the emergence of rich and

Figure 2. Relationship between Regional Manufacturing Populations and Transport Costs



poor countries was part of a common process of uneven development in which initial advantages in certain regions had accumulated over time, giving them a privileged economic position while relegating the rest of the world to a subordinate role as hewers of wood and drawers of water. In the past 10 years worries have largely reversed: now advanced countries seem to fear that newly industrializing economies will undermine the North's prosperity.

The Theoretical World, with Two Economies

New economic geography models can shed light on both concerns. The models suggest that both the differentiation of the world into high-wage industrial core and low-wage nonindustrial periphery, and a subsequent period of dispersal of industry and convergence of wages, can be explained by an ongoing process of declining trade costs.

The basic concepts were introduced by Venables (1995). He assumed, in contrast to the regional model described in the previous section, that factors were completely immobile between countries. However, a possibility for cumulative processes was introduced by making a distinction between a constant-returns agricultural sector and an increasing-returns manufacturing sector that both uses and produces intermediate inputs. The basic idea is that intermediate goods producers in a region with a large manufacturing sector will have superior access to the large markets afforded by downstream producers (backward linkage), while these producers in turn will have the advantage of better access to the intermediate goods produced in their own region (forward linkage). In the original formulation the upstream and downstream components of manufacturing were treated as separate sectors; in subsequent work, including Krugman and Venables (1995) and Puga and Venables (1997), the same differentiated goods were assumed to enter into consumption and production, allowing a consolidation of the sector into a common manufacturing aggregate.

Suppose now that we imagine a world consisting of two initially identical regions, with varying costs of transporting manufactured goods between them. If transport costs are high, each region will essentially be self-sufficient, and the regions will therefore be symmetric in outcomes as well as initial conditions. But now imagine gradually falling transport costs. It now becomes increasingly possible for firms to export their manufactured goods to the other region. Yet because of transport costs, production in whichever region has the larger manufacturing sector (because of any small difference, or simple historical accident) will benefit from better access to both markets and suppliers. Thus when transport costs drop below some critical level a process of differentiation between regions will take place, with manufacturing concentrating in a core while the periphery is relegated to primary production.

The impact of this process depends on the size of the manufacturing sector—more specifically, on the share of manufactured goods in spending. If this share is low, the region that becomes the core does not get a significantly higher wage rate from that role. But if the share is sufficiently large (in a two-region model, if it exceeds half of total spending on traded goods), the core ends up with higher wages

than the periphery, and the process of differentiation can be immiserizing for the peripheral region. This simple approach, then, offers a possible justification for claims that the backwardness of the South is not something that developed in isolation: it is a necessary consequence of the process that also produced the industrialization of the North.

Perhaps more surprisingly, the same model predicts that a continuing decline in transport costs—loosely speaking, the continuing process of globalization—eventually produces a reversal of fortune. The reason is that the peripheral region has a competitive advantage in the form of lower wages. At first this advantage is more than offset by the North's superior access to markets (backward linkage) and inputs (forward linkage). But as transport costs fall, the importance of these linkages also declines. So there is a second critical point at which industry finds it profitable to move to lower-wage locations.

This is a surprisingly satisfying result: by imagining a hypothetical history in which a single driving variable—transport costs—follows a monotonic path through time, we are able to derive an evolutionary path for the world economy in which the inequality of nations and the division of the world into primary and industrial producers first spontaneously emerges, then dissolves. Understandably, then, Venables and I referred to the original paper as the “history of the world, part I.”

The Real World, with Many Economies

I will return shortly to the question of how much of the history of the real world such an analysis actually captures. First, however, it is useful to use the geographic theories of the world economy as an occasion to discuss the “spatial” aspects of modeling.

The analysis in Krugman and Venables (1995), like much international trade theory, imagines a world with just two discrete locations, themselves modeled as points. It involves space only to the extent that there are assumed to be transport costs between these points. To a serious geographer, of course, this is grossly inadequate: the spatial relationships both between and within countries should be taken into account. Indeed, as a first approximation a geographer might even want to ignore national boundaries, asking how an undifferentiated, “seamless” world economy might develop a spatial structure.

To do this in general is probably impossible. Indeed, as soon as one goes even a bit beyond a two- or three-location world, the whole exercise tends to bog down in uninformative taxonomy. But it is possible to gain considerable insight by focusing on particular, unrealistic, but convenient “geometries” for the world.

One particular geometry that is useful despite its artificiality is what we might call the “racetrack” economy: a large number of regions located symmetrically around a circle, with transportation possible only around the circumference of that circle. This setup has two useful properties. First, the economy is one-dimensional, which greatly simplifies both algebra and calculations. Second, because there are no edges and hence no center, it is a convenient way to retain the feature that all sites are identical—which means that any spatial structure that emerges represents pure self-organization.

If one takes a racetrack version of the Krugman-Venables model (1997) and starts it with an almost but not quite uniform distribution of manufacturing across space, what happens is a spontaneous differentiation into manufacturing and agricultural regions. The size and spacing of these regions are predictable, even if the initial deviation from uniformity is random. The reason for this predictability was, it turns out, explained in a seemingly different context—morphogenesis in theoretical biology—by, of all people, Alan Turing (1952). But the question of which parts of the world take on which role remains arbitrary, a function of small initial advantages that determine the “phase” of the regional development pattern (that is, how the alternating bands of industry and agriculture are rotated around the circle).

Extending this sort of analysis to more realistic geometries turns out to be startlingly difficult. Still, the racetrack analysis is at least suggestive of the reasons that patterns of development and underdevelopment are regional—why, for example, all of northwestern Europe shared in the industrial revolution—rather than confined within national boundaries.

What about the larger story of the rise and fall of international inequality? Surely the forces covered in this approach do not tell the full story, or perhaps even more than a small part of the real story. In particular, if one tries to put realistic shares of North-South trade in gross world product into the model, it is difficult to make either the initial divergence of incomes as the world divides itself into industrial and primary-producing regions, or the later spread of industry, have impacts on real income in either region of more than a few percent. There may be ways to make the story take on greater significance—say, by introducing some interaction between patterns of trade specialization and external economies in domestic production. But at this point it would be premature to take the interesting and suggestive “history of the world” as more than a possible story about part of what actually happened.

Regional Inequality in Developing Countries

It is often observed that many developing countries suffer from significant economic dualism—in which a relatively high-wage, high-income economy appears to exist within a much less developed economy—and that this dualism has a strong geographic dimension. Although a lot of development economics continues to treat countries as dimensionless points, in other contexts the contrast between Mexico City and Chiapas, or between São Paulo and Brazil’s northeast, looms large.

It is not difficult to convert the core-periphery analysis discussed above into a story of regional divergence. One need only relabel the workers of that model with mobile factors such as capital and skilled labor, and presume that unskilled labor is a (relatively) immobile factor, so that it takes on the role of the farmers. The story can be made more realistic, adding complications but no essential differences, by allowing the mobile and immobile factors to be substitutes in production. With sufficiently strong scale economies and transport costs, the resulting core-periphery equilibrium can have large wage differentials for the immobile factor.

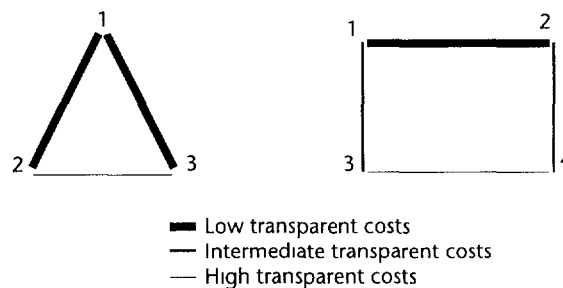
In words, this story says that Brazil's south is a more attractive place to produce than its north because of the concentration of purchasing power and availability of intermediate inputs in the south; and that because of this attraction those factors of production that can move have concentrated in the south, sustaining the concentration of markets and suppliers that creates the south's advantage. As in all the models discussed in this article, the original source of the south's advantage need not lie in any inherent superiority of its resources or location: it could simply be the result of historical accident.

While this is a coherent story, some modeling of regional inequalities has suggested an additional source of those inequalities: self-reinforcing advantages of market access through transportation networks. One simple version of this story was laid out in Krugman (1993) and is illustrated in the left side of figure 3. The figure shows three locations; the width of the lines between the locations is an inverse indicator of transport costs (that is, thicker lines mean lower costs, just as thicker means better on a road map). As drawn, location 1 is obviously a transport hub in the sense that it is cheaper to get from 1 to either of the other locations than it is to go between those locations. It is easy to show that other things being equal—that is, given the same market sizes and availability of locally produced inputs—this will make location 1 more attractive for producers subject to increasing returns. So a transport hub will be a favored location for industry. (Like many observations in the new economic geography, this is a painfully obvious point that somehow just was not in the literature before.)

But why should transport costs be lower between location 1 and other locations than between those other locations? One obvious answer is that if industry is concentrated in location 1, there will be more trade between 2 and 1 than between 2 and 3, and so on. And if there are increasing returns in transportation—as there surely are—this will mean lower per unit transport costs along the more heavily used routes.

Clearly, we have another example here of a self-reinforcing process: a location that for whatever reason has a concentration of production will tend to become central in terms of the transport network, which will reinforce its advantage as a pro-

Figure 3. Transparent Costs, Transparent Hubs, and Industrial Concentration



duction location, and so on. Krugman (1993) shows that this process can produce a core-periphery pattern of industrialization even if we suppress the factor mobility that drives the standard models of such patterns.

A slightly different role for favored transport access is illustrated on the right side of figure 3. Here we see four locations, with transport costs lower between 1 and 2 than between either of those locations and the rest of the economy, and with transport costs between locations 3 and 4 particularly high. This pattern might emerge, again in the presence of increasing returns in transportation, if locations 1 and 2 both had large concentrations of industry. The effect, of course, would be to make locations 1 and 2 more attractive places to do business, reinforcing their advantage. A concrete example: part of São Paulo's advantage is its good access to Rio de Janeiro, including frequent plane flights, and vice versa. This is natural between Brazil's two largest cities, but further reinforces the tendency of activity to concentrate in those two cities.

Just as in the global economy models discussed earlier, models of regional inequality can easily show a nonmonotonic response to declining transport costs. Initially such declines can promote the formation of core-periphery patterns. To take a classic example, the stark division of Italy into affluent north and less affluent south took shape when railroads were introduced. Railroads made it possible for factories in the north to supply the needs of agricultural markets in the south, causing deindustrialization in the south. Moreover, the railroad network did more to connect the already industrialized regions of the north than those of the south, reinforcing the advantage of those northern locations in terms of access to markets and inputs.

Eventually, however, sufficiently low transport costs (even on a small scale of transportation) can lead to a spread of industry: once it is inexpensive to transport inputs wherever they are needed and export products from any location, the lower factor costs of the periphery become increasingly significant. (In Brazil there is currently some relocation of industry to the northeast, where wages are about one-third the levels in São Paulo. This is one of the factors often blamed for rising unemployment in traditional industrial areas.) Of course, regional inequality may also be strongly influenced by government policy—including trade policy, as described below.

Policy and Primacy

A striking feature of many developing countries is the existence of one huge urban concentration, normally the capital city. Why are urban giants in developing countries so large?

Empirical studies of "primacy" identify two strong factors determining the size of the largest city: urban population as a whole and, more interestingly, political structure—primary cities are smaller in federal and decentralized systems than in highly centralized systems. Thus Mexico City is still larger than Shanghai, because of China's decentralization.

The role of political centralization in primacy is fairly obvious at one level: it results from the direct demand and employment created by the government appa-

ratus and from the more subtle advantages of access to government officials. (When one asks Japanese executives why they are willing to pay the high cost of keeping their headquarters in central Tokyo, access to officials is usually the first thing they mention.)

The type of analysis described in this survey suggests, however, that beyond these direct effects one might well expect a multiplier effect, perhaps even a catalytic effect of political centralization (see the section on geography and policy, below). That is, whatever initial concentration of demand and advantages of access are conveyed to businesses in the capital will be magnified through the usual circular processes involving market size, access to suppliers, transportation advantages, and so on. Such magnification effects may explain the extraordinary strength of the relationship between political centrality and primacy (for example, the fact that Tokyo is substantially larger than New York even though Japan has only half as many people as the United States).

There may also be other important policy linkages. Hanson (1992) notes that Mexico's trade liberalization in the late 1980s seemed to be associated with a dramatic decentralization of manufacturing away from Mexico City—not only with the growth of new export centers near the U.S. border, but also a spinning out of industries producing for the domestic market. In Krugman and Livas Elizondo (1996) an effort was made to justify this observation in terms of a formal model. The article envisaged a domestic economy with two locations and mobile manufacturing; the necessary centripetal force was supplied by backward and forward linkages, the centrifugal force by land rents. However, these two locations were assumed to trade (but not have factor mobility) with a large third region, the rest of the world.

The point we then made was that the importance of the linkages supporting population concentration *within* this country would depend on its trade policy. Suppose that the country was strongly protectionist and hence did little external trade. Then domestic producers would mainly sell to domestic consumers and buy inputs from other domestic producers. The result would be strong linkage effects that would tend to promote and sustain a concentration of manufacturing in only one location. But if trade were liberalized, domestic producers would sell much of their output abroad—and hence have less incentive to locate near the large domestic market—and would also buy many of their inputs from abroad—and hence have less incentive to locate near domestic suppliers. Meanwhile, high land rents would still create an incentive to locate away from other producers. Numerical examples confirm that high trade barriers would tend to foster concentration of manufacturing in a single Mexico City-type location, while reduced trade barriers would tend to cause such concentrations to unravel. (An interesting question would be whether Brazil's trade liberalization has similarly contributed to the apparent shift of manufacturing away from its traditional centers in the south. If so, it would be a cleaner example of our story than the case of Mexico, since proximity to the border is not an issue—indeed, given the Mercosur trade union the border issue actually cuts the other way.)

For what it is worth, cross-sectional regressions by Ades and Glaser (1995) find evidence that inward-looking trade policies foster the creation of urban giants,

although other factors appear to be more important. However, one may question whether the highly nonlinear stories told by the models can be tested very well by such regressions. (Empirical work in this area is generally difficult for that reason.)

Chance and Necessity

At the beginning of this article I described two approaches that both go under the rubric of geography but seem to take diametrically opposed positions: the type of model described above, in which there are multiple equilibria and the geographical pattern of production depends on historical accidents, and the approach recently promoted by John Luke Gallup and Jeffrey Sachs (see elsewhere in this volume), in which differences in natural geography exert powerful influences on economic development. But I also suggested that this may be a false dichotomy.

To illustrate this point, consider Mexico City. The concentration of population and production in the Valley of Mexico has deep historical roots, essentially environmental in nature: before the Spanish conquest the Aztecs practiced a highly productive form of agriculture made possible by the existence of a large lake, which supported a dense local population (by pre-industrial standards). It was natural that this location should become the site for Mexico's main urban center. But the valley no longer contains a lake, or for that matter any agriculture to speak of. Today Mexico City is there because it is there, its existence sustained by the kinds of circular processes discussed in earlier sections. So in one sense the location of Mexico's primary city was dictated by natural geography. Yet those geographic advantages are no longer relevant in any direct sense, and they have been able to cast such a long shadow over the future only because the geography of the economy has such strong self-reinforcing features that a concentration of population, once established, tends to persist and even grow. (The role of the Erie Canal in giving New York City its dominant position is a classic first-world example of the same proposition.)

Put another way, in many cases aspects of natural geography matter a lot not because natural features of the landscape are crucial, but because they inspire self-reinforcing agglomerations. So it is precisely the aspects of the economy that in principle allow history-dependent, multiple-equilibria stories to be told that in practice give exogenous geography such a strong role.

In formal models of economic geography, especially when one allows the geography of the economy to evolve over time, it often turns out that small nonhomogeneities in the landscape have dramatic effects on the outcome. Thus in the core-periphery models of the first two sections, giving one of the regions a small advantage in the size of its agricultural base removes the arbitrariness of which region will become the core and which the periphery as transport costs fall below the critical level. This means that a small difference in inherent advantage can produce a large difference in outcomes. (It also turns out that small inherent differences strongly bias the outcome when one starts with some random allocation of mobile factors.)

Most recent work making this point has concentrated on the effect of natural differences in transport costs on urban location—explaining why, for example, most

great cities are ports, even though in the modern world few large cities derive much of their income or employment from that role (Fujita and Mori 1996a).

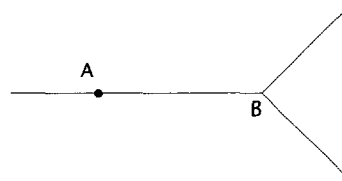
It is possible to imagine a variant on the models developed earlier in which all factors except land are mobile. In such a model it is possible, provided the economy is not too large, to have a self-sustaining “von Thünen” spatial pattern in which manufacturing is concentrated at a single location surrounded by an agricultural hinterland. However, if one imagines a gradually increasing population, eventually it becomes profitable for some manufacturing to locate away from the original center, and new cities emerge.

But where do they emerge? Figure 4 represents a version of Fujita and Mori’s analysis. Suppose the economy is in a long, narrow valley (making it effectively one-dimensional), with the original city at location A. If a fork is put in the valley at location B, it is effectively a point with superior access to the rest of the “world” than other locations—which makes it a stylized representation not only of the role of a river or road junction, but of a port as well.

As the population expands, location A’s agricultural hinterland will expand as well, eventually pushing up both forks of the valley beyond location B. And eventually a new city will emerge. Where? Location B is a very likely location, in the following sense: imagine choosing alternative initial positions for location A (or varying any other parameter of the model) and asking where the next city will emerge. In general, any possible location will be chosen for at most one location of the original city. But because of the special advantages of location B (which turn out to generate a cusp in the market-potential function that determines location choice), there is a nonzero-length *range* of initial city locations that will lead the second city to emerge there. So natural geography will often (though not always) dictate the city site. Yet once the city is established those natural advantages will be much less important a reason for the “lock-in” of its location than the self-sustaining advantages of an established concentration of activity.

The paradox that natural geography may matter so much precisely because of strong circular causation has important implications for the interpretation of correlations between natural advantages and actual economic geography. These correlations may say more about the processes that have produced the geography we see than about what might be possible in the future. To take the Fujita-Mori analysis as an example: the historical role of ports as sites around which cities crystallize explains why most of today’s large cities are ports. But because the importance of

Figure 4. Formation and Location of New Cities



the port was *only* that of serving as a springboard, and is not a major current source of advantage, it need not be the case that future cities also be ports. If, say, an inland city were constructed as a deliberate national policy and supported effectively, it might become self-sustaining even though its location does not fit any of the criteria that characterize today's major cities.

To put a sharper point on it: the current pattern of world economic geography shows a strong association between per capita income and essentially Western European conditions—temperate climate, absence of malaria, much of the population close to the coast or navigable rivers, or both. But this pattern may mainly reflect the catalytic role of these factors in the past and need not imply that an inland country (which now has access to good roads and cheap air transport) with a hot climate (but now has access to modern cooling technology) and environmental conditions that once made it malarial (but not now thanks to mosquito eradication programs) cannot break free of its low-level trap and move to a better equilibrium. All of this brings us to policy.

Geography and Policy

This will, necessarily, be a short conclusion. At this point little effort has been made to draw policy conclusions from the new economic geography literature. The main goal for the moment is to explain why.

In principle, the sort of economy envisaged by the models sketched out in this article should be a prime target for government intervention. There is no presumption here that the market will get it right. Moreover, the models suggest that under some circumstances small policy interventions can have large and perhaps lasting effects. Finally, because cumulative processes of concentration tend to produce winners and losers, perhaps at the level of nations, there is an obvious incentive for policymakers to try to make sure that their nation emerges as one of the winners.

Nonetheless, those of us working on these models have been extremely cautious about drawing policy implications. Mainly this reflects a strong sense of how difficult it is to go from suggestive small models to empirically based models that can be used to evaluate specific policies. The long debate over the applicability of the theory of strategic trade policy, which eventually led mainly to an appreciation of just how hard it is to map reality into even sophisticated models of imperfect markets, is fresh in the minds of many of the relevant theorists. And new geography models—in which the crucial effects are general equilibrium rather than merely partial equilibrium—are likely to be even harder to make operational.

There is also, to be honest, concern (at least on my part) that some of the less pleasant aspects of the history of strategic trade policy will be repeated: the frantic efforts of interested parties to recruit reputable economists to endorse questionable interventionist policies. Admittedly, that temptation was admirably resisted by all the major players in the new trade theory—but it was not an experience one wants to encourage.

But there is also a special consideration that makes policy conclusions difficult in the geographic literature. Consider table 1 again, bearing in mind that in most cases

all the entries will be relevant. What is immediately striking is that there are external effects on both sides. So there is a market failure case to be made both that any given agglomeration is too big (look at the congestion and pollution) and too small (think of the linkages and spillovers that would come with more activity). One may have opinions—I am quite sure in my gut, and even more so in my lungs, that Mexico City is too big—but gut feelings are not a sound basis for policy.

One recommendation is safe, however. Because geography is such a crucial factor in development, and there are undoubtedly strong policy implications of some sort, it is an important subject for further research.

References

- Ades, A., and E. Glaser. 1995. "Trade and Circuses: Explaining Urban Giants." *Quarterly Journal of Economics* 110: 195–227.
- Christaller, W. 1933. *Central Places in Southern Germany*. Jena: Fischer Verlag.
- Davis, D., and R. Weinstein. 1997. "Empirical Testing of Economic Geography: Evidence from Regional Data." Harvard University, Cambridge, Mass.
- Dicken, P., and P. Lloyd. 1990. *Location in Space: Theoretical Perspectives in Economic Geography*. New York: HarperCollins.
- Dixit, Aunash, and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67: 297–308.
- Fujita, Masahisa, and Paul Krugman. 1995. "When Is the Economy Monocentric: von Thünen and Christaller Unified." *Regional Studies and Urban Economics* 25 (August): 505–28.
- Fujita, Masahisa, and Tomoya Mori. 1996a. "The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub-Effect." *Journal of Development Economics* 49 (April) 93–120.
- . 1996b. "Structural Stability and Evolution of Urban Systems." *Regional Science and Urban Economics*.
- Fujita, Masahisa, Paul Krugman, and Anthony Venables. Forthcoming. *The Spatial Economy*
- Fujita, Masahisa, Tomoya Mori, and Paul Krugman. 1999. "On the Evolution of Hierarchical Urban Systems." *European Economic Review* 43 (February). 209–51.
- Harris, C. D. 1954. "The Market As a Factor in the Localization of Production." *Annals of the Association of American Geographers* 44: 315–48.
- Hoover, Edgar M., and Raymond Vernon. 1959. *Anatomy of a Metropolis*. Cambridge, Mass.: Harvard University Press.
- Karaska, G., and D. Bramhall, eds. 1969. *Locational Analysis for Manufacturing*. Cambridge, Mass.: MIT Press.
- Krugman, Paul. 1991a. *Geography and Trade*. Cambridge, Mass. MIT Press.
- . 1991b. "Increasing Returns and Economic Geography." *Journal of Political Economy*
- . 1993. "On the Number and Location of Cities." *European Economic Review* 37 (April): 293–98.
- Krugman, Paul, and R. Livas Elizondo. 1996. "Trade Policy and the Third World Metropolis." *Journal of Development Economics* 49: 137–50.
- Krugman, Paul, and Anthony Venables. 1995. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics* 110: 857–80.

- . 1997. "The Seamless World: A Spatial Model of International Specialization and Trade." Massachusetts Institute of Technology, Cambridge, Mass.
- Lösch, August. 1940. *The Economics of Location*. Jena: Fischer Verlag.
- Pred, A. R. 1966. *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800–1914*. Cambridge, Mass.: MIT Press.
- Puga, D., and Anthony Venables. 1997. "The Spread of Industry: Spatial Agglomeration in Economic Development." CEPR Working Paper 1354. Centre for Economic Policy Research, London.
- Samuelson, Paul. 1954. "The Transfer Problem and Transport Costs." *Economic Journal* 64 (254): 264–89.
- Turing, Alan. 1952. "The Chemical Basis of Morphogenesis." *Philosophical Transactions of the Royal Society of London* 237: 37.
- Venables, Anthony. 1995. "Equilibrium Locations of Vertically Linked Industries." *International Economic Review*.
- von Thünen, J. 1826. *The Isolated State*. London: Pergamon.
- Weber, A. 1909. *Theory of the Location of Industries*. Chicago: University of Chicago Press.
- Weibull, M. 1995. *Evolutionary Game Theory*. Cambridge, Mass.: MIT Press.

Comment on “The Role of Geography in Development,” by Paul Krugman

Paul A. David

The models of spatial resource allocation and trade in Paul Krugman’s article live up to expectations. They exhibit all the parsimonious elegance, expositional lucidity, and deft topicality that have characterized their creator’s many contributions to the literature of economics. They reveal, in the sparest ways conceivable, how various pairings of centripetal and centrifugal forces can result in simple but structured equilibrium configurations in the spatial distribution of specialized production. And their formal details lay bare the essential character of the mechanisms through which sunk cost hysteresis allows past events to exercise a tenacious influence on the geography of economic development.

Krugman’s modeling arts are deployed on this occasion to propel development economists more deeply into space—that is, to make them consider more carefully the ideas and implications of the new economic geography. His approach is well chosen, as it undoubtedly appeals to the predilection of many economists to favor simple, abstract models that “account for” idealized empirical regularities. And in the complex and potentially messy world of practical development policy, any clear, sharp, and reasonably robust insights that theorists can supply should be welcomed. The only thing that is holding the flow of my encomiums in check is the thought that encouraging the discipline’s inordinate fondness for a vision of the world in which order and simplicity reign might be not only delusional in the case of development economics, but also positively mischievous. That is, the exceptional pedagogical virtues of Krugman’s presentation and models may actually limit their utility as vehicles for carrying forward economic research on historical processes of regional development.

In what follows I elaborate on three facets of my friendly demurrer from the excessive clarity of Krugman’s article. First, I comment on the relationship between the underlying empirical phenomena and the “stylized facts” addressed by the new economic geographers. Next, I point to some dynamic supply-side processes that warrant greater attention than they have received. Their existence creates a sharp contrast

Paul A. David is senior research fellow at All Souls College, titular professor of economics and economic history in the University of Oxford, and professor of economics at Stanford University.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

between the clarity and elegant simplicity of Krugman's spatial models, which might be likened to line drawings in the style of Henri Matisse, and representations of actual historical processes, which bear a closer resemblance to the dense, intricate, interweaving motions captured on the canvases of Jackson Pollock. Finally, I counter what strikes me as an overly optimistic view of the relationship between the work of theorists and the political economy of regional development strategies. Such a view might be drawn from Krugman's worries that misguided policy measures could be improperly attributed to the modeling results obtained by the new economic geographers. But given the inefficiencies that are incipient in long-standing regional government practices of subsidizing the relocation of industrial plants, more than mere self-restraint would seem to be called for on the part of the new economic geographers.

NEGs, POGs, and Stylish Naked Models in Space

One of the consequences of a theory-driven social science is that the unrelenting competition to fashion ever more stylish explanatory theory inexorably drives down the ratio of data to models by enlarging the denominator. Of course, the effort that first has to be put into specifying a clear, common explanatory goal for the contestants tends to work in the same direction by reducing the numerator. Among Nicholas Kaldor's enduring but problematic contributions to economics was his introduction of the term *stylized facts*, by which we now refer to the results of the latter, data-reducing activities. Idealized empirical regularities of that variety are both "meta facts" and "quasi facts"; they are the outcome of repeated smoothing operations, in the course of which a large part of the original informative detail has been polished away.

Much of the applied theory created by the new economic geographers (NEGs) has focused on rationalizing a collection of abstract but nonetheless intriguing topographical regularities that supposedly obtain ubiquitously across time and space: the spontaneous organization of market territories into hexagonal configurations, the emergence of regional systems of cities differentiated by size and volume of commercial interactions according to their economic functions, the idealized geography of industrialization in which compact core zones of manufacturing are surrounded by peripheral regions where agricultural and other extractive activities remain dominant, the emergence of urban hierarchies characterized by rank size regularities described by mathematical power laws, and so on. These "factoids," however, were not gathered for the purpose of checking the implications of NEGs' models. Rather, they are the fruits of much prior empirical investigation and subsequent data reduction carried out by another tribe altogether—the plain old geographers (POGs, among whom for these purposes I include the "regional scientists" who flourished during the 1950s and 1960s). Having been left behind by a now-vanished race, these abstract empirical constructions are rather like the imposing monuments of Easter Island. And, to extend the analogy, they have posed an arresting explanatory challenge to the newly arriving NEGs.

A first point worth noting about this situation is that it presents an exception to the classic Duhem-Quine problem, the epistemological problem of circularity that

often plagues empirical validation of scientific hypotheses. This difficulty arises when empirical data are obtained through procedures that leave them laden with (some would say contaminated by) elements of the body of theory that they are being used to test. There is no denying that the stylized facts cited to validate the NEG's models represent a theory-filtered view of the underlying realities. But in this case the filters can be said to have been largely uncontaminated by NEG's theory. This is so because the "stylization" work on these geographic observations was carried out under the influence of an antecedent body of analytical inquiries—namely, the classic location theories of Alfred Weber, August Lösch, and Walter Christaller, and the gravity-potential models of Walter Isard.

That may be read as the bright side of the coin. On the flip side, it could be noted that the NEG's program to date has been occupied primarily with the decanting of old empirical wine into new theoretical bottles, albeit vessels of a shape that is particularly pleasing to modern economists. While this is an acceptable beginning, it should not be the ultimate goal of what has been described as a significant new departure in economic research. New theoretical structures should carry new and otherwise unanticipated empirical implications. Thus one might reasonably ask NEG's to indicate what fresh observational efforts are called for by their theoretical results, and not simply applaud them for achieving close observational equivalence with POG's theories.

A further and related point about NEG's fixation on this set of stylized facts is one that has been raised by some contemporary geographers (see Martin and Sunley 1996 and Clark 1998). Their fear is that the revealed preference of economists for deterministic mathematical models that offer readily characterized, closed-form solutions may channel future empirical work along the same lines as the old; that it reinforces a tendency to suppress details about important inhomogeneities that are present in locational processes. These contemporary geographers worry that the stylishly seductive new economic geography models, adorned with suitably stylized facts, may lead the empirical attention of neophyte geographers away from processes that are harder to analyze in simple formal models, even though those processes are important determinants of the spatial allocation of resources.

There is good cause for concern at least on this latter score. Processes such as those involving the effects of technological knowledge spillovers in the location of innovative activities, and feedbacks responsible for spatial inhomogeneities in the diffusion of technical and organizational practices, are evidently less readily susceptible to simpler sorts of mathematical modeling. On that account, more than any other, they seem to be at risk of dropping out of sight in the new economic geography literature, especially if that literature continues to follow the canon of taste established by Krugman's expositional contributions. Indeed, the parade of alluringly "sparse and bare" models might succeed in turning an entire generation of young geographers away from the empirical tasks of carefully describing the variety and heterogeneity of spatial dynamic processes. That the NEG's stylized facts largely derive from the work of quantitative geographers studying Western Europe and North America heightens my worries on this count. To the extent that new eco-

conomic geography models may be said to be grounded empirically, the regularities in question should not be presumed to hold equally in regard to the experiences of South and Central America, the Indian subcontinent, China, or Africa.

What the economic geography of development therefore would seem to need most at the present juncture is an *increase* in the ratio of data to models, rather than a further reduction. The requirement is not simply for more data, or for more geographically diversified observations, but also for richer, multidimensional data on resource allocation in space and time. For this reason the remainder of my comments seek to encourage empirical as well as analytical efforts to discern and quantify the variety of interdependent processes that shape the geographic dimensions of regional development. Admitting greater complexities into our paradigmatic representations of dynamic geographic phenomena is a way to pose relevant as well as interesting challenges for model builders, and so is likely to result in the more rapid advance of empirical and theoretical knowledge proceeding hand in hand.

Recognizing Complicating Realities

Reductionism has a useful place in scientific analysis, but its place is not everywhere. Economic life is not formed by ordering pairs from Krugman's menu of locational forces—one item from the list of centripetals and the other from the list of centrifugals. A multiplicity of dynamic processes usually are under way concurrently and, that being the reality, one must venture beyond minimal heuristic models in order to identify which combinations are operative where, and how they interact.

Input Market Externalities and Industrial Localization

Although Krugman's models have been directing our attention to localized demand-side externalities, there are externalities operating on the supply side as well. At least since Alfred Marshall, economists have recognized supply-side externalities among the potent forces in the localization of industry; modeling these explicitly involves micro-level specifications of the way that input costs may be affected (see, for example, David and Rosenbloom 1990). Some recent advances along those lines have been obtained through stochastic simulation of Markov random field models. This approach has provided insights into the factors governing the stability of the complex, self-organized spatial patterns of industrial specialization that emerge, for example, when externalities affect the relative recruitment and training costs of workers with different skills in a system of interconnected local labor markets (David, Foray, and Dalle 1998). An analogous spatial process, involving localized knowledge spillovers in the social transactions among research scientists and engineers, has been shown by the use of similar methods to result in the formation and growth of geographic clusters of firms engaged in intensive research and development (Cowan and Cowan 1998).

Externalities affecting business services, and intangible input markets more generally, appear to be comparatively important among the mechanisms underlying the persistence of industrial localization patterns, once the latter begin taking shape.

Unlike the typical income-elastic tangible consumption goods that figure prominently in the demand-side externalities envisaged by Krugman's models, specialized input services that become relatively abundant in some locales are less likely to be rendered widely tradable by the reduction in transport costs applicable to tangible commodities. (Much communication of uncodified technological knowledge appears to be conducted by face-to-face contacts, and thus is not immediately affected by general reductions in telecommunications costs.) For this reason, whereas models like that of Krugman and Venables (1995) exhibit a second bifurcation point under monotone declines in transport costs, urban-industrial agglomerations need not reach the "dispersion to low-wage regions" stage so quickly.

Localization of the Effects of Transport Cost Changes

New economic geography models typically treat transport cost changes as taking effect everywhere to the same degree, whereas in reality they are likely to operate in ways that are far from spatially ubiquitous or geographically neutral. That there are scale economies in transport technologies is recognized by Krugman, who points out that in a dynamic process transport cost reductions can be driven by the growth in trade volumes. Yet the question remains whether such changes yield ubiquitous effects, permitting many sites to benefit concurrently and to the same degree from the expansion of demand. Alternatively, spatial differentials in the growth of shipment volumes might be a potent, self-reinforcing mechanism favoring particular patterns of spatial agglomeration. Under the latter conditions the advantages of initial agglomeration can more readily be retained during the passage from one regime of transport and communications technology to the next.

A variety of dynamic mechanisms may cause asymmetric reductions in production costs at alternative sites, thereby contributing to the "geographically unbalanced" growth of demand. There is, however, a strong bias in such processes toward reinforcing preexisting concentrations in the spatial distribution of economic activity. Because infrastructure facilities are lumpy and capital markets are often imperfect, enhanced transport and communications capacity tends to be provided first where markets are thickest. Once in place, however, such excess capacity in infrastructure tends to induce further localized activity and traffic growth, further entrenching an initial and otherwise possibly evanescent spatial pattern of economic production and distribution. The benefits of lower transactions costs due to "thick markets" externalities, taken in conjunction with the technical indivisibilities of elements of physical infrastructure supporting communications and exchange, constitute a powerful positive feedback mechanism that imparts strong historicity to the spatial evolution of economies.

Is Geography Destiny? Or Is Geography History?

For the reasons just considered—namely, the presence of complementarity and substantial nonconvexities—the detailed timing of innovations and responses is likely

to matter a lot in determining the geography of economic systems. That many dynamic processes in the economy are path dependent, in the sense that their evolution cannot shake free from the influence of their history, is a bothersome complication for many model builders. But it is an aspect of reality that theorists and empirical analysts in this field hardly can afford to ignore.

From the outset of his work in this area Krugman has acknowledged the powerful and persisting effects of historical events on the current industrial organization (see Krugman 1991, especially pp. 98–100 and appendix B). In the current article he reemphasizes the point that positive, self-reinforcing dynamics may give the initial natural geographic endowment of a region leverage over the future course of its spatial development. Consequently, environmental endowments should not be regarded simply as a class of location-determining factors that stands apart from Krugman's list of centripetal, self-reinforcing forces. Instead it is necessary to allow for the possibility—indeed, the probability—that the initial natural endowments of certain sites were subsequently amplified through positive feedback mechanisms, and thereby transformed into a sequence of new and quite different sources of geographic advantage. The resource allocation effects of those inherited circumstances, however, often are too readily attributed to the present direct influences exerted by geographic and climatic conditions (see Gallup and Sachs in this volume). Due to the serial correlation of specific sites' and territories' geographic coordinates, in cross-sectional studies of the spatial distributions of population and production per capita it is no trivial problem to distinguish econometrically between the lagged endogenous effects of past geographic situations and effects that may be presently operative.

The foregoing is a heuristically useful vision of the workings of history in geography, one that clarifies the main point at the expense of considerable oversimplification: the exogenous natural environment, having come first in the sequence, is presumed to be dominant in fixing the early location of production and demand generation. Thereafter, a variety of positive feedback processes are seen as coming into play, causing fluid factors of production to crystallize around those initially selected sites. In this world sunk cost hysteresis is the key to understanding the geography of urban locations. Towns that were founded to exploit the advantages provided by natural harbors, defensible hilltops, or dry cool highlands free of malaria-carrying mosquitoes, survive as cities that go on expanding under the ensuing influence of agglomeration economies long after those original conditions have ceased to be economically significant.

While tales from that tidy, illustrative world provide a salutary corrective for the predisposition of many economists to seek the cause of present conditions exclusively in the present, if not in expectations of the future, something more has to be said. The geographic dimension of economic history is not simply a meditation on the influence of initial conditions; the timing of events along the path of a region's evolution also matters for its spatial development. Quite apart from climatic and demographic shocks, political boundary changes, and other essentially exogenous disturbances, one must allow for the already noted fact that endogenous technological and organizational innovations can have strong non-neutral geographic effects.

Moreover, technological change is subject to its own positive feedback dynamics, and therefore may become localized and channeled in directions that will reinforce some features of the spatial allocation of resources while disturbing others (see Antonelli 1995).

The State in Space

The state's influence on space is treated somewhat skittishly by Krugman's article. His section on policy and primacy recognizes the role of government policy along the two dimensions that matter most within a framework that emphasizes the role of demand-side externalities. Protective tariffs make the location of domestic demand more relevant for manufacturing firms, and the dwelling place of politicians and bureaucrats fixes the location of core home demand, around which agglomerations of producers for the protected market will arise.

Thus both functions of the state are portrayed as contributing to the rise of megacities in the developing world. The implication seems to be that trade liberalization and perhaps smaller government expenditures in the capital city will inhibit the emergence of urban giants, with their attendant costs of congestion and environmental degradation. In evaluating both the diagnosis of giantism and the prescription for its treatment, it would be important to quantify the factors that determine the relative strengths of the two effects. To what length does effective protection of domestic production have to have been pushed? How big do government expenditures have to be relative to regional aggregate demand? What would happen if the state were to subsidize internal transport facilities while protecting domestic industry, as the U.S. government did in the 19th century? Is the role of contemporary politics in urban primacy so potent, or might not the present condition of places like Mexico City be better understood as the outcome of historical processes with deeper roots, and correspondingly greater momentum?

Possibly in acknowledgment that such questions remain unanswered, Krugman sounds a clear warning against drawing policy conclusions from the new economic geography models that he has reviewed here. This commendable caution is understandable in light of the author's allusions to the embarrassing rash of strategic trade and industrial policy recommendations that claimed scientific foundations in the new trade theory. But does this effort go far enough to protect NEGs from being held liable for wrong-headed policy measures affecting the geographic distribution of economic activities? Will having theorists attach to their models a label that says "trying this at home may be dangerous to your country's economic health" be sufficient to block dubious interventionist strategies from gaining powerful political support? History, if nothing else, instructs us otherwise.

Regional Development Strategies

The leading players in the regional industrial development game should not be imagined as gangs of shadowy opportunists, waiting to seize on a fashionable theory as

cover for pursuing their private interests at public expense. Nor are they rent-seeking phalanxes of government bureaucrats in far-off developing countries. In the United States mayors and city councils, as well as governors and legislators of states, have long been active in pursuing local and regional job creation policies. Tax rebates, infrastructure subsidies, and worker training programs—all bought (ultimately) with taxpayers' money—typically are offered to induce “foreign” direct investment in the form of plant relocation. Western European governments also have constructed elaborate systems of regional development assistance, with essentially the same aims in mind (Martin 1993; Collier 1994).

None of these players waited for recommendations of such strategies by POGs and economic experts in regional science, let alone blessings from Krugman and other NEG. It could be said that the concepts of Marshallian externalities and agglomeration economies were influential, as these have been around for a long time. Yet the basic underlying policy idea seems to have been grasped readily enough by “local boosters” in large and small towns without the benefits of its elucidation by academic economists. By comparison, what does call for closer economic analysis is the nature of the outcomes that are likely to ensue when government authorities at a number of locations simultaneously get the same obvious idea. In other words, economics has something worth saying about the downside of “locational tournaments.”

The Economics of Locational Tournaments

The nub of the matter is that rivalries among sites may lead to regional failures to fully exploit the economies of agglomeration. Because accessible product market size and supplies of factors of production are usually limited in the near term, contests among local governments to attract footloose firms have a quasi zero-sum character: given positive feedback forces, getting an early lead over one's urban rivals can cumulate into a sustainable advantage. When the effects of endogenously determined provision of transport and communications facilities are added to other localizing externalities, one locale's lead in attracting industry and trade can impose absolute disadvantages on its near neighbors. Hence such regional contests will appear to each of the players to have a tournament-like payoff structure in which runners-up may get little or nothing while the winner could take all (see David 1984).

With many local government authorities able to enter the game, the inefficiencies that are most likely to emerge are those characteristic of monopolistically competitive equilibria, rather than of the forms of urban giantism on which Krugman's discussion is focused. Leaving distributional issues aside for the moment, the problem is simply that the net site rents that would be created by (successful) agglomeration at a small number of sites may remain less than fully realized. Given the tournament structure of the payoffs, the market for the “development bonds” being issued by the multiplicity of competing localities (to pay for the subsidies offered to firms who relocate) resembles a lottery—since the eventual availability of sufficient tax revenue

to redeem the bonds depends on success in attracting enough business activity to the site. The high risk premiums that would attach to the debt issues would curtail the amount of funds that could be spent up front for local infrastructure or other subsidies. This constraint on the ability of any single local authority to substantially outspend its rivals raises the likelihood that a contest among them ultimately would spread an industry among too many sites, none of which is able to achieve critical density quickly enough to attract the remaining mobile firms of the industry.

This outcome clearly would be socially inefficient, having dissipated at least part of the potential agglomeration economies and, in addition, possibly leaving taxpayers of numerous communities burdened with debt charges incurred for the provision of infrastructure capacity that remained underutilized. There also are some problematic redistribution implications: competitive bidding by local authorities tends to transfer most of the present value of future site rents (net of taxes to amortize the bond issues) to the owners of firms who are in a position to choose among alternative locations and who can make larger, durable capital commitments. Thus regional locational tournaments tend to transfer wealth from small property-owning taxpayers to large national corporations, whereas at the international level the transnational corporations operating in footloose industries are especially well positioned to pocket the rents created by agglomeration.

Recognizing these aspects of the role of the state in space highlights the enormous potential that exists for multiple equilibria in geographic resource allocation, and for outcomes that may be globally inefficient without reaching the excesses of congestion and negative environmental externalities associated with urban giants. Many episodes in the economic history of urban commercial rivalries, stretching back to the mercantilist policies of medieval European towns (see Hecksher 1955, vol. 1), exemplify the perverse effects of locational competition of the former sort. By taking this reality into account we of course will add yet another layer of complexity to the challenges already facing new economic geography model builders. Indeed, it seems that the field still holds quite a bit of unexploited territory for applied game theorists to enter—just in case the ratio of data to models begins to rise into the range above unity!

References

- Antonelli, Cristiano. 1995. *The Economics of Localized Technological Change and Industrial Dynamics*. Dordrecht and Boston: Kluwer Academic Publishers.
- Clark, Gordon. 1998. "Stylized Facts and Close Dialogue: Methodology in Economic Geography." *Annals of the Association of American Geographers* 88 (1): 73–87.
- Collier, John. 1994. "Regional Disparities, the Single Market and the EMU." In John Michie and James Grieve Smith, eds., *Unemployment in Europe*. London: Academic Press.
- Cowan, Robin, and William Cowan. 1998. "On Clustering in the Location of R&D: Statics and Dynamics." *Economics of Innovation and New Technology* 6 (2–3): 202–29.
- David, Paul A. 1984. "The Economics of Locational Tournaments." Silicon Valley Research Project Discussion Paper 1. Stanford University, Center for Economic Policy Research, Palo Alto, Calif.

- David, Paul A., and Joshua L. Rosenbloom. 1990. "Marshallian Factor Market Externalities and the Dynamics of Industrial Localization." *Journal of Urban Economics* 28 (3): 349–70.
- David, Paul A., Dominique Foray, and Jean-Michel Dalle. 1998. "Marshallian Externalities and the Emergence and Spatial Stability of Technological Enclaves." *Economics of Innovation and New Technology* 6 (2–3): 147–82.
- Heckscher, E.F. 1955. *Mercantilism*. 2nd ed. London: George Allen and Unwin.
- Krugman, Paul. 1991. *Geography and Trade*. Cambridge, Mass.: MIT Press.
- Krugman, Paul, and Anthony Venables. 1995. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics* 110: 857–80.
- Martin, Ron L. 1993. "Reviving the Economic Case for Regional Policy." In R.T. Harrison and M. Hart, eds., *Spatial Policy in a Divided Nation*. London: Jessica Kingsley.
- Martin, Ron L., and Peter Sunley. 1996. "Paul Krugman's Geographical Economics and Its Implications for Regional Development Theory: A Critical Assessment." *Economic Geography* 72 (3): 259–92.

Comment on “Geography and Economic Development,” by Paul Krugman

James E. Rauch

In his article Paul Krugman emphasizes the importance for development patterns of transportation costs for goods, both within and between countries. I agree that transportation costs are important for understanding the impact of geography on economic development, but I would like to highlight the importance of transportation costs for people rather than goods. Though I believe these are important both within and between countries, I will focus on the latter.

I also agree with Krugman that differences in initial conditions due to geographic location can set in motion cumulative processes that operate through transportation costs and that greatly magnify the initial differences. Following Arthur Lewis (1978), the initial condition I emphasize is the difference in food productivity between Western Europe (especially Britain) on the one hand and China and India on the other that was in effect propagated throughout the temperate and tropical worlds, respectively, through the mass migrations of the second half of the 19th century. This initial condition interacted—and continues to interact—with the need for manufactures trade and the flow of technological know-how associated with manufactures trade to be facilitated by movement (but not permanent migration) of people between countries.

Trade and Industrialization

Let me begin by summarizing Arthur Lewis’s argument in his book *The Evolution of the International Economic Order* (1978). He notes that (pp. 14–15):

The development of the agricultural countries in the second half of the nineteenth century was promoted by two vast streams of international migration. About fifty million people left Europe for the temperate settlements. ... About the same number—fifty million people—left India and China to work mainly as indentured laborers in the tropics on plantations,

James E. Rauch is professor of economics at the University of California at San Diego.

Annual World Bank Conference on Development Economics 1998
©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

in mines, or in construction projects. ... In Britain, which was the biggest single source of European migration, the yield of wheat by 1900 was 1,600 lbs. per acre, as against the tropical yield of 700 lbs. per acre. The European also had better equipment and cultivated more acres per man, so the yield per man must have been six or seven times larger than in tropical regions.

This difference in the opportunity cost of labor between tropical and temperate countries was reflected in local wages (pp. 15–16): “in the 1880s the wage of a plantation laborer was one shilling a day, but the wage of an unskilled construction worker in Australia was nine shillings a day.”

The cumulative process that this pattern of migration set in motion was the failure of the tropics to industrialize because they lacked domestic mass markets for consumer goods. Lewis (p. 10) recognizes that, “If the domestic market is too small, it is still possible to support an industrial sector by exporting manufactures and importing food and raw materials. But,” he argues, “it is hard to begin industrialization by exporting manufactures. Usually one begins by selling in a familiar and protected home market.”

The word I want to emphasize here is *familiar*. It is one thing to export a homogeneous product like sugar. If traders see that the price differential between domestic and foreign markets is sufficient to cover customs and transport costs, they ship the product. Prices convey all the relevant information. It is another thing to export a differentiated product like garments or shoes. Traders cannot act on the price differential between domestic and foreign shoes because the products bundle together different characteristics. Your shoes may be a poor fit (so to speak) for foreign consumers and thus sell (if they sell at all) for a price well below your expectations. Learning about foreign markets is an expensive and continuous process, because styles and specifications are always changing.

This point is best illustrated by Japan, the only country not populated by Western Europeans to successfully industrialize during this period. Japan solved the problem of breaking into foreign markets by developing general trading companies known as *sogo shosha*. These companies were unique among industrial and developing countries—in both size and scope—until imitators formed in the Republic of Korea in the 1970s and Turkey in the 1980s. In their book on the *sogo shosha*, Yoshino and Lifson (1986, p. 23) write of their operation in the late 19th and early 20th centuries:

Particularly important ... was the role the *sogo shosha* played in providing export opportunities for the myriad small Japanese firms in cottage industries, which, like their counterparts in developing countries today, faced many problems in trying to break into the world market. The *sogo shosha* fed them market information, helped them design products, extended credit, and, most important, developed foreign outlets for their products.

Transportation Costs, Travel, and Technology Transfer

What is the role of transportation costs for people in the argument so far? For permanent migrations, psychic transportation costs were probably at least as important as physical ones in explaining why Western European migrants settled in temperate regions and Chinese and Indian migrants settled in tropical regions. Political barriers to Chinese and Indian immigration in temperate regions were also a factor. Within the tropics, most Chinese migrants settled in Southeast Asia, while Indian migration was more dispersed, presumably due to the integration of India with the British colonial empire. I argue below that the concentration of overseas Chinese in Southeast Asia ultimately set in motion its own cumulative process.

Unlike permanent migration, the movement between countries of people engaged in sales, market research, factory visits, and so on as part of obtaining and renewing contracts for trade in manufactures is repetitive. Physical transportation costs, especially the opportunity cost of time, are thus much more important.

As Krugman points out, transportation costs have fallen considerably over the years, as much for people as for goods. Yet bilateral trade continues to be sensitive to distance. My own calculations using so-called gravity equations show an elasticity of bilateral trade with respect to distance falling from 0.9–1.1 in 1970 to 0.7–0.8 in 1990, still implying that when one doubles the distance between two countries trade between them falls roughly two-fifths, all else being equal (Rauch 1999). There is an interesting lack of difference in these distance elasticities between homogeneous (mainly primary) products and differentiated (mainly manufactured) products, despite the fact that insurance and freight costs of transportation as a percentage of value are only about half as large for differentiated products. My calculations show that if one adjusts for the difference in physical transportation costs, the sensitivity to distance of bilateral trade in differentiated products is greater than for homogeneous products, presumably reflecting the greater need for movement of people to facilitate this trade.

One way to avoid expensive and time-consuming factory visits and market research is to have contacts abroad who know your products and know the markets you want to penetrate. Overseas Chinese are famous for having established such an international information-sharing network. My research with Vitor Trindade shows that, after controlling for all the usual variables (including distance), the product of trading partners' ethnic Chinese population shares has an economically large and statistically significant positive effect on their bilateral trade—much larger for differentiated than for homogeneous products (Rauch and Trindade 1999). The aggregate trade promotion effect of the overseas Chinese network is greatest where there are the most overseas Chinese, as in Southeast Asia. Now that the ethnic Chinese islands of Hong Kong (China), Singapore, and Taiwan (China) have become technologically advanced, the overseas Chinese network has also become an important means of technology transfer. Indeed, manufacturing trade is important not only because it enlarges market size but also because it is associated with a flow of technological know-how.

Economists have typically modeled technology transfer as an arm's-length phenomenon. Firms are not taught the new technology. Rather, they engage in imitative activity on their own (Grossman and Helpman 1991), employ machinery and equipment that embodies foreign knowledge (Coe, Helpman, and Hoffmaister 1997), license the new technology, and so on. However, it is difficult to learn new technology through these mechanisms (Pack and Westphal 1986). There is growing evidence, mostly collected by World Bank researchers, that for developing country firms in particular a major and perhaps predominant source of technology transfer (and transfer of managerial know-how) is industrial country buyers: producers seeking cheaper suppliers of inputs and distributors seeking cheaper suppliers of final goods. For example, Egan and Mody (1992) surveyed U.S. buyers operating in developing countries, including "manufacturers, retailers, importers, buyers' agents, and joint venture partners" (p. 322). They found that:

Buyers also render long-term benefits to suppliers in the form of information on production technology. This occurs principally through various forms of in-plant training. The buyer may send international experts to train local workers and supervisors ... Buyers may also arrange short-term worker training in a developed country plant. (p. 328)

Rhee, Ross-Larson, and Pursell (1984) surveyed Korean exporters of manufactures. Their findings were similar to those of Egan and Mody:

The relations between Korean firms and the foreign buyers went far beyond the negotiation and fulfillment of contracts. Almost half the firms said they had directly benefited from the technical information foreign buyers provided: through visits to their plants by engineers or other technical staff of the foreign buyers, through visits by their engineering staff to the foreign buyers ... (p. 61)

The Rhee, Ross-Larson, and Pursell survey was conducted in 1975. More recently, Korea and other advanced East Asian countries have played the role for developing countries that foreign buyers used to play for them. The role of Korea in developing garment exports from Bangladesh is an especially interesting case that is studied in Rhee and Belot (1990). This case is part of the broader phenomenon of "triangle manufacturing" (Gereffi 1999) in East Asia: economies such as Korea and Taiwan (China) continue to accept and fulfill the orders of industrial country buyers for labor-intensive goods, but have outsourced the actual production to countries with lower wages.

To summarize, exports from developing to industrial countries are often accompanied by transfer of technological know-how from industrial to developing countries through movement of engineers and other skilled personnel. The cost of this movement increases with distance. Distance from the temperate industrial world thus becomes an obstacle for productivity increase in the tropics. The emergence of

technologically advanced economies in East Asia has mitigated this problem for developing countries in their geographic neighborhood.

Policy Implications

Does this analysis of the effect of geography on economic development have any policy implications? We know that many tropical countries have been able to industrialize despite small domestic markets and lack of large nearby ones to which they could export. They industrialized by raising domestic prices for manufactures sufficiently to make industrialization profitable despite these obstacles. But even in these cases, trade is still necessary to facilitate the flow of technological know-how needed to stimulate productivity growth in these industries. The problem of distance and transportation costs for people cannot be avoided.

I find the early Japanese example of how to conquer distance using trading companies suggestive. These companies and, much later, those of the Republic of Korea and Turkey, were initially encouraged by government policy (Rauch 1996). While formation of general trading companies is risky and probably beyond the capacity of most developing countries, smaller, more specialized trading companies could still be helpful.

Here the World Bank has again led the way with research on small trading companies in Hong Kong (Rhee and Soulier 1989). (Hsing 1999 is a more recent, and equally interesting, study of small trading companies in Taiwan, China's fashion shoe industry.) Rhee and Soulier find that export through trading companies is the norm in Hong Kong, whereas in most developing countries even small manufacturers attempt to export directly. Regarding what makes Hong Kong export trading companies effective, they conclude:

Intermediation by export trading companies [ETCs] is needed because of local producers' imperfect information about external markets and foreign buyers' imperfect information about export manufacturers. As highlighted in our Hong Kong survey, the most important resource that ETCs have is their deep knowledge about external markets/buyers and local production capabilities/producers. Without such information, ETCs can hardly be effective in matching potential overseas buyers to local producers. This depth of information must be renewed on a daily basis and be practical and specific, to respond to the ever-changing world market situation. (p. 25)

While the lessons from Hong Kong about what makes a trading company effective are important, the lessons for how to create them are limited. Rhee and Soulier find that 63 percent of the export trading company managers in their sample had experience with other trading companies, reminding us that Hong Kong is special by virtue of its long trading tradition. Without such a tradition, it appears from the experience of other developing countries that we cannot count on the spontaneous

emergence of traders with the requisite deep knowledge of foreign buyers and local capabilities. Whether there is a catalytic role for government or foreign technical assistance to play in their emergence merits further study.

References

- Coe, David T., Elhanan Helpman, and Alexander W. Hoffmaister. 1997. "North-South R&D Spillovers." *Economic Journal* 107 (440): 134–49.
- Egan, Mary Lou, and Ashoka Mody. 1992. "Buyer-Seller Links in Export Development." *World Development* 20 (March): 321–34.
- Gereffi, Gary. 1999. "International Trade and Industrial Upgrading in the Apparel Commodity Chain." *Journal of International Economics* 48 (1).
- Grossman, Gene, and Elhanan Helpman. 1991. "Endogenous Product Cycles." *Economic Journal* 101 (408): 1214–29.
- Hsing, You-tien. 1999. "Trading Companies in Taiwan's Fashion Shoe Networks." *Journal of International Economics* 48 (1).
- Lewis, W. Arthur. 1978. *The Evolution of the International Economic Order*. Princeton, N.J.: Princeton University Press.
- Pack, Howard, and Larry E. Westphal. 1986. "Industrial Strategy and Technological Change: Theory Versus Reality." *Journal of Development Economics* 22 (1): 87–128.
- Rauch, James E. 1996. "Trade and Search: Social Capital, Sogo Shosha, and Spillovers." NBER Working Paper 5618. National Bureau of Economic Research, Cambridge, Mass.
- Rauch, James E. 1999. "Networks Versus Markets in International Trade." *Journal of International Economics* 48 (1).
- Rauch, James E., and Vitor Trindade. 1999. "Ethnic Chinese Networks in International Trade." University of California at San Diego Working Paper, Department of Economics.
- Rhee, Yung Whee, and Therese Belot. 1990. *Export Catalysts in Low-Income Countries: A Review of Eleven Success Stories*. World Bank Discussion Paper 72. Washington, D.C.
- Rhee, Yung Whee, and Christine Soulier. 1989. "Small Trading Companies and a Successful Export Response: Lessons from Hong Kong." Industry and Energy Department Working Paper, Industry Series 16. World Bank, Washington, D.C.
- Rhee, Yung Whee, Bruce Ross-Larson, and Garry Pursell. 1984. *Korea's Competitive Edge: Managing the Entry into World Markets*. Baltimore, Md.: Johns Hopkins University Press.
- Yoshino, M. Y., and Thomas B. Lifson. 1986. *The Invisible Link: Japan's Sogo Shosha and the Organization of Trade*. Cambridge, Mass.: MIT Press.

Floor Discussion of “The Role of Geography in Development,” by Paul Krugman

A participant asked how the Internet has affected the role of geography in development. James E. Rauch (discussant) said that while the Internet can lower initial search costs, people still have to travel to visit factories and close deals. According to a recent article in the *New York Times*, higher airfares have limited the scope of where some businesspeople feel they can travel. But at the same time, researchers have found that business travel increased in the United States during the 1980s despite improvements in communications technology. Thus, Rauch concluded, it seems that the Internet can complement but not replace the need for business travel.

A participant described his recent empirical work linking sector developments and regional analysis. Developments in Chile’s food sector, for example, had very different regional implications depending on the industry or subsector—with effects in the same region ranging from accumulation to marginalization. The participant then asked Paul Krugman (presenter) about the empirical work being done on these conflicting effects.

Krugman said that Gordon Hanson has been doing the most direct applications of such models, actually fitting market potential functions across space, and that the results are impressive. Other researchers are also trying to isolate movements and locations in a relatively model-free way.

Along those lines, and in response to a comment by Paul A. David (discussant), Krugman added that he did not believe the problem with the new economic geography was that it relied on neat models and messy data. To some extent the field has the opposite problem—that is, there are strong empirical regularities in the data that are extremely hard to reproduce in the models, the most famous example being the rank size rule of power law on city sizes. But it is also true that hierarchy models like those developed by Walter Christaller work well in the data and work for a small subspace of the parameters in the theoretical models—and there is something wrong there. Krugman thought that if empirical work is limited, it is not because researchers are narrow-minded and want the world to be neater than it is. It is because researchers are

This session was chaired by Shengman Zhang, managing director and corporate secretary at the World Bank.

Annual World Bank Conference on Development Economics 1998
©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

trying to figure out how to get their models to produce the regularities that are in the real data.

A participant noted that Krugman had offered powerful models on intraspatial and interspatial interactions. But the models appeared to be static in terms of their parameter values—transport costs—and whether there were increasing returns to scale. To be applicable to development theory, the participant felt that new economic geography theory should be integrated with new economic growth theory, so that the speed of change in transport costs could be endogenized.

Krugman replied that a lot of work is being done to integrate geography models with new growth theory models, with the aim of developing models more dynamic than the essentially static models in his article. In terms of transport costs, however, there is no ongoing, real-time change. Krugman said that his models and those of other researchers adjust transport costs a little bit, then let the economy settle into a new equilibrium.

Geography and Economic Development

John Luke Gallup and Jeffrey D. Sachs with Andrew D. Mellinger

Location and climate have large effects on income levels and income growth through their effects on transport costs, disease burdens, and agricultural productivity, among other channels. Geography also seems to affect economic policy choices. Many geographic regions that have not been conducive to modern economic growth have high population densities and are experiencing rapid increases in population. At particular disadvantage are regions located far from coasts and ocean-navigable rivers, for which the transport costs of international trade are high, and tropical regions, which bear a heavy burden of disease. Moreover, a large portion of population growth over the next 30 years is expected to occur in these geographically disadvantaged regions.

Two centuries after the start of modern economic growth, much of the world's population remains mired in poverty. Some benefits of modern development, especially gains in life expectancy and drops in infant mortality, have spread to nearly all parts of the world (though huge discrepancies remain). But in material well-being—as measured by gross domestic product (GDP) per capita adjusted for purchasing power parity (PPP)—the yawning gaps show few signs of narrowing. In 1820 average GDP per capita in Western Europe was about 3 times that in Africa; by 1992 it was more than 13 times as high. More stunning, GDP per capita in Africa in 1992 was \$1,284 (measured in 1990 PPP-adjusted dollars)—the same as in Western Europe in 1820 (Maddison 1995). Asia has experienced significant income growth over the past 30 years, with average incomes rising from \$1,212 in 1965 to \$3,239 in 1992.¹ But in Africa incomes were identical in 1971 and 1992, and in Latin America and the Caribbean average income grew just 7 percent between 1974 and 1992.

John Luke Gallup is research fellow at the Center for International Development, Harvard University. Jeffrey D. Sachs is director of the Center for International Development. Andrew D. Mellinger is research associate at the Center for International Development.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

The Relationship between Geography and Economic Development

Examination of a global map of GDP per capita reveals two striking relationships between geography and economic development (map 1). First, nearly all countries in the geographic tropics (between 23°45' north latitude and 23°45' south latitude) are poor, and almost all countries in the mid- and high latitudes are rich.² Second, coastal countries generally have higher incomes than landlocked countries. Indeed, none of the 29 landlocked countries outside Europe enjoys a high per capita income.

The distribution of population density, measured as people per square kilometer, also reveals important information (map 2).³ First, there is no simple relationship between population density and income level.⁴ Some densely populated areas are rich (Western Europe) while others are poor (China, India, Indonesia). Some sparsely populated areas are rich (Australia, New Zealand) while others are poor (the Sahel of Africa). Across countries there is only a weak positive correlation between population density and GDP per capita. Second, the Eurasian landmass is more densely populated than the rest of the world. (This seems to be a function of human history in addition to underlying geophysical and biogeographical conditions, as we show below.) Third, coastlines and areas connected to the coast by navigable rivers are more densely populated than hinterlands (regions more than 100 kilometers from the coast or a navigable river leading to the ocean).

To elucidate some of the causes of population density, and the subtle relationship between population density and income, we calculate GDP density—GDP per capita times population density—measured as GDP per square kilometer (map 3). In line with maps 1 and 2, the coastal, temperate countries of the Northern Hemisphere have the world's highest GDP densities. Four of these areas—Western Europe, northeast Asia (coastal China, Japan, and the Republic of Korea), and the eastern and western seaboard of the United States and Canada—are the core economic zones of the modern world.⁵ These regions are the overwhelming providers of capital goods in global trade, contain the world's financial centers, and generate a large portion of global production.

The parts of the United States, Western Europe, and northeast Asia that lie within 100 kilometers of the coast contain just 3 percent of the world's inhabited land area and 13 percent of its population—yet they account for at least 32 percent of global GDP (measured in PPP terms).⁶ Excluding coastal China, the core coastal regions contain a mere 9 percent of the world's population but produce at least 30 percent of global GDP. Eleven economies in North America, Western Europe, and East Asia contain 14 percent of the world's population yet are responsible for 88 percent of global exports of capital goods (machinery and transport equipment; WTO 1995).⁷

To take a closer look at these patterns, we examine GDP per capita in the 150 countries with populations greater than 1 million in 1995. Together these countries contained 5.65 billion people in 1995, 99.7 percent of the global total. For purposes of discussion, we define a tropical country as one where at least half the land area falls within the geographic tropics. About 41 percent of the world's population lives in the 72 tropical countries; 59 percent lives in the 78 nontropical coun-

tries. Among the tropical countries the simple (unweighted by country population) average GDP per capita in 1995 was \$3,326; among the nontropical countries it was \$9,027. A simple test of the difference of means across the two groups is significant at the $p < 0.001$ level.⁸

It is convenient to divide the nontropical countries into two groups, temperate-zone and subtropical. In subtropical countries at least half the land area is made up of tropical or subtropical ecological zones but more than half the land area is outside the geographic tropics.⁹ There are 15 subtropical countries, with a mean income of \$7,874, and 63 temperate-zone countries, with a mean income of \$9,302. If we consider only the countries that were not socialist after World War II the geographic divide is even sharper: average per capita income is \$3,685 in nonsocialist tropical, \$9,248 in nonsocialist subtropical, and \$14,828 in nonsocialist temperate-zone countries.

Only 2 of the 30 richest economies (ranked by 1995 PPP-adjusted GDP per capita) are tropical, and both are tiny: Hong Kong (China) and Singapore. Of the 27 others, 4 are subtropical and 23 are temperate. The two tropical economies account for just 1.0 percent of the combined population of the 30 richest economies. Adjusting for the fact that part of the population in 4 of the 30 richest economies—Australia, Chile, Taiwan (China), and the United Arab Emirates—lives in tropical regions increases the tropical share of population in the 30 richest economies to 2.3 percent.

As noted, nearly all the world's landlocked countries are poor; the exceptions in Western and Central Europe are closely integrated with the regional European market and able to trade at low cost.¹⁰ Of the 29 non-European landlocked countries with populations greater than 1 million, the richest is Botswana (38th place), which owes its pride of place to well-managed diamond mines. The second richest is Belarus (68th place). Income differences between landlocked and coastal countries outside Western and Central Europe are striking: the landlocked countries have an average income of \$1,771, while the coastal countries have an average income of \$5,567 ($p = 0.001$). The difference in economic density is even greater, since landlocked countries tend to be sparsely populated (59 people per square kilometer compared with 207 people per square kilometer in coastal countries).

Of course, geography is not everything. Even geographically favored countries such as the temperate-zone, coastal Democratic People's Republic of Korea and the well-situated Czechoslovakia failed to thrive under socialism. Still, development seems to favor temperate-zone countries—especially the 23 nonsocialist, Northern Hemisphere countries that have not been ravaged by war, where average per capita income is \$18,000. Using a multiple regression estimate for the 78 nontropical countries, we find that being located in the Southern Hemisphere reduces average per capita income by \$3,590, being located in the subtropics reduces income by \$4,785, being landlocked reduces income by \$5,190, and being socialist reduces income by \$10,053.

To reveal some of the geographic factors associated with development, we identified major geographic characteristics for selected regions (table 1).¹¹ This analysis

Table 1. Characteristics of Selected Regions

<i>Region</i>	<i>GDP per capita (U.S. dollars)</i>	<i>Population (millions)</i>	<i>Land area (millions of square kilometers)</i>	<i>Land in tropics (percent)</i>
Sub-Saharan Africa	1,865	580	24	91
Western Europe	19,230	383	3	0
East Asia	10,655	1,819	14	30
South Asia	1,471	1,219	4	40
Transition economies ^a	3,902	400	24	0
Latin America and the Caribbean	5,163	472	20	73

a. Eastern Europe and the former Soviet Union

Source: See appendix.

reveals that Sub-Saharan Africa, the poorest region, has several characteristics associated with low income: a high concentration of land in the tropics; a population concentrated in the interior (only 19 percent of Africans live within 100 kilometers of the coast), with more than a quarter living in landlocked countries (the highest of any region); lack of proximity to core markets in Europe; and low population densities in coastal and interior regions. By contrast, Western Europe, the richest region, is nontropical; its population is heavily concentrated near coastal areas, with almost no population in landlocked areas; and population density is moderate. In South Asia and the transition economies of Eastern Europe and the former Soviet Union the population is concentrated in the interior rather than the coast. India's great mass of population, for example, lives in the Gangetic Valley, often hundreds of kilometers from the coast. South Asia is partly tropical and densely populated—indeed, it is the most densely populated region in the world—while the transition economies are nontropical and in the least densely populated region. Latin America is highly tropical and has low population densities and a moderately coastal population. The United States (not included in table 1) has two enormous advantages for development: a large portion of population near the coast (38 percent of Americans live within 100 kilometers of the coast, or 67 percent if river systems leading to the ocean are included) and a temperate-zone landmass.

These patterns prompt several questions. How much has geography mattered for economic growth, once economic policies and institutions are controlled for? If geography mattered in the past, how much does it still matter today? Are there persistent advantages to early developers through agglomeration effects, learning by doing, and the like, or do latecomers have the advantage of the possibility of rapid growth through technological diffusion, capital imports, and other forces of convergence?

We believe that geography—along with economic and political institutions—continues to matter for economic development. Thus geographic considerations should be taken into account in econometric and theoretical studies of cross-country economic growth, which have almost completely neglected geographic themes.¹² Our broad conclusions, described below, can be summarized as follows:

<i>Population within 100 kilometers of coast (percent)</i>	<i>Population within 100 kilometers of coast or ocean-navigable river (percent)</i>	<i>Landlocked population (percent)</i>	<i>Distance to core market (kilometers)</i>	<i>Coastal density (population per square kilometer)</i>	<i>Interior density (population per square kilometer)</i>
19	21	28	6,237	40	22
53	89	4	922	109	125
43	60	0	3,396	381	91
23	41	2	5,744	387	287
9	55	21	2,439	32	16
42	45	3	4,651	52	18

- Relative to temperate regions, tropical regions are hindered in development by higher disease burdens and limitations on agricultural productivity.
- Relative to hinterlands, coastal regions and regions linked to coasts by ocean-navigable waterways are strongly favored in development.
- Landlocked countries may be particularly disadvantaged by their lack of access to the sea, even when they are no farther than the interior parts of coastal countries, because cross-border migration of labor is more difficult than internal migration, infrastructure development across borders is much harder to arrange than similar investments within a country, and coastal countries may have military or economic incentives to impose costs on landlocked countries.
- High population density seems to be favorable for economic development in coastal regions with good access to internal, regional, and international trade. This may reflect increasing returns to scale in infrastructure networks or the enhanced division of labor in settings with high population densities. In hinterlands, by contrast, high population density appears to undermine development.
- Recent population growth is strongly negatively correlated with a country’s potential for economic growth. That is, populations are growing fastest in countries that are least likely to experience rapid economic growth. More generally, there is no strong historical relationship between population growth and a region’s potential for modern economic growth, because population densities seem to have been driven more by agricultural productivity than by conditions for modern industry and services.

It is worth mentioning the relationship between our approach and the recent creative and important work on economic geography by Paul Krugman, Anthony J. Venables, and others (see elsewhere in this volume). The new economic geography follows the new trade theory by showing how increasing returns to scale, agglomeration economies, transport costs, and product differentiation can lead to a highly differentiated spatial organization of economic activity (including cities, hubs and spokes, international division of labor between industry and agriculture, and so on), even when the underlying physical geography is undifferentiated. New economic

geography models illustrate the possibility of self-organizing spatial patterns of production based on agglomeration effects rather than on differences in climate, transport costs, or ecology.

Our starting point, by contrast, is that physical geography is highly differentiated and that these differences have a large effect on economic development. The two approaches can, of course, be complementary: a city might emerge because of cost advantages arising from differentiated geography but continue to thrive because of agglomeration economies even when the cost advantages have disappeared. Empirical work should aim to disentangle the forces of differential geography and self-organizing agglomeration economies.

Geography and Models of Economic Growth

Although econometric studies of cross-country performance have neglected geography over the past decade, economists have long noted the crucial role of geographic factors. Indeed, Adam Smith (1977) gave considerable attention to the geographic correlates of growth.¹³ Smith saw geography as a crucial complement to economic institutions in determining the division of labor. His logic started with the notion that productivity depends on specialization and that specialization depends on the extent of the market. The extent of the market in turn depends on the freedom of markets and the costs of transport. And geography is crucial in transport costs:

As by means of water-carriage a more extensive market is opened to every sort of industry than what land-carriage alone can afford it, so it is upon sea-coast, and along the banks of navigable rivers, that industry of every kind naturally begins to subdivide and improve itself, and it is frequently not till a long time after that those improvements extend themselves to the inland part of the country. (p. 25)

Given the crucial role of transport costs, Smith notes that:

All the inland parts of Africa, and that part of Asia which lies any considerable way north of the Euxine [Black] and Caspian seas, the ancient Sycythia, the modern Tartary and Siberia, seem in all ages of the world to have been in the same barbarous and uncivilized state in which we find them at present. The sea of Tartary is the frozen ocean which admits of no navigation, and though some of the greatest rivers in the world run through that country, they are at too great a distance from one another to carry commerce and communication through the greater part of it. There are in Africa none of those great inlets, such as the Baltic and Adriatic seas in Europe, the Mediterranean and Euxine seas in both Europe and Asia, and the gulphs of Arabia, Persia, India, Bengal, and Siam, in Asia, to carry maritime commerce into the interior parts of that great continent. (p. 25)

Great thinkers such as Fernand Braudel (1972, 1981–84) and William McNeill (1963, 1974) and important historians such as E. L. Jones (1981) and Alfred Crosby (1986) have placed geography and climate at the center of their explanations for Europe's preeminent success in economic development. Braudel points to the key role of Mediterranean and North Atlantic coastal countries as the creative centers of global capitalism after the 15th century. McNeill similarly stresses Europe's great advantages in coastal trade, navigable rivers, temperate climate, and disease patterns as fundamental conditions for its takeoff and eventual domination of the Americas and Australia. Crosby details the advantages of the temperate zones in climate, disease ecology, and agricultural productivity. Two important essays, one by Douglas Lee (1957) and one by Andrew Kamarck (1976), synthesize these arguments in excellent surveys on tropical development—surveys that have been largely ignored by the formal modelers of economic growth.

One of the most interesting recent attempts to ground very long-term development in geographic and ecological considerations comes from ecologist Jared Diamond (1997), who asks why Eurasians (and people of Eurasian origin in the Americas and Australia) “dominate the modern world in wealth and development” (p. 15). He disposes of racial explanations not just on moral grounds but also on rigorous findings of the shared genetic inheritance of all human societies. His explanation rests instead on the long-term advantages of Eurasia in agglomeration economies and diffusion of technologies. Populations in the Americas and Australia were cut off by oceans from the vast majority of populations in Eurasia and Africa. Thus they could not share, through trade and diffusion, in technological advances in agriculture, communications, transport, and the like.

Diamond also argues that because plant species and domesticated animals appropriate to one ecological zone may be inappropriate elsewhere, technological diffusion works most effectively within ecological zones, and therefore in an east-west direction along a common latitude rather than in a north-south direction. Eurasia, he claims, enjoyed the benefit of its vast east-west axis heavily situated in temperate ecological zones, while Africa was disadvantaged by its north-south axis, which cuts across the Mediterranean climate in the north, the Saharan Desert, the equatorial tropics, and the southernmost subtropical regions. Diamond argues that these advantages, in addition to more contingent (that is, accidental) advantages in indigenous plant and animal species, gave Eurasia a fundamental long-term advantage over the rest of the world.

Historians have also stressed the changing nature of geographic advantage as technology changes. In early civilizations, when transport and communications were too costly to support much interregional and international trade (and any oceanic trade), geographic advantage came from agricultural productivity rather than from access to markets. As a result early civilizations almost invariably emerged in highly fertile river valleys such as those around the Nile, Indus, Tigris, Euphrates, Yellow, and Yangtze Rivers. These civilizations produced high-density populations that in later eras were disadvantaged by their remoteness from international trade. Northern Europe could not be densely settled before the discoveries of appropriate technologies and tools to fell its great forests (Landes 1998).

Similarly, as the advantages of overland and coastal-based trade between Europe and Asia gave way to oceanic commerce in the 16th century, economic advantage shifted from the Middle East and eastern Mediterranean to the North Atlantic. In the 19th century the high costs of transporting coal for steam power meant that industrialization almost invariably depended on proximity to coal fields. This advantage disappeared with the discovery of petroleum refining, the production of oil- and hydro-based electricity, and the reduced cost of bulk transport. Railroads, automobiles, air transport, and telecommunications have reduced the advantages of coastlines relative to hinterlands, but according to the evidence below, the advantages of sea-based trade remain.

Formal Models of Geography and Development

To establish some formal ideas about the interaction between geography and development, we start with the simplest model of economic growth, the AK model (known in its earlier incarnation as the Harrod-Domar model), and add transport costs. In the resulting model, growth differences across countries depend on several parameters that we find to be important in later empirical analysis. These factors include underlying total factor productivity, denoted by A , which may differ across countries because of fundamental geographic reasons (such as differences in productivity between temperate and tropical agriculture and differences in endemic health conditions among ecozones); transport costs, reflecting both distances and physical access to trade (navigability of rivers, distance from the coast); and national saving rates and, implicitly, government economic policies.

Suppose that an economy has the aggregate production function

$$Q = AK.$$

The capital stock (K) evolves according to

$$\frac{dK}{dt} = I - \delta K$$

where I is investment and δ is the rate of depreciation.

We assume for the moment that population is constant and normalized to 1, so that Q represents both output and output per capita; later we examine population growth.

The national savings rate is fixed at s (the alternative assumption of intertemporal optimization would be straightforward as well, but with little gain in realism or insight). The price of investment goods relative to final output is P_I . Thus

$$sQ = P_I I.$$

The growth rate of the economy is then

$$\gamma = \frac{1}{Q} \frac{dQ}{dt} = \frac{1}{K} \frac{dK}{dt} = \frac{sA}{P_I} - \delta.$$

Economic growth is positively dependent on the savings rate (s) and the level of productivity (A) and negatively dependent on the relative price of capital goods (P_I) and the rate of depreciation (δ).

Transport costs affect the relative price of capital goods because some investment goods must be imported. In many developing countries virtually all equipment investment is imported. To illustrate some implications of transport costs, we assume that each country produces a distinct final good and that investment is a composite of the final goods produced in various countries. The key assumption is that there are gains from trade, so that transport costs and other barriers to trade reduce growth. We do not directly model the underlying reasons for specialization in production and hence gains from trade. As is well known, specialization in production may result from different primary factor endowments, economies of scale in production, economies of specialization through learning by doing, or different technologies across countries because of investments in proprietary research and development.

Total investment depends on investment expenditure on domestic goods (I^d) and on imported investment goods (I^m):

$$I = I(I^d, I^m).$$

As a simple illustration, consider the case in which I is a Cobb-Douglas function of the underlying domestic and foreign investment goods:

$$I = (I^d)^a (I^m)^{(1-a)}.$$

The true price index of investment goods, then, is a geometric average of the price of domestic and foreign investment goods. Setting the domestic good as numeraire (that is, setting the price of I^d equal to 1), we get:

$$(2) \quad P_I = \alpha (P^m)^{(1-a)}$$

where $\alpha = a^a (1-a)^{(1-a)}$.

We denote the (exogenous) world market price of the imported good as P^{m*} and write the landed (or c.i.f.) price in the home economy as $P^m = \tau P^{m*}$, where $\tau > 1$ is the c.i.f. factor (the world price plus cost, insurance, and freight). Then from equations 1 and 2 we have a modified equation for the growth of the economy:¹⁴

$$(3) \quad \gamma = \left(\frac{sA}{\alpha} \right) \left(P^{m*} \right)^{-(1-a)} \tau^{-(1-a)} - \delta.$$

The growth rate is now inversely related to the cost of transport (τ). In this model transport costs reduce growth by raising the cost of the imported capital good. Earlier empirical studies of growth (Barro 1991, for example) have shown that the rate of growth is a decreasing function of the relative cost of investment goods. This is essentially the channel by which the costs of transport and distance enter into equation 3.

Equation 3 suggests three important points. First, underlying total factor productivity affects growth rates. Second, transport costs affect growth rates. These costs are likely to depend on several characteristics. Coastal economies will generally have much lower transport costs than hinterland economies. Countries near core economies (the main providers of capital goods) will generally have lower transport costs than distant countries, so growth is likely to diminish in direct proportion to distance from the core. Third, protectionist policies that raise the domestic price of imported capital goods, or that limit the exports needed to import capital goods, are likely to reduce long-term growth. This point cannot be overstated: countries require imports of capital goods for long-term growth.

A Model with Intermediate Goods

Suppose now that final production requires imported intermediate inputs. This assumption is of enormous empirical importance because many of the key manufactured exports of developing countries involve imported intermediate manufactured goods (fabrics, electronic components), which are then assembled domestically with low-cost labor and re-exported to world markets. The transport costs involved in the import of intermediate products and their re-export after domestic processing can affect the success or failure of manufacturing exports—even if the transport costs for investment goods are minimal.

We must now distinguish between gross output (Q) and gross domestic product (Y). In particular, we set

$$Q = \min(AK, N/\mu)$$

where N is the intermediate good imported from abroad. The final good in the home market continues to be the numeraire (with price 1), and the relative price of the imported intermediate good is $P_n = \tau P_n^*$. The gross domestic product in units of the final good is given by

$$Y = AK - P_n N = AK - \mu P_n AK$$

or

$$(4) \quad Y = (1 - \mu P_n)AK.$$

Because domestic final output in the home market is the numeraire, its price in the foreign market inclusive of transport costs is τ . Similarly, if the intermediate product

sells for P_n in the home market, its price in the foreign market is P_n/τ . Suppose that at world prices (that is, in the foreign market) the share of the intermediate good in final output is given by $\sigma = (P_n/\tau)N/(\tau Q)$. Then equation 4 can be rewritten as:

$$(4') \quad Y = (1 - \sigma\tau^2)AK.$$

All of the model goes through as before, so that the modified growth equation is now

$$(5) \quad \gamma = \left(\frac{sA}{\alpha} \right) \left(1 - \sigma\tau^2 \right) \left(P^m \right)^{-(1-a)} - \delta.$$

The key point here is that relatively small transport costs can have huge effects on output and growth when the share of intermediate inputs in final demand is large. For example, suppose that $\sigma = 0.7$. Now compare the growth rates of two countries, one with one-way transport costs equal to 5 percent of the value of gross output and one with those costs equal to 10 percent of the value of gross output. Ignore, for the moment, the transport costs for capital goods to focus solely on the effect of intermediate products. Let γ_1 be the growth rate of the low-transport-cost economy, and let γ_2 be the growth rate of the high-transport-cost economy. Then, from equation 5:

$$\gamma_1 / \gamma_2 = [1 - 0.7(1.1025)] / [1 - 0.7(1.21)] = 1.49.$$

The growth rate is 49 percent higher in the low-transport-cost economy than in the high-transport-cost economy even though the transport costs differ by only 5 percentage points. The explanation, of course, is that a “mere” 5 percentage point decrease in one-way transport costs for intermediate and final goods implies a whopping 49 percent increase in domestic value added.

The notion that intermediate inputs represent such a large portion of the value of gross output may seem unrealistic, but such is the case for many key export sectors in developing countries. In many labor-intensive industries—such as apparel and electronics assembly—the developing country imports a large portion of the value of final output. The intermediate imports are assembled by domestic workers and then re-exported to world markets. Thus the developing country is essentially selling labor services used in assembly operations rather than selling the entire product. For such assembly industries, even small increases in transport costs can render the sector noncompetitive. For this reason only developing countries with good transport access to world markets have been able to establish assembly-type industries (Radelet and Sachs 1998).

These arguments further underscore the disadvantages of the hinterland relative to the coast in economic development. Almost all modern production depends on multi-stage processing of output, with inputs often produced in many specialized enterprises, some foreign and some domestic. The low-cost transport of such intermediate prod-

ucts is crucial, especially in developing countries, where many intermediate components are imported. Only coastal areas or areas linked to the coast through navigable waterways or low-cost land transport have a chance to compete in such activities.

Divergence, Convergence, and Poverty Traps

The central feature of the AK model is the absence of convergence. Because there are no diminishing returns to investment in production, there is no tendency for growth to slow as capital deepening occurs. For this reason countries that have underlying advantages in savings rates, efficiency, transport costs, or depreciation rates will display permanently higher growth rates. Moreover, the gap between such countries and slower-growing countries will widen.

Models like the AK model also highlight another possibility. Suppose that transport costs are important in determining economic activity at one stage of history but then become less important. In the AK model the early advantage would boost economic activity, causing the favored region to jump ahead of others. Once the advantage was lost, all countries would grow at the same rate (assuming the same A , s , and δ). The early advantage would never be lost in terms of relative income levels, though growth rates would converge. In models with increasing returns to scale rather than the constant returns to scale of the AK model, the early advantage could lead to persistently higher growth rates even if it disappeared, because growth rates could be a positive function of the level of capital, which would be raised by the transitory advantage. Thus one possible interpretation of the observation that temperate-zone, coastal countries have the highest GDP is that such geographic attributes once conferred advantages, even if they no longer do so.

If the model is recast as a neoclassical growth model with diminishing marginal productivity of capital, so that $Q = AK^\beta$, with $\beta < 1$, then the conclusions reached so far have to be recast as follows. The same list of parameters (s , A , P^{m*} , τ , δ) now affect the steady-state level of Q , denoted Q^{ss} , and the steady-state capital stock, K^{ss} , but not the long-term growth rate. Because the capital stock converges gradually to its steady state, so too does the level of output. In this case the growth equation can be written as follows:

$$(6) \quad \gamma_t = \frac{1}{Q_t} \frac{dQ_t}{dt} = \lambda \left(\ln Q_t^{ss} - \ln Q_t \right).$$

Equation 6 holds that the proportionate rate of growth depends on the gap between the steady-state level of output and the contemporaneous level of output. We could, in general, derive $\ln Q_t^{ss}$ to be a function of the underlying parameters. Approximating this relationship in log-linear form, $\ln Q_t^{ss} = \beta' Z_t$, where Z_t is the vector of underlying growth-influencing parameters and β is a vector of coefficients, we end up with an empirically estimable equation that has been extremely popular in recent years:

$$\gamma_i = \lambda\beta'Z_i - \lambda\ln Q_i.$$

In this formulation growth depends positively on the parameters in question and negatively on the initial income variable. The empirical presence of the term $\lambda\ln Q_i$ has been used to examine whether there is a tendency toward convergence, as in the neoclassical model, or continuing divergence, as in the AK model, in which the level of income is not a determinant of the rate of growth.

The issue of convergence or nonconvergence depends heavily on the underlying structure of production. In an environment of increasing returns to scale at the firm level as well as gains from a greater diversity of products—as in popular models of Dixit-Stiglitz imperfect competition (Romer 1986, 1990; Grossman and Helpman 1991)—there may well be increasing or at least constant returns to the capital stock at the macroeconomic level. The marginal productivity of capital would then be constant (as in the AK model) or even increasing, as the aggregate capital stock increases. In that setting the AK model would depict the aggregate production technology better than the neoclassical model, with its assumption of declining marginal productivity of capital. Thus to the extent that scale economies and product diversity are critical, we would expect to see little convergence between rich and poor countries, and we could well see divergence. To the extent that economies of scale and product diversity are limited, we would be more likely to see convergence in income levels, controlling for other factors.

Geography and Population Dynamics

Because it is not easy to integrate population dynamics with the AK model in a meaningful way, we have to step outside that simple framework to discuss some aspects of the relationship among geography, population dynamics, and growth. We have stressed repeatedly the advantages of coastal areas for economic development. But we have not said anything about the distribution of populations across regions. In fact, the linkages are problematic for three reasons. First, vast numbers of people live in areas that are not suited to modern economic growth. Over the course of history population densities have tended to rise in areas that are conducive to growth, so coastal areas are indeed more densely populated than hinterlands. But population densities have also risen in fertile agricultural areas—for example, along inland river systems such as the Ganges, Tigris, Euphrates, and Nile—that are useful for irrigation and inland trade but not international trade. As a result population densities are high in areas that depend on subsistence agriculture and are not well suited to modern economic growth.

Second, since 1950 population growth has tended to be highest in poor, remote regions, mainly because population growth is negatively correlated with per capita income and highly negatively correlated with mothers' education and the market value of mothers' time.¹⁵ Thus the concentration of populations in problematic regions is growing—exacerbated by the tendency for improvements in public health to spread more readily than economic growth.

Third, the mismatch between economic growth and population trends has caused mass migration from hinterlands to coasts. Most migratory movements are within poor countries, leading to unprecedented inflows of people into urban areas and to the rise of megacities in developing countries. The next largest migration most likely occurs across borders of developing countries. This type of migration includes vast flows of population from landlocked to coastal countries. The third largest migration is from poor to richer countries. This migration would, of course, be vastly larger were it not for immigration controls in the richer countries. In any case the pressures for both internal and international migration will rise sharply in the coming decades as differences in income levels continue to increase.

The effects of population pressures on economic growth are likely to differ markedly between the hinterland and the coast. In the hinterland transport costs are extremely high, the division of labor is low, and output is generally characterized by decreasing returns to scale in labor in the face of limited supplies of land. Thus higher population densities will be associated with falling output per capita, a tendency seen in many African countries in the past 20 years. But on the coast, where transport costs are low and the division of labor is high, a rising population may be associated with stable or even increasing per capita incomes, even when the capital-labor ratio declines. This is because higher population densities make possible an increasingly refined division of labor.¹⁶

Thus economies are likely to bifurcate on two pathways. The hinterland will experience decreasing returns to scale in labor and high rates of population growth. The coast will see increasing returns to scale in labor and falling rates of population growth as household incomes rise. The hinterland may therefore exhibit Malthusian dynamics, while the coast shows rising incomes and falling natural population growth rates. The two systems will interact through ever-greater pressures on migration from the hinterland to the coast.

Geography and Economic Policy Choices

So far we have stressed that geography may influence growth directly through the level of productivity and transport costs. Geography can have another potent effect through its influence on the choice of economic policies. Countries that are close to markets, for example, may choose more open trade policies than countries that are far from markets. We offer a motivation for such a possibility.

Suppose that growth is given by sA/P_I and that A can be decomposed into a multiplicative policy component, π , and a purely exogenous component, θ : $s\pi\theta/P_I$. Suppose as well that $P_I = P_I(\tau)$ —that is, the price of investment goods is an increasing function of transport costs. At this abstract level we can say that the policy component of growth is a decreasing function of the ad valorem tax rate levied by the government on the private economy, $\pi = \pi(T)$. For simplicity we will use a linear functional form, $\pi = c - eT$. These taxes might be formal taxes, bribes demanded to clear customs, seizures of property, and so on. The basic idea is that government gains revenues but at the expense of a worsening policy environment.

Suppose that T is chosen once and for all to maximize the expected utility of government officials. To keep matters simple in this abstract framework, we assume that the policymaker has an intertemporal log utility function, a pure discount rate of d , and a hazard rate h of losing office.¹⁷ Expected utility is then given by

$$(7) \quad EU = \int e^{-(d+h)t} \log[TQ(t)] dt.$$

Equation 7 is maximized subject to the constraint that $Q(t) = Q_0 e^{\gamma t}$, where $\gamma = s\pi(T)\theta/P_I(\tau)$. What we have, essentially, is an optimal tax calculation on the part of the sovereign. Higher taxation yields more immediate revenue, but at the cost of slower future growth and hence lower future revenues. With simple manipulations equation 7 can be rewritten as

$$EU = \left[\frac{\log T + \log Q_0}{d+h} \right] + \left[\frac{s\pi(T)\theta / P_I(\tau)}{(d+h)^2} \right]$$

It then remains to calculate the optimal T by setting $dEU/dT = 0$. We find that

$$T = P_I(\tau) \left(\frac{d+h}{s\theta e} \right) \quad T \leq 1.$$

This is an insightful expression. The optimal tax is an increasing function of transport costs, the discount rate, and the probability of losing office and a decreasing function of total factor productivity and the responsiveness of growth to the tax rate. Basically, the sovereign is trading quick gain for future loss. To the extent that growth is low (because P_I is high or θ is low) or the future is heavily discounted (because $d+h$ is high) or unresponsive to taxation, the tax rate (T) should be set at a high level. To the extent that underlying growth is rapid or highly responsive to taxation, the tax rate should be set lower.

There are at least two implications for geography. First, good policy and good geography may tend to go together. When growth is inherently low because of adverse geographic factors and also unresponsive to policy (perhaps for the same reasons), the revenue-maximizing sovereign will impose high rates of taxation (such as protectionist policies). When the economy is inherently productive and responsive to good economic policies, the sovereign will have an incentive to impose low rates of taxation. The result is that natural differences in growth potential tend to be amplified by the choice of economic policies.

Second, the correlation between favorable underlying growth conditions and good policies leads to an important identification problem in estimating the effects of economic policy on economic performance. Suppose that a regression estimation

of growth on taxation reveals a strong negative relationship. This is usually interpreted as a demonstration that policy matters for growth. We have just seen, however, that it might also reflect the fact that growth matters for policy. It is crucial to specify structural growth relationships that include both policy and underlying geography in order to disentangle these alternatives.

Empirical Linkages between Geography and Growth

The basic theory points to two geographic factors of deep significance for growth: transport costs, measured by the parameter τ , and intrinsic productivity, measured by A . Consider transport costs first. Remarkably, despite the likely importance of transport costs for economic growth, there are no adequate measures of transport costs for a large sample of industrial and developing countries. The best that we could obtain for a large number of countries are International Monetary Fund (IMF) estimates of c.i.f./f.o.b. margins in international trade. These margins measure the ratio of import costs inclusive of insurance and freight (c.i.f.) to import costs exclusive of insurance and freight (f.o.b.). There are several problems with these measures, the most important being that they are only crudely estimated by the IMF and they depend on the composition of imports and so are not standardized across countries.

Still, the c.i.f./f.o.b. margins are informative and predictive of economic growth. Using c.i.f./f.o.b. margins for 1995 (3.6 percent for the United States, 4.9 percent for Western Europe, 9.8 percent for East Asia, 10.6 percent for Latin America, and a whopping 19.5 percent for Sub-Saharan Africa), we estimate an equation relating the margin to the country's distance from the core areas of the world economy and its accessibility to sea-based trade by including a dummy variable for non-European landlocked countries:¹⁸

$$\begin{aligned} \text{c.i.f./f.o.b.} &= 1.06 + 0.010 \text{ Distance (1,000 kilometers)} + 0.11 \text{ Landlocked} \\ &\quad (84.9) \quad (3.0) \quad (2.4) \\ n &= 83, R^2 = 0.32. \end{aligned}$$

As expected, there is a penalty both for distance from core economies and for being landlocked. Each 1,000 kilometers from the core raises the c.i.f./f.o.b. margin by 1.0 percentage points; being landlocked raises the c.i.f./f.o.b. margin by 11.1 percentage points. We show later that the c.i.f./f.o.b. margin is indeed predictive of income levels and economic growth. Interestingly, regressing the c.i.f./f.o.b. margin on the share of the population within 100 kilometers of the coast (Pop100km) yields a negative effect, but when both Pop100km and Landlocked are included, Landlocked has more explanatory power. The coefficient on Pop100km remains negative, as expected, but it drops in magnitude and is statistically insignificant.

We have seen that various regions differ markedly in terms of the locations of their populations relative to coasts and navigable rivers (see table 1). Africans tend

to live far from the coast, while Europeans are overwhelmingly coastal. The notion that coastal access has a large effect on trade and growth is plausible given what we know about the growth patterns of the most successful developing countries since 1965. Almost without exception rapidly growing developing countries have based their growth on labor-intensive manufacturing exports. And almost without exception such activities have expanded in port cities or export zones close to ports. Almost all countries that have achieved macroeconomic success in labor-intensive manufacturing exports have populations almost entirely within 100 kilometers of the coast (Radelet and Sachs 1998).

The geographic determinants of the efficiency parameter, A , are potentially much more varied. First, we can surmise that coastal access will matter for internal trade and productivity as well as for international trade. Cities are engines of growth, and most large cities (other than garrison towns and administrative capitals) develop on coasts or on rivers that lead to the ocean. Thus countries with neither coastlines nor ocean-navigable rivers tend to have less urbanization and less growth. A simple regression estimate for 149 countries shows that in 1995 ocean-accessible regions were indeed more urbanized, as were countries closer to economic core regions (LDistance). A simple regression reveals that

$$\text{Urbanization} = 132.3 + 17.1 \text{ Pop100km} - 10.8 \text{ LDistance}$$

(10.8) (3.6) (7.1)

$n = 149, R^2 = 0.29.$

A second major dimension of productivity linked to geography is the prevalence of infectious disease. Almost all of the 200–500 million annual cases of malaria occur in the tropics (map 4; WHO 1997b).¹⁹ This pattern is neither accidental nor mainly a result of reverse causation, in which poor countries are unable to eradicate a disease that is under control in rich countries. There is no effective prophylaxis or vector control for malaria in areas of high endemicity, especially Sub-Saharan Africa. Earlier methods of vector control are losing their effectiveness because of the increased resistance of mosquitoes to insecticides. Standard treatments are also losing effectiveness because of the spread of resistance to cloroquine and other anti-malarial drugs.

The geographic extent of malaria is determined mostly by the ecology of the parasites (different species of malaria *Plasmodia*) and the vectors (different species of *Anopheles* mosquitoes). In temperate-zone and subtropical environments, where the foothold of the disease (both in terms of the mosquito population and the parasite endemicity) was more fragile, malaria has largely been under control since 1946. Since then endemic malaria has retreated to the core tropics (map 5).

Using data from the World Health Organization, we construct a measure of malaria intensity (see map 4).²⁰ A simple regression of malaria intensity on ecological zones shows that it is most intense in the tropics and somewhat less intense in the subtropics (relative to all other ecozones). There is also a strong Sub-Saharan Africa effect.

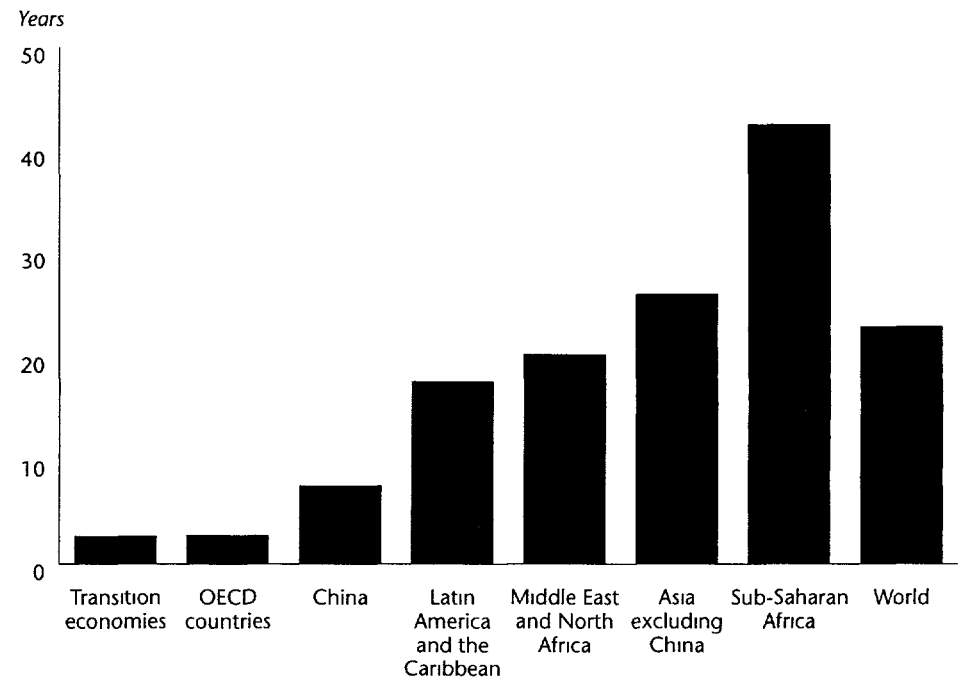
$$\begin{aligned}
 \text{Malarial intensity (scale 0-1)} = & -0.01 + 0.3 \text{ Wet tropics} + 0.5 \text{ Dry tropics} \\
 & (0.7) \quad (2.0) \quad (4.9) \\
 & + 0.4 \text{ Wet subtropics} + 0.2 \text{ Dry subtropics} \\
 & (3.9) \quad (1.5) \\
 & + 0.5 \text{ Sub-Saharan Africa} \\
 & (6.7)
 \end{aligned}$$

$$n = 148, R^2 = 0.74.$$

The pattern for malaria is a common one, fitting a range of infectious diseases whose vectors of transmission depend on the tropical climate. Many diseases that are carried by mosquitoes (dengue fever, yellow fever, and lymphatic filariasis, in addition to malaria), mollusks (schistosomiasis), and other arthropods (onchocerciasis, leishmaniasis, trypanosomiasis, Chagas' disease, visceral filariasis) are endemic in tropical ecological zones and nearly absent elsewhere. While data on disease burdens by country are generally not available, the recent massive study by Murray and Lopez (1996) on the burden of disease confirms that the heavy tropical concentration of infectious disease is an important cause of death (figure 1).

A third major correlate of geography and productivity is the link between climate and agricultural output. Estimates of agricultural productivity suggest a strong adverse effect of tropical ecozones on the market value of agricultural output after controlling

Figure 1. Infectious and Parasitic Disease as a Share of Disability-Adjusted Life-Years Lost to All Causes, 1990



Source: Murray and Lopez 1996

for inputs such as labor, tractors, fertilizer, irrigation, and other inputs. Tropical agriculture suffers a productivity decrement of 30–50 percent relative to temperate-zone agriculture after controlling as best as possible for factor inputs (Gallup 1998).

Other more prosaic geographic correlates, such as endowments of high-value natural resources (hydrocarbons, minerals, precious gems), also affect cross-country per capita income at a point in time. Lacking comprehensive measures of the international market value of such resource endowments, we settle for a rough measure of one key resource: deposits of petroleum and natural gas. Countries rich in hydrocarbon deposits per capita indeed display higher per capita income in 1995, though not necessarily higher economic growth. Indeed, Sachs and Warner (1995b) suggest that higher natural resource exports in 1970 (measured as a percentage of GDP) are negatively related to subsequent growth.

Geography and Levels of Per Capita Income

We examine the linkage between output and geography both in levels and rates of change of GDP. Suppose that countries differ in their growth rates according to a vector of characteristics, \mathbf{Z} , that includes determinants of transport costs and total factor productivity. We write a linear approximation of a growth equation similar to equation 6 such that

$$\gamma_{it} = \beta' \mathbf{Z}_i + \mu_{it}.$$

In any period $T > 0$, $Q_{it} = \exp(\gamma_i T) Q_{i0}$. Suppose that in the distant past Q_{i0} was randomly distributed across countries, with $\ln Q_{i0} = \ln Q_0 + \xi_i$, and with ξ_i independent of \mathbf{Z}_i . We can then estimate a cross-country level equation of the form:²¹

$$\ln(Q_{iT}) = \ln(Q_0) + T\beta' \mathbf{Z}_i + \varepsilon_i$$

Note that the effects of the parameters \mathbf{Z}_i on the level of Q_{iT} will tend to grow over time, relative to T , because \mathbf{Z} affects the growth rate, not merely the relative income level, of country i . Note also that if the \mathbf{Z} variables are time dependent, then Q_{iT} is a function of the entire time path of \mathbf{Z}_i . In general, for most variables of interest, we have snapshots of \mathbf{Z} rather than a time series. One objective of empirical development studies should be to create time series measures of key institutional determinants of growth (such as openness of markets and protection of property rights) to strengthen our empirical tests.

We start with the simplest specification, writing the log level of per capita income as a function of three underlying geographic variables: *Tropical*, the percentage of land in the geographic tropics; *Pop100km*, the portion of the population within 100 kilometers of the coast; and *LDistance*, the minimum log distance of the country to one of the three core regions, measured as the minimum log distance to New York, Rotterdam, or Tokyo. We estimate this relationship three times: for Maddison's (1995) GDP estimates for 1950 and 1990 and for the World Bank's

(1998) PPP GDP estimates for 1995 on the subset of countries for which Maddison's data are available (table 2). In all three regressions output is a positive function of Pop100km and a negative function of Tropicar and LDistance. As expected, the magnitude of the effects tends to increase over time. In 1950 the "penalty" for Tropicar was -0.69 , indicating that per capita income in tropical areas was 50 percent ($= \exp[-0.69]$) of that in nontropical areas, controlling for the other factors. By 1995 the effect had risen to -0.99 (or income of 37 percent of that in nontropical areas). Similarly, the benefit of a coastal population rose from 0.73 in 1950 to 1.17 in 1995. The implication is that being tropical, landlocked, and distant was already bad in 1950 and that it adversely affected growth between 1950 and 1995.

For the 1995 data we have a wider range of possible explanatory variables. We start by estimating the simple level equation for GDP per capita on a PPP basis in 1995 for the 150 countries with populations greater than 1 million. We then turn to growth equations for 1965–90. We group the explanatory variables (Z) into three broad cat-

Table 2. Relationship between GDP per Capita and Selected Variables, 1950, 1990, and 1995

Variable	(1) <i>lgdp50</i>	(2) <i>lgdp90</i>	(3) <i>lgdp95</i>	(4) <i>lgdp95</i>	(5) <i>lgdp95</i> <i>non-Africa</i>	(6) <i>lgdp95</i>
Tropicar (percent)	-0.69 (4.13)	-0.99 (5.78)	-0.99 (5.10)			
Pop100km (percent)	0.71 (4.02)	1.00 (5.43)	1.09 (5.27)	0.85 (3.63)	1.21 (4.17)	0.36 (2.53)
LDistance	-0.22 (2.56)	-0.39 (4.39)	-0.34 (3.41)			0.03 (0.55)
C.i.f./f.o.b. transport cost margin				-2.28 (2.32)	-13.50 (4.66)	
Malaria index 1994 (0–1)				-1.55 (6.60)	-1.26 (2.69)	-1.15 (7.65)
Hydrocarbons per capita				0.01 (1.84)	0.01 (1.75)	0.01 (1.85)
Socialism						-0.05 (0.31)
New state (0–3)						-0.06 (0.98)
Openness (0–1)						0.23 (7.38)
Public institutions (0–10)						0.55 (3.17)
Constant	9.07 (13.58)	11.19 (16.26)	10.98 (14.10)	10.84 (9.82)	22.64 (7.42)	6.71 (11.06)
Number of observations	129	129	129	83	52	97
R ²	0.38	0.58	0.50	0.69	0.56	0.88

Note: Numbers in parentheses are the absolute values of robust t-statistics
Source: See appendix.

egories: variables related to transport costs and proximity to markets, variables related to ecological zone, and variables related to economic and political institutions.

In regression 4 of table 2 we limit Z to a parsimonious set of four variables closely linked to geography: Pop100km, transport costs as measured by the c.i.f./f.o.b. margin, the prevalence of malaria, and the endowment of hydrocarbons per capita. These four variables alone account for 69 percent of the cross-country variation in per capita income, and all carry the expected sign and are statistically significant (although the endowment of hydrocarbons is significant only at the 10 percent level). High levels of GDP per capita are associated with low transport costs, a coastal population, a large endowment of hydrocarbons per capita, and the absence of malaria.

The strong correlation of malaria with income levels is not simply an “Africa proxy.” When we rerun the equation for the sample of countries outside Sub-Saharan Africa (regression 5 in table 2), we find similar results. When we enter both malaria and Tropicar into the regression (results not shown), the effect of the malaria variable is far more important, suggesting that the negative effect of the Tropicar variable is largely subsumed by the geographic distribution of malaria. Of course, these associations are hardly proof of causality. Not only might the explanatory variables be proxies for left-out variables—for example, malaria could be a proxy for a range of tropical diseases or other liabilities—there could also be reverse causation, in which high incomes lead to the control of malaria or to a reduction in transport costs.

Regression 6 in table 2 includes a vector of variables related to political and economic institutions: Socialism, which is a dummy variable for socialist economic institutions; New state, which measures the proportion of time under colonial rule; Openness, which measures the proportion of time between 1965 and 1990 that the country was open to international trade; and Public institutions, which measures the quality of government institutions. Consistent with many recent studies, we find that openness and the quality of public institutions are highly correlated with the level of income. Socialism is not significant, probably because of the strong collinearity with Openness and because of the smaller dataset once the Public institutions variable is included (since that variable is not available for most of the socialist economies). Newly independent countries also do not have significantly lower income levels. If the malaria index is not included, the New state variable is highly significant, which suggests that the heavy burden of disease in tropical Africa and Asia made these regions more susceptible to colonization.

Thus we find that both policy and geography variables are strongly correlated with 1995 GDP per capita. Geography may be even more important than is suggested by this equation, because there are reasons to believe that favorable geography plays a role in inducing growth-promoting institutions—such as open trade and an efficient public bureaucracy—a linkage examined briefly below.

Geography and Growth of Per Capita Income

We now examine the forces of convergence and divergence by estimating a cross-country growth equation that allows for the possibility of catching-up effects.

Specifically, we estimate a model of average annual growth during 1965–90 conditional on per capita income levels in 1965. (The dates are determined by data availability. For the growth equations, we use Summers and Heston 1994 for our measures of PPP-adjusted GDP per capita.) We test whether growth is affected by the initial income level (negatively in the case of convergence or positively in the case of divergence) as well as by geographic variables, holding constant initial income and other policy and institutional variables (table 3).

We start with a baseline equation similar to those in Barro and Sala-i-Martin (1995), in which average annual growth between 1965 and 1990 is a function of initial income in 1965, initial education level in 1965 (measured by average years of secondary school among the population), log of life expectancy at birth in 1965, openness of the economy to international trade, and quality of public institutions (regression 1 in table 3).²² We find evidence for conditional convergence and stan-

Table 3. Relationship between GDP Growth and Selected Variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>gr6590</i>	<i>gr6590</i>	<i>gr6590</i>	<i>gr6590</i>	<i>gr6590</i> (<i>TSLS</i>) ^a	<i>gr6590</i>	<i>gr6590</i>
GDP per capita 1965	-2.3 (7.70)	-2.4 (8.02)	-2.5 (8.06)	-2.6 (7.87)	-2.7 (7.60)	-2.3 (7.41)	-2.4 (7.09)
Years of secondary schooling 1965	0.3 (1.75)	0.2 (1.77)	0.2 (1.32)	0.2 (1.34)	0.2 (1.15)	0.1 (0.81)	0.1 (0.89)
Log life expectancy 1965	6.6 (7.23)	5.5 (6.21)	4.3 (4.45)	3.3 (3.60)	2.4 (1.79)	4.1 (4.53)	3.4 (3.89)
Openness (0–1)	1.9 (5.49)	1.9 (4.79)	1.7 (4.79)	1.7 (4.70)	1.7 (4.39)	1.8 (4.79)	1.8 (4.66)
Public institutions (0–10)	0.3 (3.08)	0.3 (2.63)	0.3 (3.32)	0.4 (3.92)	0.5 (3.66)	0.3 (3.20)	0.4 (3.47)
Tropical (percent)		-0.9 (2.28)	-0.6 (1.35)	-0.5 (1.09)	-0.4 (0.82)	-0.7 (1.89)	-0.5 (1.44)
Pop100km (percent)		1.0 (3.07)	0.9 (3.01)	0.8 (2.64)	0.6 (1.91)		
LDistance		0.0 (0.24)					
Malaria index 1966			-1.2 (2.15)	-2.0 (3.60)	-2.6 (3.87)	-0.9 (1.86)	-1.6 (2.89)
dMal6694				-2.5 (3.93)	-4.5 (2.12)		-1.9 (2.94)
Log coastal density						0.3 (4.91)	0.2 (4.34)
Log interior density						-0.1 (2.26)	-0.1 (1.60)
Constant	-8.9 (2.90)	-4.1 (1.17)	1.3 (0.34)	5.9 (1.57)	9.8 (1.76)	0.7 (0.19)	4.1 (1.08)
Number of observations	75	75	75	75	75	75	75
R ²	0.71	0.75	0.77	0.80	0.78	0.80	0.82

a Two-stage least squares

Note Numbers in parentheses are absolute values of robust t-statistics

Source. See appendix

standard results for the other variables: output is an increasing function of education, life expectancy, openness, and the quality of public institutions. In regression 2 of table 3 we add *Tropicar*, *Pop100km*, and *LDistance*. *Tropicar* and *Pop100km* are highly significant and of the expected sign. All other things being equal, annual growth is 0.9 percentage points lower in tropical countries than in nontropical countries. Annual growth in landlocked countries (*Pop100km* = 0) is 1.0 percentage point lower than in coastal countries. Interestingly, the *LDistance* variable is not significant, suggesting that distance to the economic core may be subsumed by other variables. Indeed, countries closer to the core economies tended to be more open and have better public institutions. If we drop those two variables, *LDistance* has the expected negative sign and is statistically significant (not shown).

In regression 3 of table 3 we drop *LDistance* and add a measure of malaria at the beginning of the period.²³ Initial malaria prevalence is strongly correlated with poor economic performance. Countries that had severe malaria in 1966 grew 1.2 percentage points a year slower than countries without *falciparum* malaria, even after controlling for life expectancy. The estimated effect of being in the tropics becomes smaller than it was without the malaria variable and insignificant. Clearly, the malaria variable is picking up the explanatory power of the tropics variable. The effect of initial life expectancy is also reduced, though it remains large and significant.

The effect of malaria is accentuated if we also account for changes in malaria over the period. Countries that had severe malaria but reduced its prevalence grew more rapidly than countries that were not able to control it. We use the malaria index for 1994 and calculate the change in malaria in the intervening period, *dMal6694*. In regression 4 of table 3 countries with severe malaria and no change in malaria are estimated to have grown 2.0 percentage points slower than countries free of *falciparum* malaria, other things being equal. Countries that had severe malaria in 1966 but eradicated it by 1994 are estimated to have grown 2.5 percentage points faster than countries with malaria, or slightly (0.5 percentage points) faster than countries without malaria.

None of the countries in the sample with 100 percent of their land area subject to *falciparum* malaria was able to eradicate it during this period, however. Reductions in malaria were largest in the countries with the least malaria in 1966 (table 4). (Temperate ecozones were effectively free of *falciparum* malaria in 1966.) Reductions in malaria occurred in desert (nontemperate) ecozones and the subtropics, while the malaria index increased slightly in tropical ecozones.

The change in malaria incidence could be partly due to economic growth—if growth provided countries with the economic and institutional resources to carry out effective control programs. To account for the possible impact of economic growth on malaria reduction, we instrument the malaria change with four subtropical ecozone variables and two desert tropical ecozone variables.²⁴ Regression 5 in table 3 shows that the estimated impact of malaria on growth increases when change in malaria is instrumented. There is no indication that faster growth is the cause of malaria reduction. Some of the countries with the largest increases in malaria, such as India and Sri Lanka, had steady, if unspectacular, economic growth over this

Table 4. Levels and Reductions in Malaria Prevalence between 1966 and 1994, by Ecozone

<i>Ecozone</i>	<i>Malaria index 1966</i>	<i>Average reduction (percent)</i>
Temperate (n = 57)	0.2	-0.2
Desert (n = 23)	27.8	-8.8
Subtropical (n = 42)	61.7	-5.0
Tropical (n = 21)	64.9	0.5

Note Countries are classified by their predominant ecozone. temperate (temperate, boreal, and polar ecozones), desert (tropical and subtropical deserts), subtropical (nondesert subtropical), and tropical (nondesert tropical). The index and average reduction are unweighted averages over countries

Source See appendix

period. Likewise, some countries with dramatic decreases in malaria, such as Namibia, had almost no economic growth. Only after controlling for other relevant variables can the effect of malaria reduction on growth become apparent.

The index of malaria, and malaria change, may be more than just a measure of malaria. It may be picking up the prevalence of other tropical diseases not well indicated by average life expectancy and tropical area. Among tropical diseases, malaria is widely recognized to be the most important. But the malaria index may also be a proxy for scourges like onchocerciasis, filariasis, and trypanosomiasis.

Malaria occurs throughout the tropics but is concentrated in Sub-Saharan Africa, where 90 percent of the estimated cases occur each year (WHO 1997b). Moreover, Sub-Saharan Africa is the only region where *falciparum* malaria predominates. When the region is left out of regression 5 (not shown), the malaria coefficients fall to about half their size in the full sample and the estimates lose statistical significance (though the change in malaria from 1966 to 1994 is significant at the 7 percent level). With the loss of 15 Sub-Saharan countries from the sample, it is not possible to obtain accurate estimates of the effect of the tropics and the effect of malaria separately. When Tropical is left out of the non-African regression, both initial malaria and change in malaria are significant at the 5 percent level.

In regression 6 of table 3 we test for agglomeration effects. The basic idea is to see how economic growth depends on the scope of the market. A plausible measure of the scope of the market is GDP per square kilometer within the economy in the initial year, 1965. We separate GDP per square kilometer on the coast and GDP per square kilometer in the hinterland because population density on the coast is likely to be associated with an increased division of labor and increasing returns, while population density in the hinterland is likely to be associated with diminishing returns. Note that $\ln \text{GDP density} = \ln \text{GDP per square kilometer} = \ln \text{GDP per capita} + \ln \text{Population per square kilometer}$. And because $\ln \text{GDP per capita}$ is already a regressor, we can enter population density or GDP density interchangeably into the regression. For countries whose entire population is within 100 kilometers of the coast, the population density of the hinterland is entered as 0. For countries whose entire population is more than 100 kilometers from the coast, the population density of the coastal region is entered as 0. We also drop Pop100km as a separate regressor because it is highly collinear with the two population density variables.

Note that we use 1994 measures of Pop100km and 1965 population levels for the country as a whole to calculate the population densities in coastal and hinterland regions.

The regression estimate is revealing. We now find that higher coastal population density is associated with faster growth, while higher hinterland population density is associated with lower growth. Thus economies of agglomeration appear to be at play in coastal regions, though they are not powerful enough to overcome the other tendencies toward conditional convergence in income levels. Large populations appear to be a net disadvantage for hinterland economies, which must rely more on their natural base and have few opportunities to absorb population through manufacturing and international trade based in cities.

With separate inland and coastal agglomeration effects in regression 6, the estimated effect of malaria becomes less precise, slipping to a 7 percent significance level. In contrast, if initial malaria and malaria change are included in regression 7 they are both strongly significant, but the diminishing returns of population density in the hinterland lose statistical significance. We are pushing the limits of the degrees of freedom in our data, but the results suggest that both malaria prevalence and inland population concentrations are detrimental to growth.

Several conclusions emerge from these equations:

- Both geography and policy matter. Geography is not necessarily destiny, but more than good policy is needed to foster economic growth.
- The tropics are not conducive to growth. This effect seems to be strongly related to the presence of malaria, which may be a proxy for a range of tropical maladies geographically associated with that disease.
- Coastal regions are good for growth. Access to the coast seems to matter because it lowers transport costs and creates agglomeration economies.
- High population density appears to be conducive to growth on the coast and inimical to growth in the hinterland.
- Distance from core markets is not an important determinant of economic growth.

What are the regional implications of these findings? Sub-Saharan Africa is especially hindered by its tropical location, high prevalence of malaria, small portion of people living near the coast, and low population density near the coast. Europe, North America, and East Asia, by contrast, have been favored on all four counts. South Asia is burdened by its largely tropical location, large portion of people living in the hinterland, and the high population density there. The transition economies of Eastern Europe and the former Soviet Union, many of which are landlocked, have suffered from their small share of population living near the coast and low population density near the coast. But they have benefited from their lack of exposure to tropical disease. Finally, Latin America is moderately coastal, but with relatively low coastal population densities. Moreover, Latin America is moderately exposed to the problems of the tropics, including malaria.

The growth rates of Sub-Saharan Africa, South Asia, and Latin America are decomposed, using regression 7 in table 3. The deviation of each region's growth

from that of East Asia is shown in table 5. In Sub-Saharan Africa geography and health factors are estimated to reduce growth by 3.0 percentage points a year, more than policy and education factors. In South Asia geography and health factors reduce growth by -0.8 percentage points a year. And in Latin America geography and health factors explain almost none of the shortfall in growth relative to East Asia. As we suggest in the next section, this accounting may understate the real role of geography, since economic policies are likely to be a function of geography.

Geographic Effects on Economic Policies

As noted, geography may affect economic policies by altering the tradeoffs facing government. A coastal economy, for example, may face a high elasticity of output response with respect to trade taxes, while an inland economy does not. As a result a revenue-maximizing inland sovereign may impose harsh trade taxes, while a coastal sovereign would not. In this section we briefly explore this idea, focusing on the effect of geography on the choice of protection or openness to trade in 1965–90.

The first step is to check the underlying notion that the responsiveness of growth to openness depends on geography. So far we have entered the openness and geography variables in a linear manner, not allowing for interactions. To check the possibility of interactions, we estimate the basic regression equation for three sets of countries—all, coastal ($\text{Pop100km} \geq 0.5$), and hinterland ($\text{Pop100km} < 0.5$)—and check the coefficient on the openness variable. Because we lose degrees of freedom in this exercise, we estimate the barebones growth equation in which annual average growth between 1965 and 1990 is a function of initial income, openness,

Table 5. Growth Rates in Selected Regions Relative to East Asia

<i>Variable</i>	<i>Sub-Saharan Africa</i>	<i>South Asia</i>	<i>Latin America</i>
Growth	-4.2	-2.8	-3.6
Explained	-3.7	-2.0	-3.0
Initial GDP	1.4	1.0	-1.1
<i>Geography and health</i>			
Total	-3.0	-0.8	-0.2
Coastal density	-0.7	0.0	-0.5
Interior density	0.0	-0.3	0.1
Tropical	-0.1	0.1	-0.1
Malaria index	-1.0	-0.1	0.3
Life expectancy	-1.2	-0.5	0.0
<i>Policy and education</i>			
Total	-2.1	-2.1	-1.8
Openness	-1.2	-1.2	-1.0
Public institutions	-0.7	-0.9	-0.7
Secondary education	-0.2	-0.1	0.0
Residual	-0.5	-0.8	-0.6

Source: Regression 7 in table 3

malaria in 1966, the change in malaria between 1966 and 1995, and the log of life expectancy in 1965. The results for the openness coefficient are as follows: for all economies ($n = 92$), $\beta = 2.6$ ($t=6.8$); for coastal economies ($n = 46$), $\beta = 3.3$ ($t = 6.5$); and for hinterland economies ($n = 46$), $\beta = 1.4$ ($t = 2.5$). The responsiveness of growth to trade seems to be more than twice as high in coastal economies.

The next step is to see whether coastal countries are more likely to choose open trade policies. Here we use the percentage of land area within 100 kilometers of the coast (Land100km) as a measure of coastal access. We regress the extent of openness during 1965–90 on Land100km , Tropicar , and the initial income level:

$$\text{Openness}_{6590} = -1.12 + 0.23 \text{Land100km} - 0.18 \text{Tropicar} + 0.19 \ln \text{GDP}_{1965}$$

(3.2) (2.1) (2.0) (4.5)

$n = 106, R^2 = 0.44.$

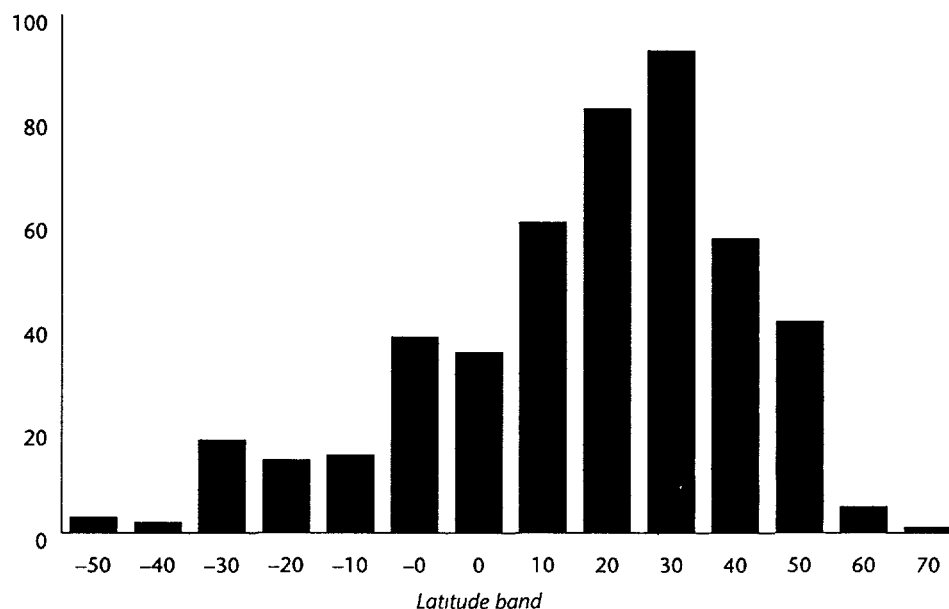
These results imply that coastal economies may be more likely to adopt open trade policies. This is certainly evident in East Asia, where economies such as the Republic of Korea, Malaysia, Taiwan (China), and Thailand opened themselves to trade in the early 1960s, long before other developing countries did so. The early liberalizers, on the whole, were coastal economies. Still, the results are at best suggestive and should be tested more carefully.

Population Distribution and Economic Activity

The distribution of population around the world is far from uniform. In some regions large expanses of land are virtually uninhabited, while almost all the land in Europe and coastal South and East Asia is farmed or occupied by towns or cities. Differences in population density at different latitudes are dramatic (figure 2).

Two types of geographic features seem to support high population densities. First are features that favor dense agricultural settlements, such as arable soil, inland rivers for local transport and irrigation, and climatic and ecological systems conducive to rice cultivation (which supports an especially high labor intensity of production relative to other grains). Second are features that support modern economic growth, such as access to the coast (and thereby to international trade). Because population densities have a long time dependence, the current distribution of world population was heavily influenced by demographic trends well before the period of modern economic growth. As a result population density across regions remained almost unchanged between 1800 (map 6) and 1994 (see map 2). Broadly speaking, global population patterns reflect suitability for agriculture more than suitability for modern economic growth. And the legacy of low population densities in the New World persists despite several centuries of emigration from the Old World.

The biggest changes have occurred in the lands of new settlement: the Americas, Australia, and New Zealand. These parts of the world had been cut off from other continents and did not have plant and animal species that could easily be domesticated to produce high yields. Thus they could not support the large population con-

Figure 2. Population Density by Latitude, 1994*People per square kilometer**Source* See appendix

centrations of Eurasia (Diamond 1997). European settlers, having devastated the sparse indigenous populations through conquest and the spreading of unfamiliar diseases, had not yet grown to large numbers by 1800.

The geographic conditions conducive to dense agrarian populations are often very different from those conducive to economic growth. In particular, agriculture depends more on access to fresh water than on access to the ocean. Over history this has led to high concentrations of inland populations that are now substantially cut off from international trade. And as noted earlier, dynamics of population change may exacerbate biases toward high inland concentrations. In richer regions rising incomes have reduced fertility, making population growth self-limiting. But in rural areas with poorer economic growth prospects, population growth rates are among the highest in the world.

Regions where geography supports high population densities but not economic growth are the sites of the most severe and intractable poverty. Inland China, north-central India, central Asia, and inland Africa are all far from world trade and dependent on labor-intensive agriculture, with significant disadvantages for modern economic growth. Severe endemic disease burdens, especially in Africa, add to the geographic obstacles. The role of geography in shaping the distribution of population can be seen in the simple population growth identity. Current population depends on the population at some point in the past and the growth rate during the intervening period:

$$(8) \quad P_{i1} = P_{i0}e^{r_i T}$$

where P_{i1} is the current population density in location i , P_{i0} is past population density, r_i is the instantaneous growth rate, and T is time elapsed between period 0 and period 1. The population growth rate in each location, r_i , depends on geography as well as on initial population. If r_i is not allowed to depend on the initial population density, then current population is always exactly proportional to past population. The population growth rate is given by

$$r_i = g \ln P_{i0} + \gamma \ln X_i$$

where g and γ are parameters and X_i is a vector of geographic characteristics. Taking logs of equation 8 and adding an error term ε_i ,

$$(9) \quad \ln P_{i1} = (1 + gT) \ln P_{i0} + \gamma T \ln X_i + \varepsilon_i$$

We can regress the current population density on an initial population density (say, in 1800) and geographic characteristics, all in logs, using equation 9. The first coefficient will tell us the degree of persistence in population density. If the coefficient is less than 1, g is negative: higher-density regions grow slower. This is a measure of “convergence” of population density across space, similar to convergence in economic growth equations. The second set of coefficients tells us the impact of geography on population density given the initial density (that is, the impact of geography on population growth in the past 200 years). If we take the initial point at the time of the first appearance of humans ($T' \cong 500,000$, according to Diamond 1997), then P'_{i0} and

$$(10) \quad \ln P_{i1} = \gamma' T' \ln X_i + \varepsilon_i$$

The coefficient vector estimated from this specification tells us the unconditional impact of geography throughout time on current population densities.

Equations 9 and 10 are estimated using the following geographic features: accessibility to coast and rivers, elevation, prevalence of malaria, soil qualities and water availability, and ecozones (table 6). (The variables and their sources are explained in the appendix.) Several clear patterns emerge:

- Proximity to inland or ocean-navigable rivers is an important predictor of population density, more important than proximity to the coast.
- Good soil and water supply are important factors in population density.
- Population densities are highest in the moist temperate ecozone.
- Population density is greater at high altitudes in the tropics but lower at high altitudes in the temperate zone.
- The prevalence of malaria is positively correlated with population density.
- There is tremendous persistence in relative population density over the centuries, but there is also some convergence toward a more uniform density. (The coefficient on population density in 1800 is positive and less than 1.)

- After taking into account all geographic factors, population densities are much higher in Eurasia and lower in the Americas, Australia, and New Zealand.

The tendency of populations to cluster near non-ocean-navigable inland rivers (rather than on the coast or near ocean-navigable rivers) is directly contrary to the importance of these rivers for economic output. This tendency, together with the clustering of populations in areas of good soil and water (especially soil and water suitable for rice cultivation) and in more agriculturally productive ecozones, suggests that an area's suitability for agriculture has been the driving force behind population distribution. The distribution of population near rivers rather than near the coast is striking when the regressions are done by region (not shown). In South Asia, the former Soviet Union, Western Europe, and East Asia population densities near coasts are lower than elsewhere, conditional on their distance to rivers. In Western Europe and East Asia populations had the good fortune to cluster near rivers navigable to the sea, while South Asian populations have poor access to

Table 6. Impact of Geography on Log Population Density, 1994

<i>Variable</i>	<i>Impact of geography on population density given population in 1800 (1)</i>	<i>Overall impact of geography on population density (2)</i>
Log population density 1800	0.612 (79.63)**	
Eurasian continent		1.136 (28.15)**
Lands of new settlement		-0.998 (23.73)**
Log distance (kilometers) to Coast	-0.040 (2.26)*	-0.145 (8.13)**
Ocean-navigable river	-0.113 (6.28)**	-0.188 (10.2)**
Inland river	-0.324 (26.48)**	-0.235 (18.4)**
Log elevation in temperate zone		
Less than 1,000 meters	0.018 (1.35)	0.096 (6.77)**
1000–2,000 meters	0.859 (9.63)**	1.108 (11.67)**
More than 2,000 meters	-1.361 (12.83)**	-0.279 (8.47)**
Log elevation in tropics		
Less than 1,000 meters	-0.034 (6.47)**	0.034 (7.28)**
1,000–2,000 meters	1.801 (6.04)**	2.133 (6.2)**
More than 2,000 meters	0.846 (3.2)**	1.296 (3.15)**
Malaria (fraction of area)	0.109 (2.48)*	2.104 (4.72)**

Table 6. Impact of Geography on Log Population Density, 1994 (continued)

Variable	<i>Impact of geography on population density given population in 1800 (1)</i>	<i>Overall impact of geography on population density (2)</i>
Soil and water		
Rice-growing land (fraction of area)	1.064 (10.44)**	1.335 (13.41)**
Soil suitability (0–100)	0.127 (21.47)**	0.134 (22.25)**
Log stream density ^a (number per cell)	0.144 (17.77)**	0.192 (23.42)**
Ecozones (relative to moist temperate) ^b		
Polar and boreal	-2.273 **	-3.202 **
Desert	-1.224 **	-1.840 **
Dry temperate	-0.176 **	-0.409 **
Very wet temperate	-1.204 **	-1.539 **
Dry tropical	-0.380 **	-0.494 **
Wet tropical	-0.932 **	-1.257 **
Number of observations	13,976	14,418
R ²	0.74	0.73

* Significant at the 5 percent level

** Significant at the 1 percent level

Note: The regressions included constants that are not reported. Numbers in parentheses are absolute values of robust t-statistics.

a Stream density is the count of streams in each 1 degree cell from satellite data

b Includes 36 ecozones (moist temperate ecozone is excluded). The reported estimates are the average of ecozone coefficients in each ecozone group. "****" means that all the constituent coefficients were statistically different from zero at the 1 percent level.

Source: See appendix

waterborne trade. Latin America is the only region with a higher concentration of population on the coast than near rivers. The highest population densities that are far from coasts and from ocean-navigable rivers are in central Africa, South Asia (especially the Gangetic Plain), inland China (with heavy concentrations in the river valley systems and Manchuria), and Central Asia, including Iran, Iraq, Anatolia, and countries near the Caspian Sea (map 7).

The much higher population densities at higher altitudes in the tropics could be due to the less hostile disease environment there, since many tropical disease vectors are sensitive to altitude and temperature. But this explanation seems inconsistent with the positive correlation between population density and malaria. In individual regions population densities are (statistically) significantly lower in malarial areas in all regions except Africa, where population is much denser in malarial areas. If Africans in malarial areas have built up partial immunities to malaria, they may not seek to avoid infec-

tion by moving to nonmalarial areas that may have other disadvantages (such as lack of water). Given the strong negative correlation between malaria and income levels, the higher population density in malarial areas is extremely worrisome.

There may be long-term geographic explanations for the high population densities in Eurasia. As noted, Diamond (1997) argues that Eurasia's east-west axis, which runs along rather than across ecozones, has allowed the movement of crop varieties, ideas, and goods. He speculates that Eurasia had the best selection of native plants and animals, while the lands of new settlement (the Americas, Australia, and New Zealand) had the least conducive flora and fauna for original domestication. Until the modern era new lands were physically isolated from the rest of the world and so were cut off from the diffusion of technology, ideas, and trade that permeated Eurasia. Africa also suffered from isolation and from a north-south axis that hindered the diffusion of Eurasian innovations and crop varieties.

Population in the lands of new settlement heavily reflects the equilibrium between the economic productivity of those regions and income levels in the European sending countries. Much of the indigenous population was exterminated by conquest and disease soon after its first contact with Europeans; since then incomes of the settlers have generally been somewhat higher than incomes of Europeans considering migration to distant lands. Although incomes in North America, Australia, and New Zealand have remained on par with those in Western Europe, the costs of distance mean that population densities are much lower than in Europe (the eastern United States comes closest to being an exception). Africa, by contrast, resisted European conquest until the end of the 19th century, largely because of the high rate of mortality from malaria and tropical diseases for would-be European explorers and conquerors (Curtin 1989). Once quinine was applied as a prophylaxis for malaria, Europe subjugated most of Africa. But except in the temperate southern and northern extremes, few Europeans settled in Africa. Thus the low productivity of the land in Africa was reflected in low population densities *and* low incomes.

Research Directions and Policy Implications

One skeptical reviewer of an early draft of this article said, "Fine, but we knew all this in seventh-grade geography." We have three responses. First, that is not really true. Seventh-grade geography does not attempt to quantify the advantages or disadvantages of various regions in a systematic way, holding constant other determinants of economic performance. Second, even if it is true, that learning is lost somewhere on the way to graduate school. The vast majority of papers on economic development and growth in the past decade have neglected even the most basic geographic realities in cross-country analyses. In considerable writing on Africa, for example, many socioeconomic variables have been tested for their effects on growth. But little attention has been paid to the implications of the large portion of landlocked countries, the disease environment, the harsh climate and its effects on agriculture, or the implications of low population densities in coastal areas. Third, the policy implications of these findings, if the findings are true, are staggering. Aid

programs should be rethought, and the crucial issue of migration should be brought into much sharper focus.

The research agenda needs to be reshaped in light of the importance of geographic variables. We know precious little about the underlying relationships between climate and agricultural productivity, disease vectors, and public health. Not only do we not know the costs of malaria in terms of economic development, we barely know the quantitative extent of the disease. Cause-of-death data are not available for most developing countries and even fewer data are available on illness. We lack basic data on transport costs that are comparable across countries and, even more important, within countries between hinterlands and urban areas. By neglecting geographic variables, we may overstate the role of policy variables in economic growth and neglect some deeper obstacles (although, because policy variables are often so poorly measured in cross-country work, there is an inherent downward bias due to measurement errors).

The following research questions deserve much greater scrutiny:

- How do transport costs differ across countries? How much of these differences are related to policy (port management, road maintenance), market structure (pricing by shipping cartels), or physical geography (allowing inland, coastal, or oceanic trade)? How are transport costs likely to change as a result of new information technologies, better intermodal transport, and other trends?
- What is the burden of disease on economic development? What are the channels of effects—that is, the direct and indirect costs of infant and child mortality, adult morbidity, and premature death? What are the main channels of morbidity—direct effects, interactions with other diseases, interactions with nutrition? To what extent do different burdens of disease reflect policy (that is, provision of public health services), resource availability for health spending, or intrinsic geographic factors such as the ecology of disease vectors?
- To what extent are differences in agricultural productivity a result of policy (taxation of agricultural inputs and outputs), quality of inputs, scope and scale of agricultural research, and intrinsic geophysical and biological conditions?
- How are fertility decisions affected by geography? Do Sub-Saharan Africa's high fertility rates result from low population densities in rural areas, limitations of nonagricultural activities in the hinterland, policy decisions or limitations (such as lack of adequate family planning), or institutional arrangements (such as communal land tenure, which may lead to externalities in family size)?

Gaining a better grasp of these issues will lead to a next step of analysis that looks at the extent to which transport costs, disease burdens, agricultural productivity, and population growth and density affect overall economic performance. Consider, for example, the relatively straightforward issue of transport and communications costs. It might be assumed that falling transport costs would favor the hinterland, which is burdened by high transport costs. But Krugman and Venables (1995) have shown that a drop in transport costs from high to moderate levels can hurt a high-cost region by giving even greater benefit to a medium-cost region. Consider our simple

setup in equation 3. Suppose that there are two economies with different transport costs and that $\tau_i = \exp(d_i\mu)$, where d_i is the distance of economy i from the core economy and μ is a transport cost parameter that declines over time. Suppose that there is a “near” and a “far” economy, with $d^n < d^f$. When μ is high, both economies have zero growth. When μ is 0, both have equal and high growth. It is when μ takes a mid-range value that growth rates differ, with the near economy growing faster than the far economy. Thus even when we know how transport costs differ and how they are likely to evolve, we need an accurate spatial model to understand the implications.

Of course, the policy implications of these geographic considerations must be informed by clearer research results. Even now, however, we can identify several areas of public policy that almost surely should be adjusted. First, the special problems of landlocked countries and hinterland populations within coastal countries deserve closer scrutiny. The 28 landlocked countries outside Europe, containing 295 million people in 1995, are among the poorest in the world, with an average per capita income of \$1,673. In many cases the infrastructure linking these countries to world markets is seriously deficient. Coastal countries harass landlocked countries, neglect the road networks that would link them to the coast, or impose punitive effective taxation through transit and port charges. In some cases heated political clashes have taken place between interior and coastal countries. Bolivia and Chile, for example, still lack diplomatic relations 119 years after the War of the Pacific cost Bolivia its coastline. Aid programs to improve transport infrastructure linking landlocked countries to ports almost necessarily require the cooperation of several countries. Crucial infrastructure aid for Rwanda, for example, includes the repair and maintenance of the Kenyan road from Nairobi to Mombasa, which transports Rwandan and Ugandan tea to the Indian Ocean. Such cross-national needs are hard to coordinate and are often neglected by country-based donor efforts. Policymakers should also pay more attention to transport conditions for hinterlands within national economies, such as Uttar Pradesh in India, where more than 140 million people live several hundred kilometers from the coast.

Second, policymakers should examine the likelihood and desirability of large-scale future migrations from geographically disadvantaged regions. Suppose that sizable populations face local cost or disease conditions that prohibit economic growth. The result is likely to be growing pressures for mass migration, first within countries, then across national borders, and finally internationally. We have not yet studied the linkage between geography and migration, though it is painfully evident that the linkage is strong. Consider landlocked Bolivia, for example. Some 15–20 percent of Bolivians live in neighboring countries, especially Argentina. About a third of Burkinabés live in Ghana, Côte d’Ivoire, and elsewhere. Throughout southern Africa there are large, relatively uncontrolled population movements across national boundaries. The consequences of these movements are becoming increasingly complex and often deleterious, as AIDS and other diseases spread and environmental and public health institutions prove unable to cope.

Third, to the extent that the arguments in this article are correct, they cast a dramatic light on current population trends. We have shown that future population

increases are likely to be largest precisely in the most geographically distressed economies. According to United Nations population projections, the highest projected growth rates between 1995 and 2030 are in regions that are least coastal, most tropical, and most distant from the core economies (map 8). If we regress the projected population growth on geographic characteristics, we find that

$$\begin{aligned} \text{Annual population growth, 1995–2030 (projected, in percentage points)} = \\ -2.1 + 0.91 \text{ Tropical} + 0.41 \text{ LDistance} - 0.74 \text{ Pop100km} \\ (3.34) (6.42) \quad (5.31) \quad (4.63) \\ n = 147, R^2 = 0.59. \end{aligned}$$

Inland countries (Pop100km = 0) have projected annual population growth rates that are 0.74 percentage points a year higher than those in coastal countries (Pop100km = 1). Tropical countries have projected population growth rates that are 0.91 percentage points higher than those in nontropical countries.

Population pressures in inland and tropical areas are likely to intensify pressures for mass migration. Thus a more urgent review of population policy is required. A certain calm has descended over this policy area on the questionable grounds that population growth does not matter for per capita income growth. We have seen the half-truth of this assessment: it may be true for coastal countries engaged in the international division of labor, but it is most likely untrue for the geographically distressed regions where population increases will be most dramatic.

Finally, donors should reexamine the balance of aid between policy-based lending to individual governments—which is the most popular form—and greatly enhanced aid for basic research on tropical agriculture and tropical public health. Our results suggest that the tropics are damned not just, or even mainly, by bad policies but by difficult inherent conditions. If this is the case, the relentless focus on policy reform may be misguided. A more effective approach to controlling malaria might do more to improve the economic environment—and incidentally, might improve policy by enhancing the incentives for good policies facing the sovereign. Many of the core issues in tropical health and agriculture are prime examples of international public goods that require a concerted scientific and financial commitment far beyond that available from any single government. The coordinated agricultural research aid effort is seriously underfunded; the situation in tropical public health is even more desperate.

Appendix. Data Definitions and Sources

This appendix defines the variables used in the analysis and gives the data sources.

Dependent Variables

GDP PER CAPITA. PPP-adjusted GDP per capita in 1950 and 1990 are from Maddison (1995, tables D1 and F4). PPP-adjusted GDP per capita in 1995 are from

World Bank (1998). When World Bank (1998) data were not available, U.S. Central Intelligence Agency (CIA) data were used. Data from CIA (1996) were used for Afghanistan, Albania, Cambodia, Cuba, Eritrea, Iraq, Democratic People's Republic of Korea, Kuwait, Liberia, Libya, Somalia, Sudan, Tanzania, former Yugoslav Republic of Macedonia, and Yugoslavia (Serbia/Montenegro). Data from CIA (1997) were used for Bosnia and Herzegovina, Bhutan, Brunei, Djibouti, Equatorial Guinea, French Guiana, Myanmar, and Taiwan (China). Additional data for countries in map 1 are from CIA (1997).

GDP GROWTH. Instantaneous growth rates of PPP-adjusted GDP per capita from 1965 to 1990 are from the Penn World Tables 5.6 (Summers and Heston 1994).

Transport Cost and Market Proximity Measures

LAND100KM. The proportion of a country's land area within 100 kilometers of the coast, excluding coastline in the Arctic and sub-Arctic regions above the winter extent of sea ice. Data are from the National Geographic Society (NGS 1995) for science, and the World Data Bank II (ESRI 1996b) for digital coastlines.

POP100KM. The proportion of a country's population in 1994 within 100 kilometers of the coast (as defined for Land100km). The data for population distribution in 1994 come from the first detailed geographic information system world population dataset (shown in map 2), described in Tobler and others (1995).

POP100CR. The proportion of a country's population in 1994 within 100 kilometers of the coast or an ocean-navigable river, excluding coastline in the Arctic and sub-Arctic regions above the winter extent of sea ice and the rivers that flow to it. Rivers were classified as ocean navigable mainly according to descriptions in Rand McNally (1980), Britannica Online (1998), and *Encyclopedia Encarta* (1998). (Precise information on the classification of river systems is available from the authors.) Ocean-navigable rivers (displayed in map 7) are calculated from digital coastlines in ESRI (1996b) and rivers in ESRI (1996a). The population data are as for Pop100km.

COASTAL DENSITY. Coastal population/coastal square kilometers = (population x Pop100km)/(land area x Land100km). The unit is people per square kilometer.

INTERIOR DENSITY. Interior population/interior square kilometers = [population x (1 - Pop100km)]/[land area x (1 - Land100km)]. The unit is people per square kilometer.

LANDLOCKED, NOT IN EUROPE. Excludes landlocked countries in Western and Central Europe (Austria, the Czech Republic, Hungary, former Yugoslav Republic of

Macedonia, Slovakia, and Switzerland). Includes the Eastern European countries of Belarus and Moldova.

LDISTANCE. The log of the minimum Great-Circle (air) distance in kilometers to one of three capital-goods-supplying regions: the United States, Western Europe, or Japan, measured as distance from the country's capital city to New York, Rotterdam, or Tokyo.

C.I.F./F.O.B. SHIPPING COST MARGIN. The ratio of c.i.f. import prices to f.o.b. import prices as a measure of transport costs from International Monetary Fund data (Radelet and Sachs 1998).

Other Geographic Variables

TROPICAR. The proportion of a country's land area within the geographic tropics. Calculated from ESRI (1996b).

MALARIA INDEX 1966. Index of malaria prevalence based on a global map of the extent of malaria in 1966 and the proportion of *falciparum* malaria, from the World Health Organization (WHO 1967). The percentage of each country's land area subject to malaria was calculated from the digitized 1967 map shown in map 5 ("some risk" areas excluded). The intensity of malaria is captured by the percentage of cases that were the malignant *falciparum* strain in 1990 (WHO 1992). For African countries without 1990 *falciparum* data, we used WHO (1997a) data (in which almost all African countries with malaria are described as predominantly *falciparum*, which we classified as 100 percent). The index is the product of the percentage of land area subject to malaria times the percentage of *falciparum* malaria cases.

MALARIA INDEX 1994. Constructed in the same way as the malaria index for 1966, based on a global malaria map for 1994 (WHO 1997b) and the percentage of *falciparum* malaria in 1990.

HYDROCARBONS PER CAPITA. Log of British thermal units per person of proven crude oil and natural gas reserves in 1993, from the World Resources Institute (WRI 1996).

SOUTHERN HEMISPHERE. Indicator for countries wholly below the equator, as well as Brazil, Democratic Republic of Congo, Republic of Congo, Ecuador, Gabon, Indonesia, and Kenya.

LAND AREA. Area in square kilometers, excluding submerged land. All data are from World Bank (1997) except for data on Mexico and Taiwan (China), which come from CIA (1997).

Other Economic, Social, and Political Variables

OPENNESS. Proportion of years a country was open to trade during 1965–90, based on criteria established in Sachs and Warner (1995a). An economy is considered open if it meets minimum criteria on four aspects of trade policy: average tariffs must be lower than 40 percent, quotas and licensing must cover less than 40 percent of imports, the black market premium must be less than 20 percent, and export taxes must be moderate.

PUBLIC INSTITUTIONS. The quality of public institutions is based on an index created by Knack and Keefer (1995) that averages five indicators of the quality of public institutions, including the perceived efficiency of the government bureaucracy, the extent of government corruption, the efficacy of the rule of law, the presence or absence of expropriation risk, and the perceived risk of repudiation of contracts by the government. Each country is scored on these five dimensions, based on surveys of business attitudes within the countries. The subindexes on the five measures are then summed to produce a single, overall index that is scaled between 0 and 10.

NEW STATE. The timing of national independence (0 if before 1914, 1 if between 1914 and 1945, 2 if between 1946 and 1989, and 3 if after 1989), from CIA (1996).

SOCIALISM. Equal to 1 if the country was under socialist rule for a considerable period during 1950–95, based on Kornai (1992).

LIFE EXPECTANCY 1965. Data are from United Nations (1996).

YEARS OF SECONDARY SCHOOLING 1965. Data are from Barro and Lee (1993).

URBANIZATION. Percentage of population living in urban areas in 1995, from World Bank (1998).

WAR-TORN. Countries that participated in at least one external war during 1960–85, from Barro and Lee (1994), with additional countries classified by the authors.

POPULATION. Total population in millions, from World Bank (1997).

One Degree by One Degree Population Database

The data for population in 1994 come from the first detailed geographic information system world population dataset (seen in map 2), described in Tobler and others (1995). We aggregated the 5' by 5' cells to 1° by 1° cells, creating about 14,000 observations. The world population distribution in 1800 comes from McEvedy and Jones (1978), mostly on a countrywide basis. The geographic data come from a

variety of sources. The incidence of malaria for each 1° cell was digitized from a WHO (1997b) map for 1994. The distance of each cell from the coast was calculated from World Data Bank II coastal boundaries (ESRI 1996b). These boundaries were edited to remove the coasts north of the winter extent of sea ice, as in Land100km above. Ocean-navigable rivers were classified as in Pop100cr above. Inland rivers are rivers classified as navigable by ESRI (1996a) but with no outlet to the sea, as well as rivers navigable to the sea but not navigable by ocean-going vessels. Elevation data are derived from the the U.S. National Aeronautics and Space Administration ETOPO world elevation database (NOAA 1988). Land used for rice-growing was derived from the ArcAtlas database on agriculture (ESRI 1996a). Soil depth and stream density (a count of the streams in each 1° cell from satellite data) come from NOAA (Sellers and others 1997). Soil suitability for rain-fed crops was derived from FAO (1995). A classification of land areas into 37 ecozones comes from Leemans (1990).

Notes

1. These figures refer to the unweighted average of GDP per capita in China, Hong Kong (China), Indonesia, the Republic of Korea, Malaysia, the Philippines, Singapore, Taiwan (China), and Thailand.
2. The term *geographic tropics* refers to the area between the Tropic of Cancer (23°45' north latitude) and the Tropic of Capricorn (23°45' south latitude), the band in which the sun is directly overhead at some point during the year.
3. All data are described in the appendix. Population data are based on geographic information system modeling, with a resolution of 5 minutes by 5 minutes. Some of the underlying database is less refined, however, with population interpolated to the 5 minute by 5 minute grid.
4. Among the 150 countries with populations greater than 1 million, the correlation between population density and GDP per capita in 1995 was 0.32.
5. We include the U.S. and Canadian regions bordering the Saint Lawrence Seaway and Great Lakes.
6. We have not yet assembled subnational GDP data. Therefore, to make the calculation we assume that GDP per capita is identical in all regions within a country. This approach understates the size of GDP in coastal regions because GDP per capita tends to be higher in those regions.
7. These 11 economies, in declining order of shares in global exports, are the United States (20.0), Japan (20.0), Germany (14.6), France (6.4), the United Kingdom (5.8), Italy (4.9), Canada (4.7), the Republic of Korea (3.2), Taiwan, China (2.9), Belgium (2.7), and the Netherlands (2.5). Other major exporting economies that are closely linked to the core production system include Singapore (4.3), China (2.7), Mexico (2.3), Malaysia (2.2), and Hong Kong, China (0.7).
8. The estimated significance level of the t-test is precise only if the errors are normally distributed, the two groups are statistically independent, and they are drawn from a random sample (not the whole population). The small estimated *p* values nevertheless indicate that the means are far apart. This caveat applies to the *p* values reported below.
9. There is an inconsistency in our classification because *tropics* is based on land area in the geographic zone but *subtropics* is defined according to ecozone (Leemans 1990). We rely on the geographic definition of the tropics both for convenience and because of its empirical rel-

evance in the regression estimates. Because there is no comparable geographic definition of the subtropics, we fall back on an ecozone definition.

10. The landlocked countries in Western and Central Europe are Austria, the Czech Republic, Hungary, the former Yugoslav Republic of Macedonia, Slovakia, and Switzerland.

11. Using geographic information system data, we calculate the coastal population in two ways. First, we take all land area within 100 kilometers of the open sea (except coastline in the Arctic and sub-Arctic regions) and measure the population in that area. Second, we identify river systems that accommodate ocean-going vessels on a regular basis and add land areas within 100 kilometers of such rivers. Due to the classification difficulties for navigability and the greater empirical relevance of coastal access alone, population near navigable rivers is not used in the rest of the article. (Information on the classification of river systems is available from the authors.)

For distance to “core” markets, we experimented with a number of measures, all of which produced similar outcomes. We therefore choose the simplest: the shortest distance between the country capital and New York, Rotterdam, or Tokyo. The construction of these variables is described in more detail in the data appendix under Pop100km, Pop100cr, and Ldistance.

12. The key recent reference on cross-country growth is Barro and Sala-i-Martin (1995) which, like hundreds of recent cross-country growth studies, makes no reference to economic geography. One recent (and nearly lone) exception is Hall and Jones (1996), which notes that countries’ economic productivity (measured by GDP per capita) increases with distance from the equator.

13. Smith does not discuss culture and economic development in detail in the *Wealth of Nations*, but it seems clear that, like other thinkers of the Scottish Enlightenment, he viewed human nature as universal and did not view culture as a primary differentiating factor in economic development. For example, Smith never bemoans the lack of entrepreneurial zeal in one place or another as an explanation for poor economic performance. Later thinkers, such as Max Weber, put great stress on culture, though the alleged linkages are difficult to document and test. More recently, Landes (1998) argues that culture, in addition to geography and institutions, should be given pride of place in explaining differences in economic performance.

14. Suppose that instead of Cobb-Douglas, the investment function is a constant elasticity of substitution (CES) function of the underlying investment in each country: $I = (\sum \mu_j I_j^{-\sigma})^{-1/\sigma}$. The elasticity of substitution is $\sigma = 1/(1+\varepsilon)$. In that case the price index P_I is also a CES of the prices of the individual investment goods, of the form $P_I = [\sum \mu_j (\tau_j P_j^*)^{1-\sigma}]^{1/(1-\sigma)}$. Thus if the investment function has an elasticity of substitution of σ , the price index has an elasticity of substitution $1/\sigma$. When σ equals 1, so that the investment function is Cobb-Douglas, the price index is also Cobb-Douglas, with

$$P_I = \Pi v P_j^{a_j} = \Pi v (P_j^*)^{a_j} \tau_j^{a_j}, \quad v = a_1^{a_1} a_2^{a_2} \dots K a_n^{a_n}.$$

The growth rate of the home country depends on a geometric weighted average of transport costs from each of its capital suppliers, with the weight equal to the share of investment goods from j in the total investment expenditure of the home country. When the elasticity of substitution among investment goods is infinite, the CES price index has zero elasticity and takes the form $P_I = \min\{\tau_j P_j^*/\mu_j\}$. What counts in that case is not a weighted average of prices, but rather the lowest price adjusted for the productivity of the various investment goods. The relevance for geography would be as follows: assuming that the efficiency-adjusted price of capital goods is equal in all markets (that is, P_j^*/μ_j is the same across all markets), growth would depend on the minimum distance to one of the capital goods suppliers, rather than the average distance to all the capital goods suppliers. In practice, transport costs would depend on the minimum distance to a major market—the United States, Western Europe, or Japan—rather than the average distance to the United States, Europe, and Japan. In our empirical work we find that the specification of minimum distance to a major capital goods supplier outperforms the average distance to all major capital goods suppliers.

15. Caldwell (1982), in particular, has argued that population pressures are likely to remain high in rural areas while falling sharply in urban areas. According to Caldwell, children represent net economic assets in peasant rural areas because they can assist in household production from an early age, do not generally require high expenditures on education, and can be counted on to care for parents in old age. In an urban setting, however, children represent net economic costs because they are likely to attend school rather than contribute to household production and, because of urban mobility, are much less reliable as social security for elderly parents. Moreover, the opportunity costs of raising children are much higher, especially if women are part of the urban labor force.

16. The standard model of traditional agriculture is based on a neoclassical production function of the form $Q = Q(L, F)$, where L is labor input and F is a vector of other farm inputs, including land. Output per worker falls as L rises relative to F . For the modern sector the increasingly popular model of differentiated production makes Q a function of n intermediate products X_i , each produced with labor L_i , so that $Q = (\sum X_i^\gamma)^{1/\gamma}$, with $0 < \gamma < 1$. Under conditions of monopolistic competition, free entry, and costless introduction of new product varieties, it is typically shown that X_i is fixed by profit-maximizing firms at a given production run x , with $L_i = a + bx$. For a total labor force $L = \sum L_i$, we have $n(a + bx) = L$, or $n = L/(a + bx)$. Production is then $(\sum x^\gamma)^{1/\gamma} = n^{(1/\gamma)}x = L^{(1/\gamma)}x(a + bx)^{-(1/\gamma)}$. The result is that production shows increasing returns to labor. Output per worker is therefore a rising function of L .

17. There is an instantaneous probability, h , of losing office. The probability distribution for tenure in office is then $f(t) = he^{-ht}$, and the mean time in office is $1/h$.

18. The data refer to unweighted country averages for the respective regions for countries for which IMF data are available.

19. The great range of uncertainty about the number of cases is indicative of the lack of concerted study and monitoring of malaria by international organizations in recent years.

20. Because of lack of reliable data on the incidence or prevalence of malaria in the most severely affected countries, our index is necessarily approximate. We digitized a world map of the extent of malaria in 1994 (shown in map 5) from WHO (1997b) and used the geographic information system to calculate the percentage of a country's land area subject to malaria, excluding areas of "some risk." To quantify the differing intensity of malaria, we collected WHO (1992) data for 1990 on the percentage of malaria cases that are the malignant *falciparum* species of malaria, which has the most severe symptoms, is the most resistant to drugs, and is responsible for almost all malaria deaths. The malaria index is the product of the percentage of land area times the percentage of *falciparum* cases. As discussed below, *falciparum* malaria is predictive of low economic growth, but non-*falciparum* malaria is not.

21. The random term ε_i is a function of the initial error term ζ_i , as well as of the intervening sequence of disturbances $\{\mu_{it}\}_{t=0, \dots, T}$. Ordinary least squares estimation would require that the error terms be independent of Z_i .

22. Growth is measured as $(1/25) \times (\ln \text{GDP}_{1990} - \ln \text{GDP}_{1965}) \times 100$.

23. The malaria index at the start of the 1965–90 period is constructed in a manner similar to that for the malaria index for 1994. We digitized a malaria map for 1966 (shown in map 5) from WHO (1967) to calculate the percentage of a country's land area subject to malaria. The malaria index is the product of the percentage of land area times the percentage of *falciparum* cases in 1990. The mix of *falciparum*, *vivax*, and other species of malaria in a given ecozone is not likely to depend much on vector control or public health measures, nor does it change substantially over time. For instance, Sub-Saharan Africa has always had almost 100 percent of the malignant *falciparum*, while the temperate regions that once had malaria had almost no *falciparum*.

24. The first stage regression of ecozones on the change of malaria is

$$\begin{aligned} d\text{Mal6694} = & 0.01 (0.63) - 0.19 (0.99) \text{ Subtropical thorn woodland} \\ & + 0.12 (1.38) \text{ Subtropical dry forest} \\ & - 0.24 (3.29) \text{ Subtropical moist forest} - 1.79 (1.37) \text{ Subtropical rain forest} \\ & - 0.55 (1.33) \text{ Tropical desert} + 0.46 (0.35) \text{ Tropical desert scrub} \end{aligned}$$

$$n = 75, R^2 = 0.19.$$

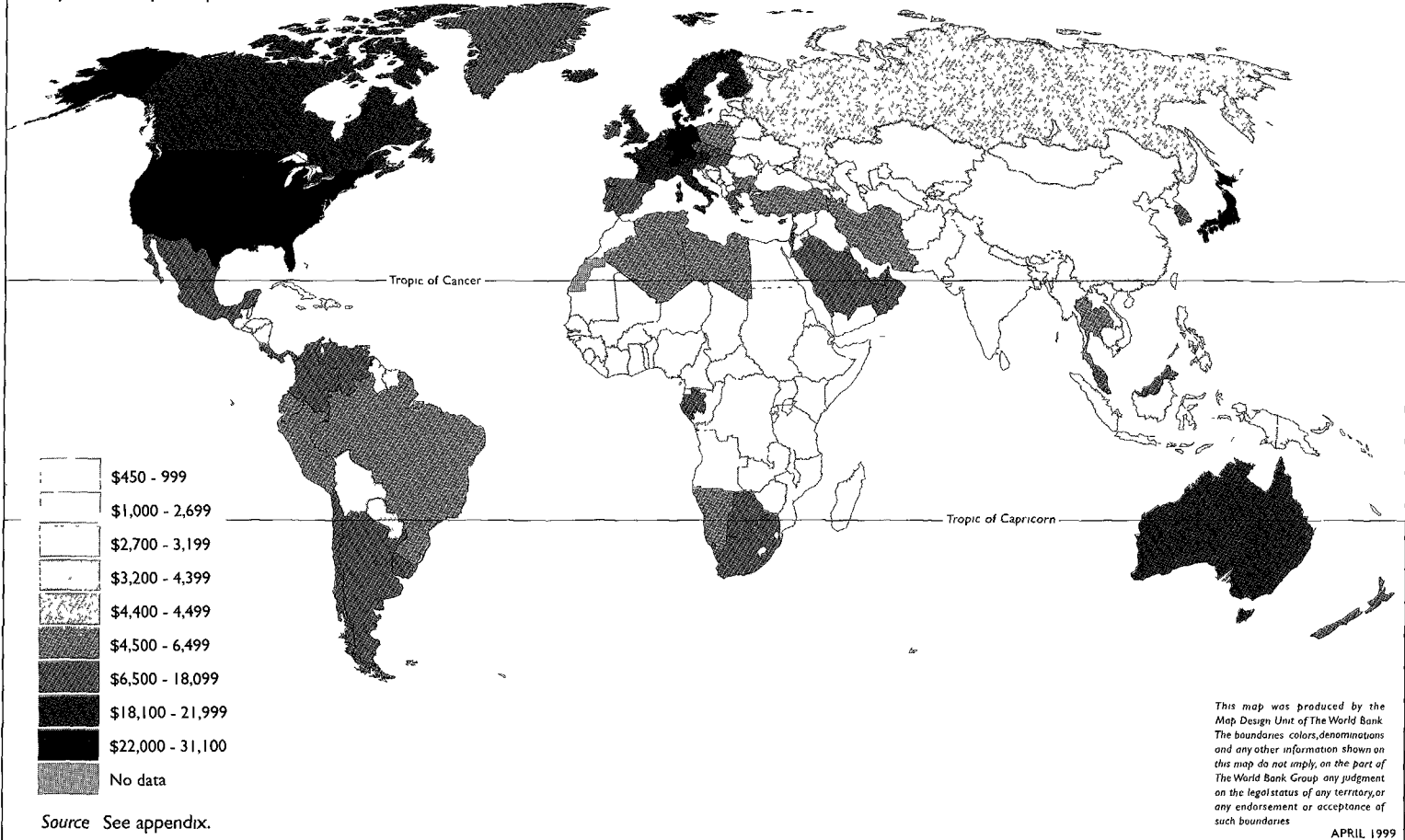
References

- Barro, Robert J. 1991. "Economic Growth in a Cross-section of Countries." *Quarterly Journal of Economics* 106(2): 407–43.
- Barro, Robert J., and Jong-Wha Lee. 1993. "International Comparisons of Educational Attainment." *Journal of Monetary Economics* 32: 363–94.
- . 1994. "Data Set for a Panel of 138 Countries." <http://www.nber.org/pub/barro.lee/ascii/readme.txt>.
- Barro, Robert J., and Xavier Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw Hill.
- Braudel, Fernand. 1972. *The Mediterranean and the Mediterranean World in the Age of Philip II*. New York: Harper & Row.
- . 1981–84. *Civilization and Capitalism, Fifteenth–Eighteenth Century*. London: Collins.
- Britannica Online. 1998. <http://www.eb.com>.
- Caldwell, John. 1982. *Theory of Fertility Decline*. London: Academic Press.
- CIA (Central Intelligence Agency). 1996. *The World Factbook*. Washington, D.C.
- . 1997. *The World Factbook*. <http://www.odci.gov/cia/publications/factbook/index.html>.
- Crosby, Alfred W. 1986. *Ecological Imperialism: The Biological Expansion of Europe, 900–1900*. Cambridge: Cambridge University Press.
- Curtin, Philip D. 1989. *Death by Migration: Europe's Encounter with the Tropical World in the Nineteenth Century*. Cambridge: Cambridge University Press.
- Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. New York: W. W. Norton.
- Encyclopedia Encarta*. 1998. <http://encarta.msn.com>.
- ESRI (Environmental Systems Research Institute). 1992. *ArcWorld User's Guide and Data Reference*. Redlands, Calif.
- . 1996a. *ArcAtlas: Our Earth*. Redlands, Calif.
- . 1996b. *ArcWorld Supplement Data Reference and User's Guide*. Redlands, Calif.
- FAO (Food and Agriculture Organization). 1995. *The Digital Soils Map of the World Version 3.5: Notes*. CD-ROM. Rome.
- Gallup, John Luke. 1998. "Agricultural Productivity and Geography." Harvard Institute for International Development, Cambridge, Mass. <http://www.hiid.harvard.edu/pub/other/geo-dev.html>.
- Grossman, G. M., and E. Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.
- Hall, Robert E., and Charles I. Jones. 1996. "The Productivity of Nations." NBER Working Paper 5812. National Bureau of Economic Research, Cambridge, Mass.
- Jones, E. L. 1981. *The European Miracle: Environments, Economies, and Geopolitics in the History of Europe and Asia*. Cambridge: Cambridge University Press.

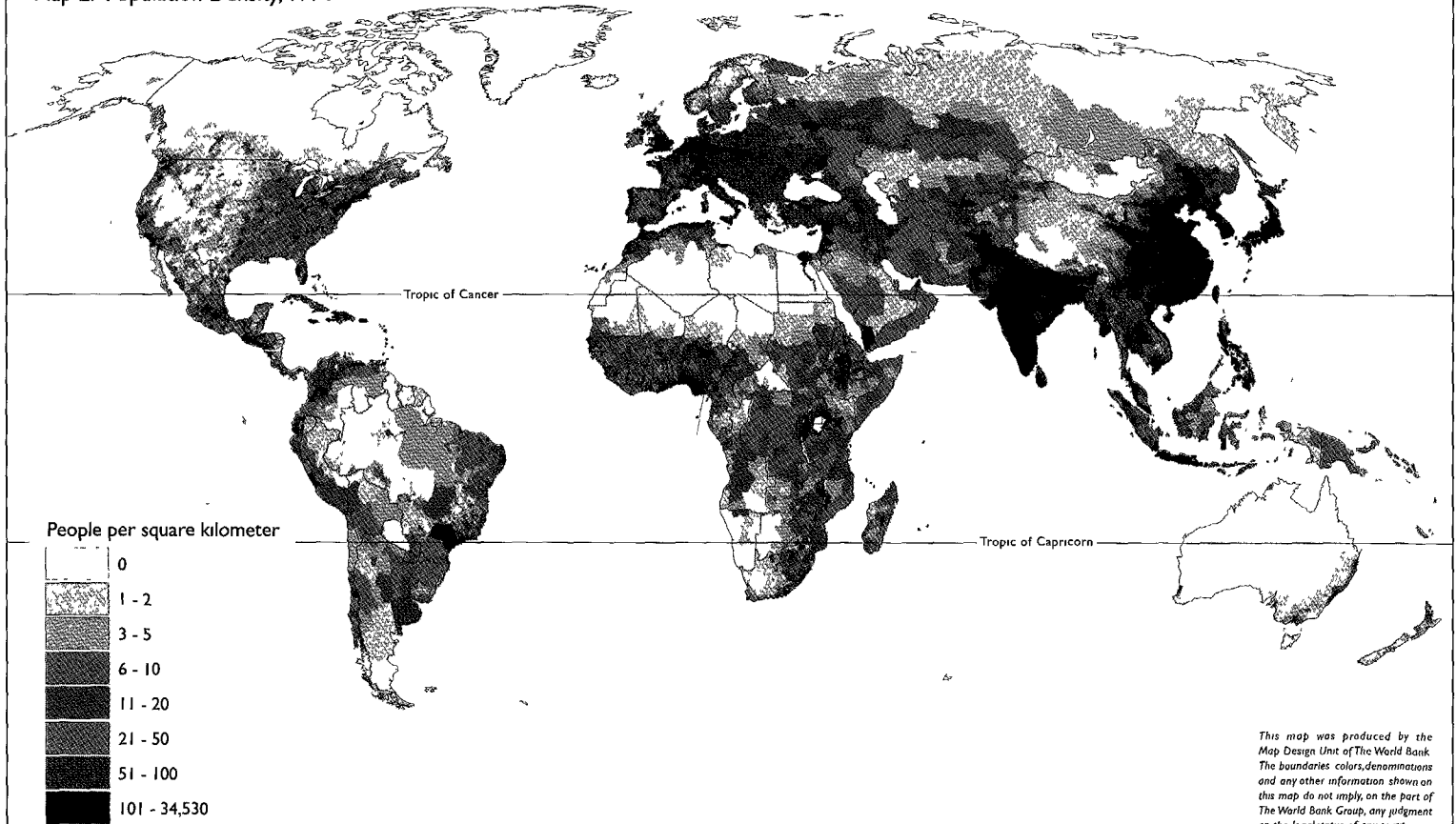
- Kamarck, Andrew. 1976. *The Tropics and Economic Development: A Provocative Inquiry into the Poverty of Nations*. Baltimore, Md.: The Johns Hopkins University Press.
- Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-country Tests Using Alternative Institutional Measures." *Economics and Politics* 7(3): 207–27.
- Kornai, János. 1992. *The Socialist System: The Political Economy of Communism*. Princeton, N.J.: Princeton University Press.
- Krugman, Paul, and Anthony J. Venables. 1995. "Globalization and the Wealth of Nations." *Quarterly Journal of Economics* 110(4): 857–80.
- Landes, David S. 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. New York: W. W. Norton.
- Lee, Douglas. 1957. *Climate and Economic Development in the Tropics*. New York: Harper for the Council on Foreign Relations.
- Leemans, Rik. 1990. "Possible Changes in Natural Vegetation Patterns Due to a Global Warming." Working Paper WP-90-08, IIASA, Laxenburg, Austria. Geographic information system database available at <http://www.grid.unep.ch/matgnv3.html>.
- Maddison, Angus. 1995. *Monitoring the World Economy, 1820–1992*. Paris: Organisation for Economic Co-operation and Development.
- McEvedy, Colin, and Richard Jones. 1978. *Atlas of World Population History*. New York: Penguin Books.
- McNeill, William Hardy. 1963. *The Rise of the West: A History of the Human Community*. Chicago: University of Chicago Press.
- . 1974. *The Shape of European History*. Oxford: Oxford University Press.
- Murray, Christopher J. L., and Alan D. Lopez, eds. 1996. *The Global Burden of Disease*. Cambridge, Mass.: Harvard University Press.
- NGS (National Geographic Society). 1995. *The World: Political* (map). Washington, D.C.
- NOAA (National Oceanic and Atmospheric Administration). 1988. "Data Announcement 88-MGG-02, Digital Relief of the Surface of the Earth." National Geophysical Data Center, Boulder, Colo. Geographic information system database available at <http://grid2.cr.usgs.gov/dem/>.
- Radelet, Steven C., and Jeffrey D. Sachs. 1998. "Shipping Costs, Manufactured Exports, and Economic Growth." Harvard Institute for International Development, Cambridge, Mass. <http://www.hiid.harvard.edu/pub/other/geodev.html>.
- Rand McNally and Co. 1980. *Rand McNally Encyclopedia of World Rivers*. Chicago, Ill.
- Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5): 1002–37.
- . 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, part 2): 71–102.
- Sachs, Jeffrey D., and Andrew M. Warner. 1995a. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1 (August): 1–118. Brookings Institution, Washington, D.C.
- . 1995b. "Natural Resource Abundance and Economic Growth." HIID Discussion Paper 517a. Harvard Institute for International Development, Cambridge, Mass.
- Sellers, P. J., B. W. Meeson, J. Closs, J. Collatz, F. Corprew, D. Dazlich, F. G. Hall, Y. Kerr, R. Koster, S. Los, K. Mitchell, J. McManus, D. Myers, K. -J. Sun, and P. Try. 1997. *An Overview of the ISLSCP Initiative I Global Data Sets*. http://eosdata.gsfc.nasa.gov/CAMPAIGN_DOCS/ISLSC/OVERVIEW_DOCUMENTS/overview.html.
- Smith, Adam. 1976. *An Inquiry into the Nature and Causes of the Wealth of Nations*. 2 vols. Edited by Edwin Canaan. Chicago, Ill.: University of Chicago Press.

- Summers, Robert, and Alan Heston. 1994. *The Penn World Tables, Mark 5.6*. <http://www.nber.org/pwt56.html>.
- Tobler, Waldo, Uwe Deichmann, Jon Gottsegen, and Kelly Maloy. 1995. *The Global Demography Project*. Technical Report TR-95-6. National Center for Geographic Information and Analysis. Santa Barbara, Calif.
- United Nations. 1996. *World Population Prospects 1950–2050 (The 1996 Revision)*. New York.
- WHO (World Health Organization). 1967. "Malaria Eradication in 1966." *WHO Chronicle* 21 (9): 373–88.
- . 1992. "World Malaria Situation in 1990, Part II." *WHO Weekly Epidemiological Record* 67 (23): 169–74.
- . 1997a. "International Travel and Health: Vaccination Requirements and Health Advice—Situation as on 1 January 1997." <http://mentor.who.ch/yellow/welcome.htm>.
- . 1997b. "World Malaria Situation in 1994, Part I." *WHO Weekly Epidemiological Record* 36: 269–74.
- World Bank. 1997. *World Development Indicators 1997 CD-ROM*. Washington, D.C.
- . 1998. *World Development Indicators 1998 CD-ROM*. Washington, D.C.
- WRI (World Resources Institute). 1996. *World Resources Database 1996–97*. Washington, D.C.
- WTO (World Trade Organization). 1995. *International Trade: Trends and Statistics*. Geneva.

Map I. GDP per capita, 1995



Map 2. Population Density, 1994

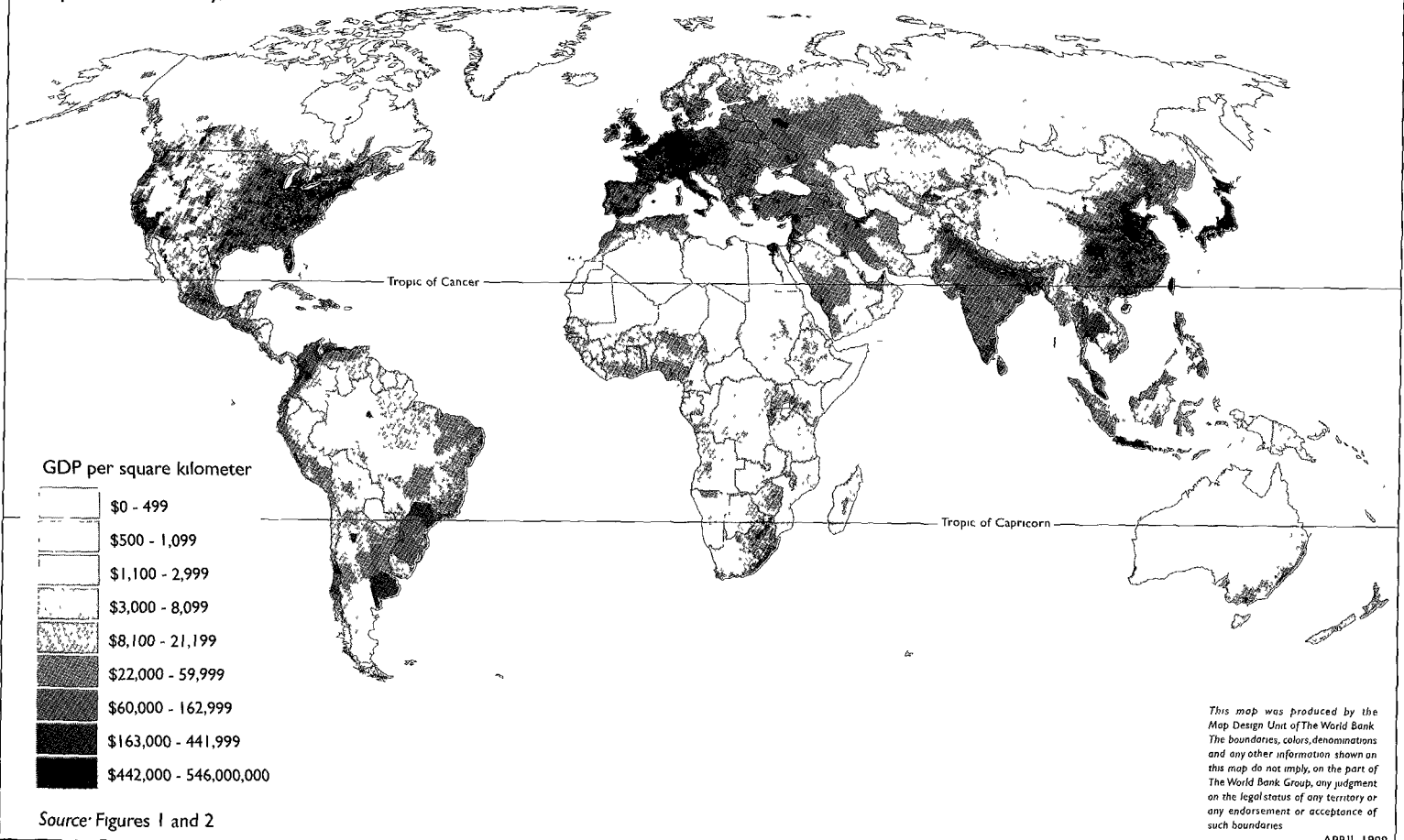


Source. See appendix.

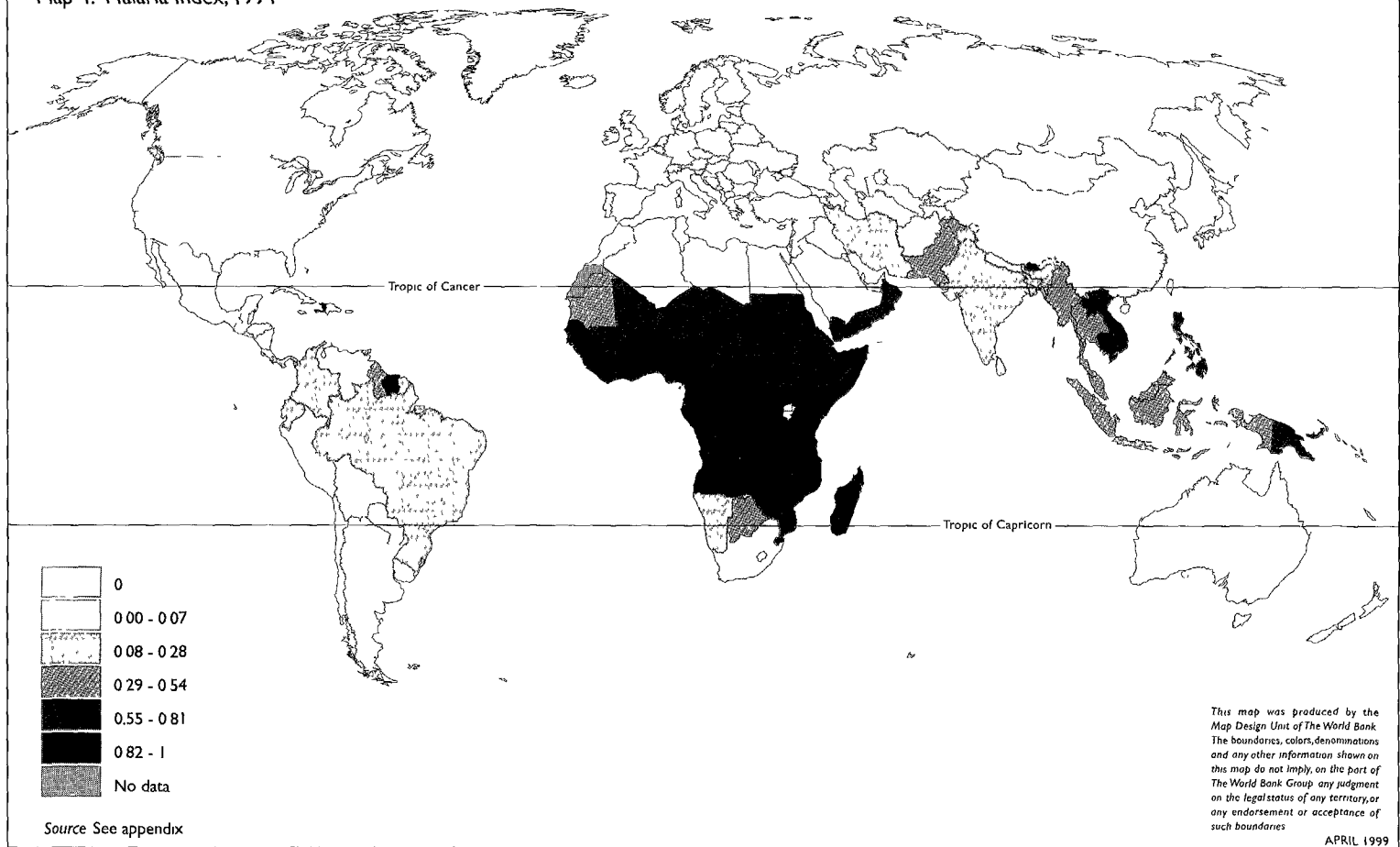
This map was produced by the Map Design Unit of The World Bank. The boundaries, colors, denominations and any other information shown on this map do not imply, on the part of The World Bank Group, any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries.

Map 3. GDP Density, 1995

IBRD 29781

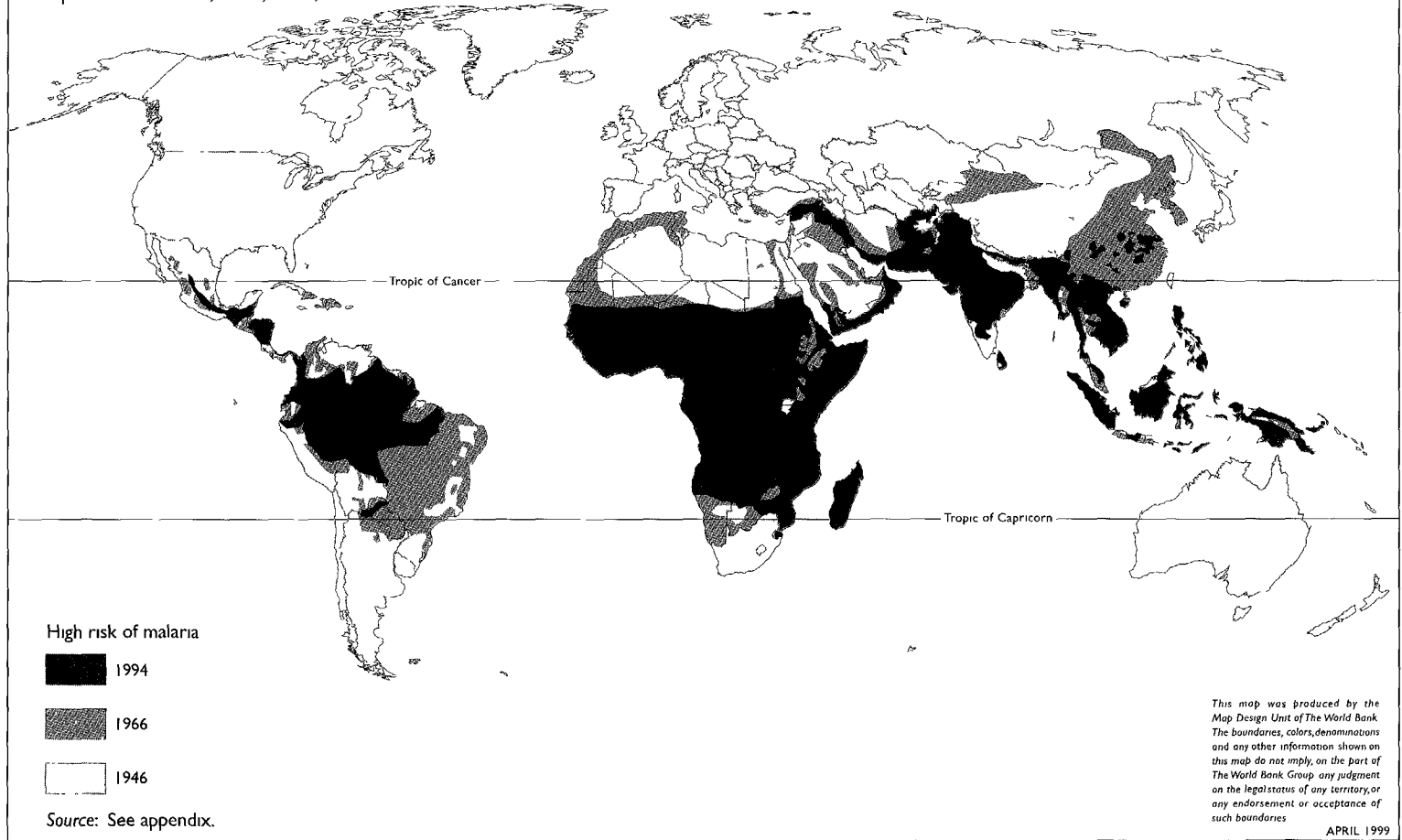


Map 4. Malaria Index, 1994

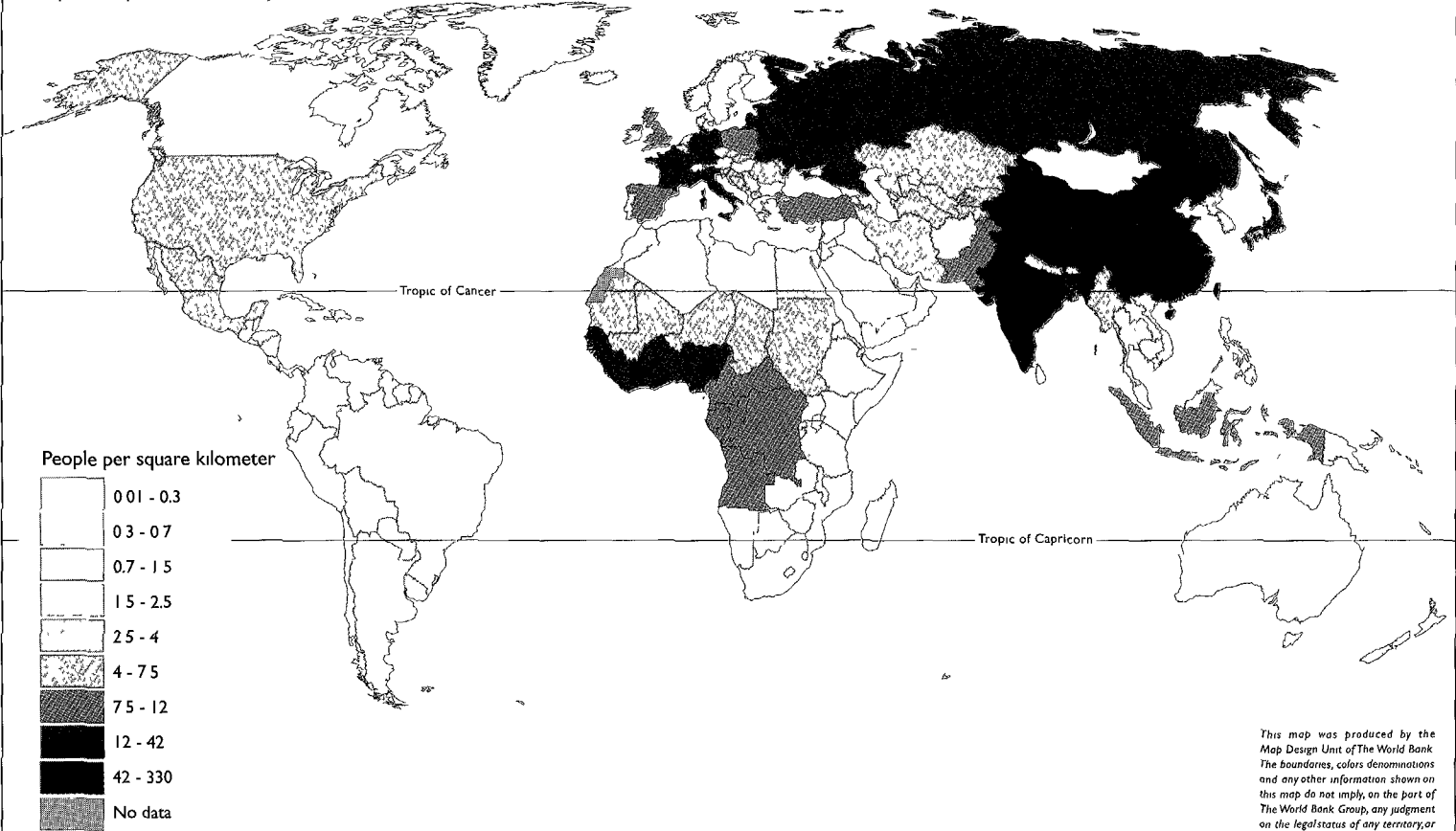


Map 5. Malaria Risk, 1946, 1966, and 1994

IBRD 29783



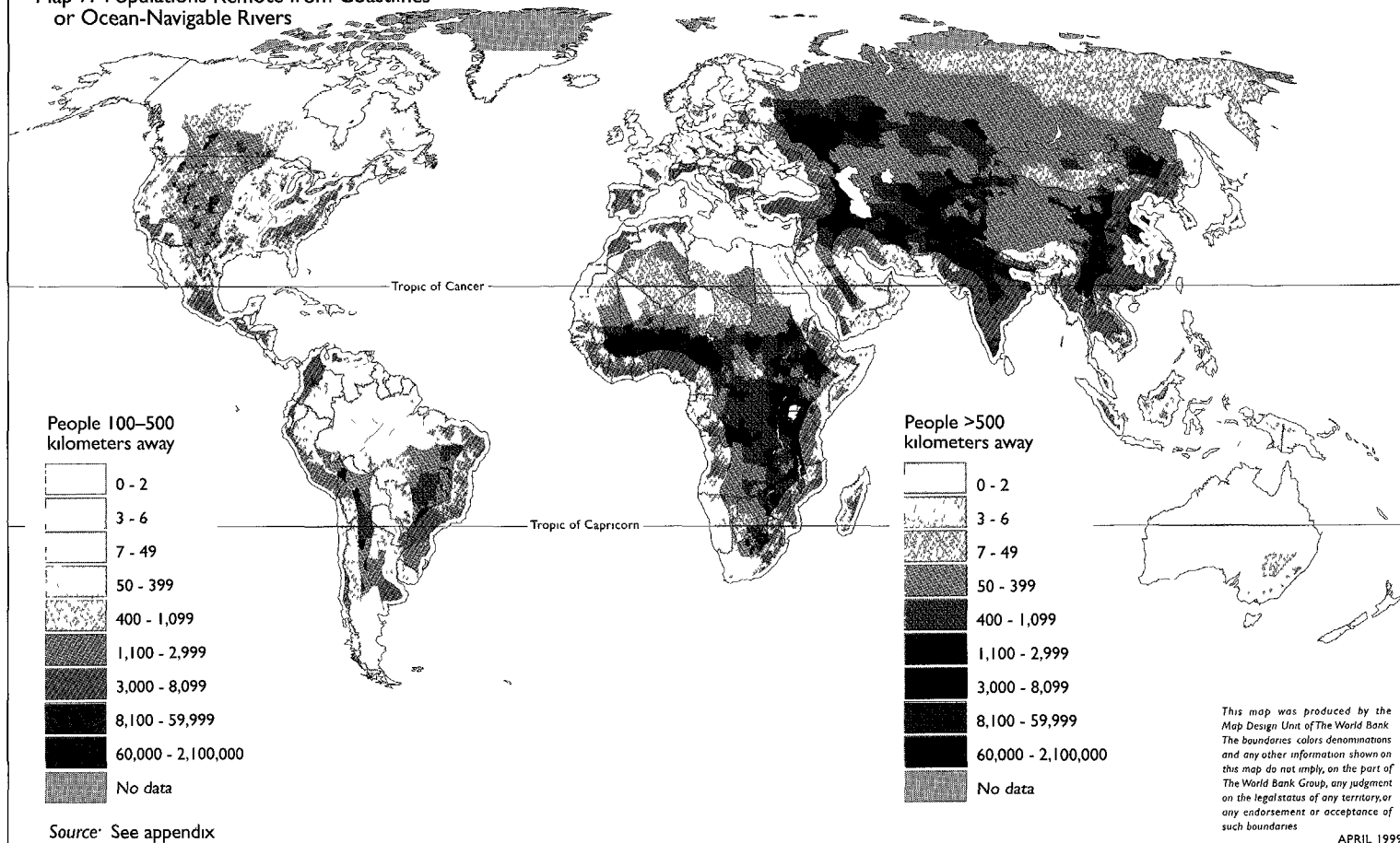
Map 6. Population Density, 1800



Source. McEvedy and Jones 1978.

This map was produced by the Map Design Unit of The World Bank. The boundaries, colors, denominations and any other information shown on this map do not imply, on the part of The World Bank Group, any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries.

Map 7. Populations Remote from Coastlines or Ocean-Navigable Rivers

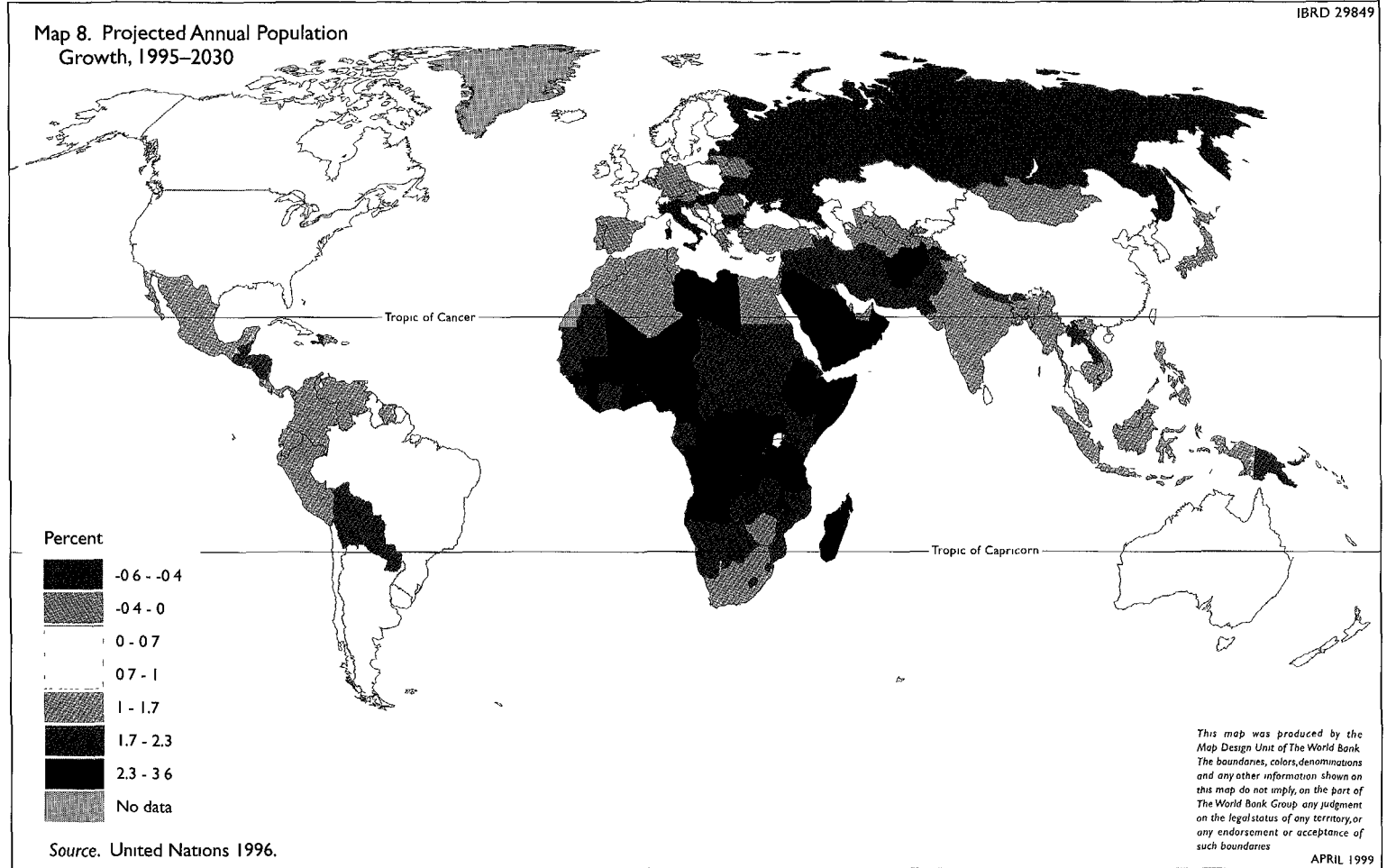


Source: See appendix

This map was produced by the Map Design Unit of The World Bank. The boundaries, colors, denominations and any other information shown on this map do not imply, on the part of The World Bank Group, any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries.

APRIL 1999

Map 8. Projected Annual Population Growth, 1995-2030



This map was produced by the Map Design Unit of The World Bank. The boundaries, colors, denominations and any other information shown on this map do not imply, on the part of The World Bank Group, any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries

Comment on "Geography and Economic Development," by John Luke Gallup and Jeffrey D. Sachs with Andrew D. Mellinger

J. Vernon Henderson

John Luke Gallup, Jeffrey D. Sachs, and Andrew D. Mellinger have written a detailed and provocative article reminding us of the importance of geography in economic development. Modern economics often overlooks the role of geography in development. But geography strongly affects development, and that may have strong implications for public policy.

Main Findings

The key finding of the article is the close relationship between climate and economic growth. The relationship is robust, appearing in comparisons of GDP per capita, in maps, and in growth regressions. The impact of climate on development appears to derive from the relationships between climate and disease, between climate and agricultural productivity, and between location and technology transfer. The results suggest a research agenda linking human capital investment and health policy to economic development in tropical areas.

In addition to climate the article focuses on the importance of long-distance transport costs for growth. The authors describe extremely high transport costs for hinterland regions and cite examples of countries that cannot substitute away from imported inputs with high transport costs. Although I believe transport costs are important, I think that the emphasis on long-distance shipping costs is overdone. First, long-distance shipping costs have dropped dramatically in the past 150 years. Second, in the growth regressions long-distance shipping cost variables have no effect on growth once standard growth determinants (such as investment rates) are controlled for.

Rather than long-distance shipping costs, I believe the focus should be on insufficient investment in transportation and communications infrastructure linking coastal and hinterland areas in developing countries. That deficiency also affects urbanization patterns. Henderson (1997) links urbanization patterns to crude measures such

J. Vernon Henderson is Eastman Professor of Political Economy at Brown University.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

as national road density per kilometer or the ratio of public to private investment. As infrastructure investment increases, the role of large cities declines because both urban primacy (the percentage of the urban population in the largest city) and the percentage of the urban population in large cities fall.

Coastal Urbanization and Growth

The article finds strong relationships among coastal access, the portion of the population living in coastal areas, urbanization, and economic growth. The authors note the problems associated with having high-density agricultural populations far from the coast and suggest that growth could involve mass migration of these populations to coastal cities. The extent to which this is true is open to debate.

Within that debate I will focus on the role of urbanization and aspects of coastal development. The authors find that countries with neither coastlines nor ocean-navigable rivers tend to have less urbanization and less growth. They seem to be suggesting that having more cities causes more growth. The authors attempt to explain national urbanization by the concentration of population on coasts and long-distance shipping costs.

Urban theory and empirical work suggest that the industrialization that is inherent in economic growth invokes a change in the composition of national output. Production shifts in favor of products with low land intensity and strong localized external economies of scale. That switch in the composition of national output and technology fosters urbanization. Urbanization accompanies growth as a desired reconfiguration of population densities that enhances efficiency. More than 65 percent of the variation in national urbanization levels is explained simply by differences in GNP per capita (Henderson 1997).

The real issue, however, is why urbanization in some developing countries is so coastal. If hinterlands contain natural resource deposits and materials, economic development should involve hinterland urbanization. An example is the United States, where population densities are high throughout large portions of the hinterlands in the eastern part of the country. These patterns seem to have developed because of the location of iron ore and coal deposits, as well as timber, natural gas, and petroleum resources. High hinterland population densities thrived in part because of substantial initial investments in canal and rail systems and later in interstate highways linking the hinterlands to the eastern seaboard. Despite a decline in heavy manufacturing and a century of migration west, the patterns persist because the established urban scale in the hinterlands provides fertile environments for location of services and high-tech production. In fact, the key resource of the eastern U.S. hinterland is its people. Access to hinterland resources (including people) provides a geographic challenge to be overcome by infrastructure investment.

Part of this issue is obscured in the article by defining much of the United States (and Europe) as coastal—that is, as being on ocean-navigable waterways. But navigable waterways are generally not just a function of pure geography. They represent

investment in infrastructure and maintenance in canals, dredging, diversion and redirection of services, and so on. There is potential for, say, the Ganges to be navigable through most of the Indian subcontinent. But that would require investment.

Thus the failure of hinterland development in some countries is partly due to insufficient investment in transport and communications infrastructure linking coastal to hinterland regions. An example on a smaller geographic scale is Indonesia, where export-induced industrialization has mostly occurred in key coastal cities, particularly Jakarta. Lack of investment in hinterland infrastructure, as well as the role of government in industrialization, has forced industrial producers to locate in Jakarta. That pattern has induced massive migration, turning Jakarta into a polluted, expensive, and congested megacity. Any extension of highways east and west of Jakarta induces almost instant and massive relocation of industrial producers into the Bobatek hinterland region. Highway extensions in the late 1980s, for example, cut metropolitan Jakarta's share of formal manufacturing employment in the Jabotabek region from 57 percent in 1986 to 44 percent in 1991 (Henderson, Kuncoro, and Nasution 1996).

Success stories of full-scale hinterland development are suggestive. Again on a smaller scale, the Republic of Korea invested heavily in hinterland communications and transport development throughout the late 1970s and the 1980s. The result was massive decentralization of industry throughout all Korean provinces. Between 1983 and 1993 the number of metropolitan areas in Korea rose from 50 to 74 and the share of the three largest metropolitan areas in national manufacturing employment dropped from 44 percent to 28 percent, with the biggest gains in the rural hinterlands (Henderson, Lee, and Lee 1998). On a larger geographic scale, Brazil has invested heavily in hinterland infrastructure for the past 25 years.

Empirical Results on Growth and Density

The population density regressions in the article are quite informative. Using a geographic information system database—something more economists should do—the authors compile a global sample of 14,000 geographic units. They can explain most of the variation in population density through geography (distance to water sources, elevation, disease, soil suitability, ecozones), with current and historical density closely tied through the common link of almost unchanged geography. The authors could comment on how density is estimated—especially in 1800, before satellites and widespread population censuses.

In contrast to the density regressions, the growth and related regressions might be described as somewhat heroic. For related regressions on open trade policy and urbanization, it is unclear which variables should appear on the left-hand side and which on the right-hand side of the equations. For the growth regressions the authors provide results on about a dozen explanatory variables (see table 3), and they surely experimented with another half-dozen or so. But these variables share a common country sample size of just 60, with the maximum sample size rising to 75 with dropped variables. The authors do not use panel methods to enhance the

degrees of freedom, even with the obvious problem of trying to sort out the role of 20 explanatory variables with 60–70 observations. There is also the usual problem that countries with populations of 1 billion people have equal weight in the regressions as countries with populations of 1 million. While the regressions are suggestive and interesting, it is difficult to take them too seriously.

Open Trade Policies

The authors develop an internal political economy model to suggest that coastal countries will choose to adopt more open trade policies than landlocked countries. An alternative is that landlocked countries do not choose closed trade policies. Rather, such policies are imposed on them by their neighbors, who prevent or make costly access to the sea. The problem of being a landlocked country is clouded by the fact that many such countries are small and in marginal hinterland areas, surrounded by powerful neighbors.

Conclusion

Gallup, Sachs, and Mellinger have written a thought-provoking article that should stir debate. I hope that the article and the debate will help policymakers focus on the roles of public infrastructure investments, health policy, and agricultural policies, to overcome the adverse effects of geography on economic growth.

References

- Henderson, J. Vernon. 1997. "Urbanization in Asia." Background paper to the Asian Development Bank's *Emerging Asia*. Brown University, Department of Economics, Providence, R.I.
- Henderson, J. Vernon, A. Kuncoro, and P. Nasution. 1996. "Dynamic Development in Jabotabek." *Indonesian Bulletin of Economic Studies* 32: 71–96.
- Henderson, J. Vernon, T. Lee, and Y-J Lee. 1998. "Externalities, Location and Industrial Development in a Tiger Economy." Brown University, Providence, R.I.

Comment on "Geography and Economic Development,"
by John Luke Gallup and Jeffrey D. Sachs
with Andrew D. Mellinger

Anthony J. Venables

John Luke Gallup, Jeffrey D. Sachs, and Andrew D. Mellinger's article asks whether geography matters for economic development. After exploring the role of geography in determining per capita income levels, growth rates, and the spatial distribution of population, the authors conclude that geography does indeed matter. They find that location and climate have sizable effects on income levels and growth rates and that geographic characteristics explain most of the variation in population density across a fine disaggregation of regions.

The questions being pursued in this research are important, and many of the authors' results are plausible. Indeed, at one level the reader's reaction to the article is that we already know that geography matters. If the null hypothesis is that rich and poor, fast- and slow-growing, densely and sparsely populated countries are randomly scattered around the globe, then we surely know that the hypothesis is rejected. Furthermore, we already know a lot about how geography matters: deserts are less densely populated than temperate areas, and tropical zones in Africa are poorer and grow more slowly than the world average.

Two General Points

Against this background what can we expect to learn from this type of study? I believe that there are at least two things. First, the empirical techniques used by the authors might reveal—and quantify—less obvious empirical regularities than those I have mentioned above. Second, we might hope to learn something about why geography matters.

Turning to the first point, the article offers many interesting findings; I shall mention just a few. The importance of the portion of the population near the coast comes through repeatedly. It is positively associated with per capita income levels and, in some specifications, with growth. And whereas high population density in the hinterlands is associated with slower growth, high population density in coastal regions

Anthony J. Venables is professor of economics at the London School of Economics.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

is associated with faster growth—indicating, the authors suggest, economies of agglomeration in coastal regions. Another important finding is that most of the standard variables used in growth estimates survive the addition of the geographic variables. While location in the tropics, prevalence of malaria, and coastal population are all significant in the growth equations, so too are initial income, education, and openness, as is usually the case in equations of this sort.

Turning to the second point, the reader would like to learn more about why geography matters. Conceptually, there are two very different reasons why geography might be important. One is that some regions may have an absolute disadvantage in their endowments—lack of natural resources, bad climate, poor land quality, low agricultural productivity, propensity to disease, and so on. The other is that a region may be located far from core economic centers. This distance penalty may affect the relative prices of different goods, the relative profitability of different activities, and perhaps the flow of new ideas and technologies into the region.

The distinction is essentially between physical and economic geography, and it is important both for thinking about the likely future spread of economic development and for policy. If geography matters primarily because of distance from economic centers, then a number of observations follow. First, new communications and transport technologies reduce economic distance, as do transport investments and trade liberalization. While the effects of technological changes are not necessarily positive for all poor regions, there is a strong expectation that, by reducing the cost of distance, they will promote development in previously remote regions. Technical progress should then reduce the problem the authors have observed.

Second, we have to ask, distance from where? The location of economic centers is endogenously determined, and new economic centers can develop. Hong Kong (China) and Singapore might have been peripheral locations in 1950, but not anymore. New economic hubs can develop, and regions can develop critical mass in a wide range of economic activities—exactly as East Asian economies have done in recent years. Such possibilities are quite positive for developing countries, because they mean that distance barriers are not insurmountable.

Third, if geography matters primarily because of distance, then there seems to be a fairly clear-cut set of policy implications. Improvements in transport and communications infrastructure, together with trade liberalization to remove artificial barriers to trade, bring regions closer economically and might promote development. But if geography matters because some regions are at an absolute disadvantage, the prospects for these regions are less positive. The degree of geographic determinism should not be exaggerated, however. Although the effects found by the authors are large, technical progress may neutralize a region's disadvantage (say, by eradicating malaria). And as the authors note, policy can play a role independent of a country's geography. Still, the policy implications are different from those outlined above. Policy should be designed to develop activities that are less adversely affected by the disadvantages of the region.

Thus there is a crucial distinction between absolute geographic disadvantage and the costs of distance from economic centers, and this distinction merits further

investigation. Many of the relationships estimated in the article contain measures of both physical geography (location in the tropics, endowments, disease) and distance, and it turns out that both measures are significant in many of the relationships. I would have liked to have seen more attention paid to the distinction and to ways of separating the two mechanisms. I think that doing so will require a good deal more econometric detective work—work that needs to be more closely related to theory.

Specific Comments

Let me now turn to more specific comments. First, the large body of literature on traditional trade theory, new trade theory, and now new economic geography could be drawn on more heavily in the authors' research. I have some problems with the very aggregate model developed in the article. The idea is that distance from the economic core may raise the price of capital goods and intermediate goods relative to the price of output, thereby reducing growth. This may be true, depending on what the output is. If output competes with imports, there is no relative price effect—the price of output will be higher, just as the prices of imported inputs are higher. But if the output is an export, there is a double effect—local firms may have to bear the transport costs on both the price of imported inputs and their exported output. So while I accept that the mechanisms put forward by the authors may be important, proper modeling of them requires a more fully developed theoretical framework than that offered.

Turning to econometrics, I have several comments. It is all too easy for a discussant to call for more econometric tests—but in this case they are necessary. First, the absence of diagnostic statistics makes it difficult to make an econometric assessment of the authors' results. In the equations on GDP per capita, for example, the parameters are quite unstable as Africa comes in and out of the sample (see table 2). This suggests that the model is misspecified—it might not be appropriate to pool Africa with the rest of the world—and the appropriateness of the assumed specification should be tested. Another example: heteroskedasticity is an obvious issue in a dataset of this type, and it should be tested for. Without seeing the results of such a test, it is difficult to gauge the significance of results (and the usefulness of the *t*-statistics).

Second, I am concerned about the exogeneity of some of the variables. This issue is addressed in some of the relationships for GDP per capita, but not in the growth equations. The problems raised by treating investment as an exogenous variable are well known (see, for example, Levine and Renelt 1992). Other variables, such as the malaria variables, also may not be exogenous.

Finally, there is an implicit arithmetic relationship between level equations and growth equations. Equations are presented for the log of levels of GDP at different dates and, as a matter of arithmetic, the first difference between these is simply a growth equation. Yet the level and growth equations are estimated separately, albeit for a somewhat different period. This approach raises conceptual issues and questions about parameter restrictions (due to the relationship between the level and growth equations) that need to be addressed.

The final analytical section of the article switches attention from the determinants of income and growth to the determinants of population density. Econometric analysis is used to identify the geographic characteristics of regions that are more or less densely populated and of regional changes in population density since 1800. Once again, many interesting features are uncovered—for example, the importance of proximity to the coast and, more important, to ocean-navigable rivers.

While the authors' results are a useful way of summarizing the historical record, I am less persuaded about their usefulness for prediction or policy. The relationships estimated are essentially equations for cumulated past birth rates, death rates, and migration, and we know that different regions have had very different historical experiences. It seems to me that for most practical or policy purposes, current birth, death, and migration rates are of much greater interest than are the historical values of these variables cumulated over the past few centuries.

Still, I heartily endorse the authors' conclusion. Geography matters, and more research is needed on the issues raised in the article. In particular, we need to refine our understanding of why geography matters so that we can use the analysis to address policy issues.

Reference

Levine, Ross, and David Renelt. 1992. "Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82(4): 942–63.

Floor Discussion of "Geography and Economic Development,"

by John Luke Gallup and Jeffrey D. Sachs
with Andrew D. Mellinger

J Vernon Henderson and Anthony J. Venables (discussants) both downplayed the importance of intermediate goods and transport costs in economic development, noted Jeffrey D. Sachs (presenter). Yet almost every low- and middle-income country that has experienced rapid growth in the past 30 years has done so through manufactures exports. And almost without exception such growth has involved a specific kind of manufactures exports—those that involve processing intermediate goods and following design specifications imported from the world's core economies. So even though domestic value added tends to rise relative to imports, the process remains closely linked to core area production.

The point, said Sachs, is that whether the exports are shoes, garments, or electronics, the logistics of the operation depend enormously on access to world markets. And with few exceptions, a specific type of economic development occurs—for example, few instances of development have been led by tropical areas. Thus the importance of intermediate goods and transport costs should not be overlooked.

If imported commodities account for 70–80 percent of the cost of final manufactured goods—and in the Subic Bay (Philippines) export zone they account for 85 percent—transport costs have to be extremely low, said Sachs. If transport costs rise by even a few percentage points, profit margins can be wiped out. Exporting areas throughout East Asia—Pusan (Republic of Korea), Subic Bay, Penang (Malaysia), Singapore, Hong Kong (China)—have depended on those few percentage points to integrate with global markets. None of them has achieved integration by producing 100 percent value added in manufactured goods by themselves.

Sachs agreed with Henderson that shipping distance may not be the vital issue. Rather, it is distance over land. For example, all of Korea is coastal, even the interior: most Koreans live within 100 kilometers of the coast. By contrast, the Central African Republic, Chad, Mali, and Niger are almost entirely hinterland—a fact that raises import costs by 40–50 percent. Higher costs mean that these countries cannot compete in manufacturing and so cannot expect export-led growth.

This session was chaired by Jean-Claude Milleron, executive director at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

In the United States, by contrast, about 75 percent of the population lives within 100 kilometers of the coast, the Great Lakes, the Saint Lawrence Seaway (which links the Great Lakes to the Atlantic Ocean), or an ocean-navigable river. With its powerful combination of coastal features and temperate weather, Sachs said, the United States is a miracle—one that is extremely well suited to economic growth.

Jean-Claude Milleron (chair) noted his satisfaction with the discussion. The World Bank's Board of Directors had recently discussed Bank research on the location of activities and the location of people, he said, and it appears that such issues are increasingly important from a global perspective. Some issues, such as unemployment, could even be viewed as a kind of discrepancy between the location of people and the location of activities. Thus, Milleron said, more research is needed on these topics.

Moreover, Milleron continued, it is crucial that tools be created to help developing countries foster productive economic activity—from attracting foreign direct investment to determining the best location for production. The World Bank is well positioned to synthesize the vast store of information on these topics. Moreover, Milleron concluded, additional empirical analysis could be conducted—say, econometric analysis of panel data that were compiled but not used in the article by John Luke Gallup, Jeffrey D. Sachs, and Andrew D. Mellinger.

*Competition and
Regulation Policy
in Developing
Countries*

Regulatory Priorities for Infrastructure Sector Reform in Developing Countries

Paul L. Joskow

In response to the poor performance of state-owned providers of telecommunications, electric power, transportation, and other infrastructure services, many developing countries have implemented or are contemplating major industry restructuring and regulatory reform programs. These programs generally involve privatization, deregulation, and initiatives to increase competition in some segments of these sectors. Yet important segments of most infrastructure sectors remain natural monopolies requiring continuing regulation, and open and nondiscriminatory access by new competitors to the network facilities controlled by these monopolies is necessary for effective competition. Accordingly, the success of infrastructure sector reform partly depends on the creation of effective regulatory institutions. Issues that must be addressed in designing those institutions include establishing regulatory goals, deciding on the structure and organization of regulatory agencies, crafting regulatory mechanisms that balance efficiency and distributional goals, setting appropriate tariffs, establishing the terms and conditions of access by competitors to bottleneck network facilities, and determining the regulator's role in promoting competition. The economic and institutional attributes of developing countries create special challenges for the choice of effective regulatory institutions.

Economical and reliable supplies of basic infrastructure services—electricity, water, telecommunications, transportation—play a crucial role in supporting economic growth and development. Traditionally, many of these services have been viewed as natural monopolies, and supply responsibilities often have been assigned to state-owned monopolies. These infrastructure monopolies often are vertically integrated and provide services that are (arguably) natural monopolies as well as services that are potentially competitive but require access to bottleneck (“essential”) monopoly network facilities to make competition feasible.

Paul L. Joskow is Elizabeth and James Killian Professor of Economics and Management at the Massachusetts Institute of Technology.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

The performance of these infrastructure sectors has been quite poor in many developing countries. Many people lack access to these services or can obtain them only by paying high prices, waiting years to obtain legal connections, or using informal mechanisms such as bribes, theft, and political influence. Technical performance in these sectors is often weak, with widespread equipment outages, large amounts of excess labor, and low productivity. Often, prices are set too low to recover the costs of providing services, and sector enterprises have poor cash flows and encounter difficulties in mobilizing the financial resources needed to maintain equipment and construct additional capacity. The availability and quality of infrastructure services are often highly politicized, and corruption is rampant. Poor infrastructure performance can be a significant drag on economic growth and development.

The reasons for poor performance are complex and diverse. Many developing countries may simply be too poor to finance and support (through user fees) extensive investments in infrastructure, which tends to be capital intensive and characterized by economies of scale and scope. In other cases political considerations have led infrastructure prices to be set too low to cover operating costs and amortize investments. Regulatory and legal institutions can represent yet another source of performance problems if they are not viewed as providing credible commitments to create reasonable opportunities for potential investors to recover their investment costs, including the risk-adjusted opportunity cost of capital. Legal restrictions on foreign direct investment further exacerbate the problems of attracting capital to certain infrastructure sectors. Performance also suffers when corruption is rampant, when property rights and contracting institutions are poorly developed, and when infrastructure sectors are used to pursue social and political goals that siphon off revenues and increase costs.

In an effort to improve performance, many developing countries have implemented or are contemplating major infrastructure sector reform programs. Reforms generally focus on three related issues: privatization, restructuring to promote competition, and regulatory reform. But while privatization, competition, and deregulation are the standard prescriptions for improving the performance of infrastructure sectors, these terms can be misleading. Important segments of most infrastructure sectors continue to be natural monopolies, and competition in the market cannot be relied on to yield satisfactory performance. Moreover, competition for the market—through concession or franchise contracts—must confront problems resulting from significant sunk costs, asset specificity, and incomplete contracts. Many segments in which competition in the market is a reasonable option for allocating resources are likely to be imperfectly competitive and characterized by some degree of market power in both the short and long run. Moreover, the effectiveness of competition will depend on policies governing the initial structure of the competitive segments, the conditions of entry into the market, and the price and nonprice terms and conditions of access to bottleneck monopoly network facilities for competing suppliers.

These considerations imply that simplistic privatization and “complete” deregulation policies are unlikely to be realistic or effective policy options in most infra-

structure sectors. Certain segments of most infrastructure sectors should and will continue to be subject to some form of price and entry regulation, because the services they supply have natural monopoly characteristics. The attributes of the regulatory institutions that are developed will affect both the performance of the residual monopoly segments and their ability to provide a platform for competition to supply complementary services that are not subject to price and entry regulation.

This article discusses issues related to the design of regulatory institutions accompanying privatization, restructuring, and expansion of competitive opportunities in key infrastructure sectors that were previously legal monopolies (in most cases, state-owned monopolies). The discussion focuses on telecommunications, electric power, and natural gas transportation systems, but some conclusions—especially as they relate to the regulation of residual natural monopoly services—have implications for railroads, water supply systems, intercity highways, and other infrastructure services. While there are lessons to be drawn from the recent reform experiences of industrial countries, these lessons cannot be applied mechanically to developing countries. Differences in development and performance in each infrastructure sector—as well as differences in basic market institutions, in income distribution and other social concerns, in legal and political institutions, and in the economic attributes of different infrastructure sectors—must be taken into account in designing the new industrial organization and regulatory institutions that will govern each sector. Not only does one size shoe not fit all feet, but trying to apply a one-size-fits-all approach can severely limit the performance improvements that sector reforms can achieve.

The Standard Reform Prescription for Regulated Infrastructure Sectors

The standard public policy prescription for reforming vertically and horizontally integrated natural monopoly industries is fairly straightforward. It has been or is being applied, in varying degrees, to telecommunications, natural gas, electric power, water distribution, and railroads around the world. Its basic elements include:

- *Privatizing state-owned enterprises to create hard budget constraints and depoliticize the sector.* Meaningful evidence is scarce on the comparative performance of public and private enterprises. In most countries, especially developing countries, basic infrastructure sectors have been publicly owned for many years, so cross-sectional data do not provide a good natural experiment that controls for differences in supply opportunities and demand attributes.¹ Where meaningful comparisons can be made, however, the weight of evidence favors private over public enterprises (Boardman and Vining 1989; Vickers and Yarrow 1988; World Bank 1997). Privatization helps depoliticize infrastructure sectors, limiting (though not eliminating) their use in pursuing social policy goals (such as employment and price stability) that are not directly related to the services they supply. Privatization

also decouples financial resources for maintaining equipment and financing new investments from the government's general budget and fiscal situation. Governments find it harder to manipulate a private firm that has private investors and debt holders and is subject to accounting and auditing standards necessary for its securities to be listed on major exchanges. When the state wants to retain ownership of infrastructure sectors, it should consider commercializing the enterprises and requiring them to be self-financing, rather than situating them under government ministries or departments.

- *Fostering competition in potentially competitive segments that historically have been integrated with natural monopoly segments in the same organization.* Historically, "natural monopoly" sectors have typically evolved with organizations that are composed of both potentially competitive segments (such as long-distance telephone service, electricity generation, production of natural gas, railroad rolling stock) where "competition in the market" may be an effective way to allocate resources, and natural monopoly segments (such as local telephone exchanges, electricity transmission and distribution networks, natural gas pipeline networks, and railroad track and switching networks) where competition in the market is not expected to yield good performance. It is often argued that vertical integration between natural monopoly segments and potentially competitive segments leads to an unnecessary expansion of monopoly from one horizontal level (the natural monopoly level) to other (potentially competitive) levels and has extended inefficient regulation to segments where market forces can and should govern better.
- *Unbundling prices for competitive services from prices for access to and use of complementary natural monopoly services to ensure effective and efficient competition.* More important, competing suppliers of competitive services must be given access to the regulated bottleneck natural monopoly segments that they require to compete. Access should be based on comparable price and nonprice terms and conditions that do not unreasonably discriminate between competing suppliers and, in particular, do not give more favorable access to the operator of the regulated monopoly network if it offers services in the competitive segments.
- *Recognizing the potential competitive problems created by vertical integration.* These problems derive from the ability of owners of natural monopoly segments to favor competitive suppliers in which they have a financial interest and discriminate against those in which they do not. If vertical integration is allowed, then effective rules are needed to ensure that unaffiliated competitors have comparable access to the bottleneck network facilities. Developing and enforcing such rules is difficult, both in theory and in practice.
- *Ensuring that customers pay fair prices for services sold in competitive segments by securing an adequate number of competing suppliers or imposing constraints on prices charged by actual and potential suppliers.* If the current ownership structure is too concentrated and entry is not an effective con-

straint on prices, the transition to competition may require horizontal separations to create more competitors or the continuing regulation of dominant firms. A reasonable level of competition, not perfect competition, should be the goal, recognizing that the costs of imperfectly competitive markets are often lower than the costs of necessarily imperfect regulation.

- *Replacing traditional cost-of-service regulation of residual monopoly services with incentive- (or performance-) based regulation.* For example, in the United Kingdom price caps are used to regulate the prices that gas and electric distribution companies can charge for “pipes” and “wires” services (see below). Alternatively, competition for the market rather than competition in the market—through the use of concession or franchise contracts with incentive provisions and suppliers selected through competitive bidding—may be used as an alternative to traditional price and entry regulation.
- *Giving retail customers direct access to wholesale markets for competitive services (such as natural gas supply) when it is technologically and economically feasible to do so.* Retail customers should be offered unbundled prices and services from the incumbent, vertically integrated, or bundled supplier and have access to the bottleneck network services required to consummate transactions with competing supplies of competitive services. The obligation traditionally placed on public utilities to plan for and supply all customers located in a designated franchise area with a full bundle of services should wither away as customers gain the ability to arrange for their own supplies in a competitive market. If suppliers of infrastructure services continue to be used as vehicles for financing social programs (universal service obligations, special low-income rates, Internet access for public schools), the associated costs must be clearly identified and included in nonbypassable access charges paid by all consumers if such programs are to be sustainable in the face of competition (Posner 1971).

Choice of Reform Models

Privatization, restructuring, deregulation, and competition are not economic goals (though they may be political goals in support of democratic governments). Rather, they are means for achieving economic goals. Accordingly, the first step in developing effective regulatory institutions and competition policies is to clearly define the public policy goals for each infrastructure sector, given its current and expected levels of performance under prevailing institutional arrangements. These goals will naturally include:

- Improving the sector’s ability to mobilize adequate financial resources to support the investments required to balance supply and demand efficiently.
- Increasing sector productivity by reducing operating costs and investing in more efficient plant and equipment.
- Bringing prices in line with costs to provide consumers with good price signals.

- Balancing aggregate revenues and costs so that the sector's capital needs can be supported by sector cash flows.
- Adjusting the prices charged for sector services so that they are compatible with the introduction of competition in competitive segments (tariff rebalancing), where prices and entry are to be deregulated and competition is to govern the allocation of resources.

The second step is to define the basic structure of the reform model that will be adopted for each sector. The specification of a reform model must go well beyond simple slogans such as privatization, deregulation, and competition. The model should define which segments of each sector will continue to be regulated monopolies and whether regulation is expected to be permanent or transitional (anticipating future deregulation of prices and entry). The regulation model should also define which segments will be opened to competition, the form of the competitive mechanisms that will be introduced, how quickly competition will be phased into the competitive segments, which segments will be privatized, and how any remaining state-owned enterprises will be organized. The specification of which segments of each sector will be opened to competition and which will continue to be regulated monopolies is a choice variable, as is the timing of the introduction of competition. In telecommunications, for example, competition can be opened up in international, long-distance, and local markets. It can be facility-based competition only or nonfacility-based competition, with competitors having the option of leasing from the incumbent network operator some or all of the network facilities they require to compete for consumers, rather than investing in parallel competing network facilities.

The third step is to define an industrial organization for the sector—that is, the extent of horizontal integration at each level of the production chain and of vertical integration between levels—that is consistent with the chosen reform model. For segments that are to remain regulated monopolies, should there be one large monopoly serving the entire country or several regulated monopolies serving specific areas? This decision has implications for the effectiveness of regulatory instruments (for example, for the use of some form of yardstick regulation) and potentially for the intensity of competition in segments that are to be opened to competition.

For example, say that the reform model anticipates that electricity and gas distribution entities will purchase electric power and natural gas on behalf of retail consumers rather than act as passive wires and pipes companies through which retail consumers make their own purchasing decisions. In that case an industrial organization with multiple distribution companies may be desirable to avoid monopsony problems in competitive segments and to provide comparative purchasing information for regulators of the distribution entities. This decision must in turn be informed by considerations of economies of scale, measured over the geographic expanse of the franchise to be served by the regulated monopoly. Some developing countries may have markets that are too small to support multiple regulated distribution monopolies serving different regions without sacrificing significant economies of scale.

In competitive infrastructure segments the number and attributes of incumbent competitors and the conditions of entry will have important implications for the intensity of competition and the associated inefficiencies and distributional consequences of any significant market power problems in these segments. For example, electricity restructuring in Argentina created a large number of private generating companies, and competition at the generation level has been intense. By contrast, electricity restructuring in England and Wales created only two large fossil fuel generators, and this initial market structure decision has led to market power problems in the unregulated generation market.

Specification of the extent of vertical integration raises more difficult problems. In many network infrastructure industries there are good reasons to expect economies of vertical integration between natural monopoly segments and potentially competitive segments. In this area policymakers have three choices:

- They can retain a vertically integrated regulated monopoly without competition in potentially competitive segments, to preserve economies of vertical integration.
- They can allow the incumbent to remain vertically integrated but also encourage new entrants to supply potentially competitive services by giving them access to the incumbent's monopoly network at fair prices.
- They can separate vertical ownership and control of suppliers in regulated natural monopoly segments from suppliers in competitive segments, giving all suppliers in the competitive segments comparable access to monopoly network facilities.

The first option is the status quo in most countries before privatization and deregulation. The second option is the approach taken in telecommunications restructuring in Europe and Japan, in electricity restructuring in many parts of the United States, and in natural gas reform in the United States, and it is the path that telecommunications reform is supposed to take in the United States. The third option is the approach taken in electricity restructuring in England and Wales, where monopoly transmission and distribution segments have been separated from the competitive generation segment.² On the other hand, in England and Wales the distribution business is separated into two components: a "wires" or pure electricity transportation business and a "supply" or retailing business that involves buying energy from generators and reselling it to retail consumers in the wholesale electricity market. While the supply business is competitive, the regulated monopoly distributors are allowed to participate in it through unregulated affiliates, in competition with independent retail electricity suppliers that also need access to the distributor's bottleneck monopoly distribution facilities. Rather than barring distributors from the competitive retailing business, cost and functional separation rules ("ring-fencing") have been put in place to prevent distributors from using their control of distribution wires to favor their supply affiliates. A similar structural model for electricity has been adopted in Argentina, New Zealand, and Norway, with variations in the extent to which vertical integration is permitted. The initial concept for telecommunications restructuring in the United States clearly separated the incumbent companies allowed to provide

local exchange service from those authorized to provide long-distance service. The local exchange carriers were, however, permitted to provide other services that made use of their local network facilities.

The second option appears to offer the benefits of competition without the potential costs of lost economies of vertical and multiproduct integration. But two things need to be kept in mind when comparing the second and the third options. First, successfully implementing the second option requires considerable regulatory effort to ensure that the terms and conditions of access to the network do not discriminate against the network operator's competitors. Second, the open network access rules often implemented to deal with potential vertical foreclosure problems may end up requiring the incumbent to behave as though it were not vertically integrated. For example, the rules could require the incumbent to separate the operations and accounts of its competitive and regulated monopoly businesses and restrict private communications between the regulated operator of the bottleneck network and its unregulated affiliates. In this case it is not clear that the potential benefits of vertical integration can be realized.

Once the basic industrial structure for the industry is defined, the next step is to decide how the reform model will be implemented. Under a "big bang" approach privatization, restructuring, and the introduction of competition occur at the same time. Electricity restructuring in Argentina and in England and Wales took this approach with fairly short transition periods. Another approach is to provide for a relatively long transition period, during which the industrial organization and associated regulatory institutions are allowed to evolve according to a planned transition program. For example, in Peru telecommunications reform is proceeding in stages. The incumbent telephone enterprise was privatized with a monopoly on local (except cellular), long-distance, and international service and regulated based on a prespecified regulatory contract for several years. At the end of the transition period competing suppliers of intercity and international services will be allowed to enter the market. Electricity restructuring in Colombia also took a gradualist approach but failed to clearly specify how it would proceed over time (Gray 1997). Thus it is important to distinguish between clearly structured transition programs and unstructured gradualism. Unstructured gradualism in infrastructure reform, as has also been practiced in Japan, is not a desirable approach.

The choice between the big bang approach and the transition approach largely depends on six factors: the performance of the existing system, the complexity of implementing a big bang approach given pricing and other imperfections that cannot be fixed instantly, the capacity of legal and political institutions to support competitive markets for infrastructure services, the speed with which reasonably competitive markets can evolve, the time required to create effective regulatory institutions, and the government's ability to credibly commit to a restructuring framework that supports private investment and competitive entry. Whichever model is followed, there is a role for effective regulation because there will almost always be residual services that have natural monopoly characteristics. The span of

regulatory responsibilities will vary depending on the sector and the reform model selected. But the basic regulatory issues will be the same.

Competition for the Market and Competition in the Market

Some analysts argue that infrastructure industries can be restructured in a way that completely eliminates the need for price and quality regulation. The key to this argument is the distinction between competition for the market and competition in the market. In natural monopoly segments competition can theoretically be exploited by putting up for competitive bidding—using medium- or long-term franchise or concession contracts—the exclusive right to provide the service with natural monopoly characteristics.

Those who take this view have been reading too much Demsetz (1968) and not enough Williamson (1976) and Goldberg (1976). Franchise bidding works fine for trash collection, ambulance service, fire protection, and other local services where the relevant assets are not specific to a particular area, service quality attributes can be readily defined, and franchise contracts are of short duration and can easily be rebid without significant price penalties. But the experience with “pure” franchise contracts for water and sewer systems, highways, and rail networks (tracks) is not particularly good (van den Berg 1997; Crampes and Estache 1997; Ruster 1997). Moreover, many regulated monopoly infrastructure sectors in industrial countries evolved from enterprises that had franchise contracts that were not sufficiently flexible for them to adapt to changing economic conditions.

The fundamental problems with relying on concession contracts or franchise bidding for infrastructure sectors with natural monopoly characteristics are the importance of significant long-lived immobile capital investments in these sectors, the incompleteness of contracts (and associated potential ex post holdup problems), and contract renegotiation and adaptation problems. Long-term franchise and concession contracts ultimately need to be adjusted over time, and the contract administrator has essentially the same job as a regulator (Goldberg 1976). The bottom line is that competition for the market does not alter the fundamental conclusion that good regulatory institutions are crucial to improving the performance of the natural monopoly segments of infrastructure sectors and to supporting the introduction of competition in the competitive segments of these sectors.

Universal Service, Income Distribution, and Taxation by Regulation

One of the standard criticisms of infrastructure sectors in both industrial and developing countries is that tariff structures and service connection and supply expansion policies have been widely used to provide subsidies and cross-subsidies that benefit some groups and burden others. For example, in telecommunications it is almost universally the case that prices for long-distance and international calls have been kept high (far above marginal cost) so that these services can help cover the fixed costs of the local network, allowing prices for local calls to be kept low. This policy has been

justified as promoting universal service, which can be viewed as desirable based on assumed network externality problems (the value of the network increases as more people can be reached over it) or simply on the broader social aim of giving all of a country's citizens access to basic infrastructure services at affordable prices.

In many countries—such as France and Spain—social and economic policy makes electricity tariffs uniform across the country despite regional variations in supply costs. Most countries have directly and indirectly subsidized electricity and telephone services in rural areas. In some countries electricity tariffs either expressly provide lower prices for low-income consumers or tilt tariffs so that consumers with low consumption pay prices below the cost of service to them (under the assumption that there is a high correlation between income and electricity consumption). Many countries have used special electricity tariffs to attract or retain electricity-intensive industries nationally or regionally. In developing countries electricity, telephone, and railroad tariffs have often been held down in response to political pressures. Widespread illegal connections to electricity distribution networks and local telephone systems and the failure of consumers to pay for services are forms of subsidy that arguably tend to benefit the poor.³ Redistributive goals are also reflected in regional or national policies that favor suppliers of local fuel (primarily coal) and domestic equipment manufacturers and construction companies or that finance certain research and development activities. Thus subsidization and cross-subsidization extend well beyond defined universal service goals or the desire to help the poor pay for modern infrastructure services.

The result is that regulated monopolies have been used to finance and implement a wide range of implicit taxes and subsidies. A regulated monopoly is an effective (though not necessarily efficient) vehicle for such policies for two reasons. First, monopoly status makes it impossible (or at least extremely difficult) for competitors to undermine the price increases needed to finance direct and indirect subsidy programs. Second, the magnitude and nature of subsidies and cross-subsidies are buried in complex tariff-setting policies and resource acquisition programs that are insulated from public scrutiny. This system of hidden taxes and subsidies makes it easier for the interest groups benefiting from them to sustain them.

The future role of infrastructure sectors in supporting some or all of these redistributive goals should be resolved as part of privatization, restructuring, and regulatory reform. As Richard Posner (1971) pointed out many years ago, competition undermines the ability to use regulated monopoly sectors as vehicles for taxing and subsidizing. Indeed, in some infrastructure sectors—particularly telecommunications and freight transportation—the cross-subsidies reflected in regulated tariffs have created pressures to allow competitive entry because potential entrants saw business opportunities in competing to supply services in segments where regulated prices were higher than the associated stand-alone costs of supply. The kinds of direct and indirect subsidies that exist today will not be sustainable in the competitive segments of infrastructure segments (electricity generation, commodity natural gas, long-distance telephone service), and government efforts to sustain them in the face of competitive entry will distort the nature and direction of competition.

The standard prescription for restructuring and regulatory reform is to remove subsidies and cross-subsidies from infrastructure sectors. The argument is that if these subsidies are socially desirable, it is better to shift the responsibility for raising revenue and distributing subsidies to the general tax system and budget process. Tariff-based subsidies that are considered necessary to help the poor are often poorly targeted and may even be regressive. More important, taxation by regulation corrodes democratic institutions, encourages corruption, undermines the system's ability to produce adequate revenues to cover the costs of supply (as interest groups press to keep tariffs low), and increases supply costs by favoring (domestic) firms even when they supply inputs at high cost or low quality and by using infrastructure sectors as employment sinks. Finally, once it is recognized that competition will undermine existing subsidies and cross-subsidies, interest groups that benefit from them, as well as incumbent regulated firms that fear they will have difficulty rebalancing tariffs to respond to competition, will have powerful incentives to oppose the introduction of competition. In short, taxation by regulation tends to be inconsistent with standard universal service goals, leads to costly inefficiencies, and creates a political environment that is resistant to reform.

True, many industrial countries have managed to develop quite nicely while using electricity, telephone, water, and transportation prices to subsidize poor and rural consumers and to finance the expansion of infrastructure services to regions where initial demand was insufficient to support the associated costs. Such policies reflect long-term economic development and social integration goals as well as political constraints. Thus cross-subsidies and taxation by regulation may not be ideal but are also not fatal to economic development. The challenge is to respond to widely shared income redistribution goals in a way that does not significantly undermine the efficiency of the infrastructure sector, that is reasonably transparent, and that makes it difficult for taxation by regulation to be abused by interest groups.

The primary rationale for using infrastructure tariffs to raise and redistribute revenues should be that they raise taxes more efficiently and redistribute income more effectively than the general tax and expenditure system. If this is not the case, the end of taxation by regulation is a legitimate and highly desirable public policy goal. Either way, effective infrastructure reform makes it necessary to identify subsidies and cross-subsidies in each sector and decide whether to end them, shift them to the general tax and expenditure system, or maintain them. The introduction of competition means that subsidies can no longer be financed by increasing the prices incumbent firms charge for competitive services to levels far above their stand-alone costs. Competitive suppliers will just enter under the umbrella created by the high prices and steal customers from the incumbent. For subsidies and cross-subsidies to be sustained in the face of competition, they have to be reflected in surcharges to prices charged for services provided by the remaining monopoly segments (electricity, water, and natural gas distribution, local access to the telephone network). To help ensure that subsidies are contained and reflect shared social values, subsidy amounts should be clearly measured and identified as sepa-

rate surcharges (taxes) on bills sent to consumers. This makes them more transparent and visible to the public.

Organization and Procedures for Regulatory Agencies

The standard prescription for designing good regulatory institutions typically includes the terms *independence*, *transparency*, *accountability*, *expertise*, and *credibility*. These terms convey important principles, the implementation of which must be sensitive to the legal and political institutions of different countries.

The concept of an independent regulatory agency pursuing its goals with open and transparent procedures is familiar only in Canada, the United States, and a few other countries. But even in the United States, where the concept of an independent regulatory commission has had nearly a century to evolve, regulatory agencies are not completely free from political pressures, and their information and staff are often inadequate for them to do their jobs effectively.⁴ The concept of an independent regulatory agency operating with transparent procedures, obligated to make decisions based on a logical evaluation of the facts in light of its statutory responsibilities, and subject to judicial review may be completely alien to countries where policy decisions are made behind the closed doors of government ministries and departments, with limited due process, transparency, or judicial review. Moreover, it necessarily takes time to create regulatory institutions and to allow them to mature, and their independence and credibility are established on the basis of both their legal foundations and their actual behavior when faced with difficult decisions that involve substantial interest group controversy (Gilbert and Newbery 1994).

In light of these considerations, I can offer only general guidelines on the attributes of good regulatory institutions. These guidelines apply in a variety of ways to individual countries, based on their economic situation and legal and political institutions (Levy and Spiller 1994). First, however, it is useful to examine why independence, transparency, accountability, expertise, and credibility are important guiding principles for the creation of regulatory institutions.

The segments of infrastructure sectors that will be served by a monopoly enterprise subject to public regulation are typically capital intensive and require significant investments in long-lived, highly immobile capital facilities. Because the prices charged by the regulated enterprises are subject to regulation, investors are naturally concerned that once the investments are sunk, the government will have an incentive to “hold up” the firm in order to redistribute quasi rents to influential interest groups. In the absence of credible commitments that rules for price setting will provide a reasonable opportunity for suppliers to recover their investment and operating costs, including an appropriate return on the investments, it is more difficult and more costly for an enterprise to attract capital. Similarly, investment decisions may be distorted to reduce the potential cost of holdups.

The lack of credible commitments is a special problem in developing countries because many lack a strong tradition and legal framework for respecting and enforcing private property rights and contracts. Moreover, many developing country gov-

ernments have allowed their infrastructure sectors to deteriorate rather than set prices high enough to cover the costs of supplying services. At the same time, in the absence of effective regulation the authorization of a monopoly supplier could lead to the exploitation of consumers through excessive prices and poor service. In addition, potential entrants into competitive segments that require access to the incumbent's network facilities to provide service are concerned about unreasonable restrictions on access. In other words, both efficiency and rent distribution issues arise when granting an enterprise a legal monopoly, and these issues must be addressed in the design of regulatory institutions.

In principle all these issues could be addressed in a straightforward manner if the regulator were perfectly informed and could write an enforceable and complete contingent claims contract that specified ex ante the terms and conditions of services supplied by the regulated monopoly. But regulators cannot be perfectly informed, and it is impossible to write a credible and complete contingent claims contract that defines ex ante how regulated firms are expected to behave and what they will be paid under all possible contingencies. Accordingly, regulatory institutions must provide credible commitments that investors will not be held up once their investments have been sunk, that consumers will be protected from excessive prices and poor service resulting from supply-side inefficiency or the exercise of monopoly power, and that other goals for the sector (such as universal service) will be achieved. These institutions must also be sufficiently flexible to adapt to changing economic conditions without undermining commitments to investors. With these considerations in mind, I offer some thoughts about the organization and procedures of regulatory agencies. (Smith 1997a, b, and c provide good discussions of these issues.)

Statutory Framework

A basic legal framework must be specified to govern the operation of each sector subject to regulation. Typically, a separate statute is used to create an organic legal framework for each sector. This framework defines the goals for the sector, the basic preferred structure of the sector (defining regulated monopoly and competitive segments), a restructuring and privatization program for the incumbent suppliers, regulatory goals and responsibilities, the structure of the regulatory institutions that will govern the sector, methods for selecting regulators, the ways in which the regulatory agency will be funded, and procedures that the regulators must use when making regulatory decisions.

Structure of Regulatory Institutions

A number of decisions must be made about the structure of the regulatory framework that will govern each sector. Should the regulatory agency govern a single industry or multiple industries? Should a single regulator head the regulatory agency or should multiple commissioners serve? How should regulatory appointments be made? How much discretion and authority to acquire information from regulated

firms should the regulator have? What kind of staff resources should be allocated to the agency? What kind of administrative procedures should be used? What opportunities for judicial review should be created?

SINGLE- AND MULTI-INDUSTRY REGULATORS. There is an ongoing debate on whether it is better to have single-industry regulatory agencies (as in England and Wales) or multi-industry regulatory agencies (as with most state public utility commissions in the United States). The primary argument in favor of the single-industry approach is that it ensures deep technical and economic expertise about the attributes of the industry within each regulatory agency's jurisdiction, leading to better regulatory decisions. But there are several arguments against the single-industry approach. It is inefficient because it duplicates staff with common skills that are needed in the regulation of two or more industries. It limits opportunities for learning based on experience with new regulatory mechanisms—such as price caps—or competition policies that may be applied first in one industry and then later in other industries under the agency's jurisdiction. And it increases the potential for capture of the regulatory agency by a single interest group—in particular, the firms that it regulates.

In my view there is enough overlap in regulatory issues, necessary staff skills, and opportunities to learn from experience to make it desirable for a single agency to regulate electricity, natural gas pipelines and distribution networks, common carrier oil pipelines, and perhaps water distribution. Telecommunications regulation involves similar issues, but also often includes allocating spectrum bandwidth and supervising obligations placed on broadcasters. The sector is also experiencing rapid technological change. Thus enough is different about telecommunications to justify a separate regulatory agency. Similarly, any residual regulation of transportation is probably best assigned to a separate agency.

SINGLE REGULATOR AND MULTICOMMISSIONER STRUCTURE AND APPOINTMENTS. Single regulator systems grant more power to a specific individual than do multicommissioner systems, which require that a majority of commissioners agree on major policy decisions. This difference is somewhat diminished when the chair of a multicommissioner system is given control over the agenda and staff resources. There is necessarily a tradeoff between the ability of a single regulator to make quick decisions and the potential that a single regulator will abuse his or her authority. It may be difficult to attract new investment when a regulator has substantial discretion, because investors recognize that policies can change dramatically with a change in the regulator.

The regulator's ability to act decisively and to abuse authority is necessarily influenced by the policy discretion given by the statute, by the procedural requirements under which the regulator operates, and by the opportunities that affected parties have for judicial review. The more specific the statutory mandate is, the more desirable a single regulator is likely to be. At the same time, the more power the regulator has, the more important are transparent administrative procedures and opportunities for judicial review. While regulators cannot (and probably should not)

be completely insulated from political pressure, agencies should be reasonably independent of changing political winds. Appointing regulators for fixed terms that are not coterminous with changes in government can contribute to the independence of the agency.

REGULATORY DISCRETION. Statutes can convey responsibilities to regulatory agencies in two ways. The first is to give the agency substantial discretion by defining general policy objectives and leaving it to the agency to develop mechanisms to achieve those objectives. The agency's discretion is limited by procedural requirements, legislative oversight, and opportunities for judicial review. The second approach is to give the regulatory agency substantially less discretion by defining more precisely what it is supposed to do and how it is supposed to do it. Here too discretion may be further limited by procedural requirements, legislative oversight, and opportunities for judicial review.

The second approach usually makes more sense in developing countries that are privatizing their utilities. For example, the statute may define a detailed regulatory contract that specifies prices and price adjustment provisions, service obligations, and other factors that are to prevail for several years, after which the agency is given more discretion to develop and apply new regulatory mechanisms. This approach establishes the credibility of the reform program and so helps to attract capital to these sectors. This approach is also compatible with a single regulator model combined with transparent opportunities for judicial review of the agency's implementation of the statute. This strategy is not costless, however. Rigid contractual terms reduce the agency's ability to make socially desirable changes in response to changing supply and demand conditions, better information, and more experience. As experience and credibility are gained and administrative and judicial review procedures are developed, the agency can be given more discretion to define regulatory mechanisms.

ADMINISTRATIVE PROCEDURES AND JUDICIAL REVIEW. Administrative procedures and opportunities for judicial review should provide affected parties with opportunities to express their views on proposed agency actions and have those views and associated facts taken into account in agency decisions. Excessive due process obligations, however, create significant opportunities for administrative delay, frustrating regulatory action of any kind. In developing countries transparent administrative procedures and opportunities for judicial review are likely to be especially important to infrastructure investors, to competitors who rely on access to network infrastructure to compete, and to consumers who want to be sure that the objectives of the statute are being realized and regulatory authority is not being abused. Basic due process obligations include requirements that agencies publicly announce the initiation of a process to make a policy decision, opportunities for interested parties to comment on major policy issues, rules restricting *ex parte* nonpublic contacts between regulators and interest groups affected by their decisions, and requirements that regulators explain and justify policy decisions based on rational evaluation of the factual record and relevant statutory obligations.

AGENCY STAFF AND INFORMATION COLLECTION AUTHORITY. A regulatory agency requires qualified staff to fulfill its obligations. Similarly, a regulatory agency cannot do an effective job if it does not have the authority to require the firms under its jurisdiction to submit cost, price, and other information necessary for assessing their behavior and performance. Moreover, an agency must have the authority to define accounting systems for regulated firms and to audit them. Information requirements for firms in competitive segments may be less stringent, and procedures can be adopted to ensure that sensitive business information is held in confidence.

Fostering adequate staff resources and skills, developing uniform accounting systems, and collecting and evaluating cost, price, and performance data can take several years. The likely availability (or unavailability) of staff skills, analytical capabilities, auditing authority, databases, and experience should be taken into account when designing regulatory mechanisms. Relatively simple regulatory rules and procedures may initially be more effective than more complicated mechanisms that are theoretically superior. Opportunities to evaluate, adapt, and enhance regulatory procedures as experience is gained can be built into the program so that investors can anticipate reasonable changes over time.

Design of Regulatory Mechanisms

Regulators of legal monopolies typically are given multiple explicit or implicit goals, which they must achieve subject to a number of constraints. These goals include:

- Making sure that the monopoly charges consumers reasonable prices for services (*the rent extraction goal*).⁵
- Inducing the monopoly to provide services efficiently (*the supply-side efficiency goal*).
- Using the level and structure of prices to induce consumers to make efficient use of the services offered by the monopoly (*the demand-side efficiency goal*).
- Providing adequate incentives to attract additional capital to the sector and to maintain the existing capital stock. Regulatory procedures should allow the monopoly to anticipate revenues that will at least cover the costs of supplying services, including a return on investment at least equal to the firm's cost of capital (*the capital attraction or firm viability goal*).
- Achieving income redistribution goals through the level and structure of prices for regulated services rather than through general tax and expenditure policies (*the income redistribution goal*).

The regulator's task would be fairly straightforward if complete exogenous information were available on the regulated firm's present and future production and cost opportunities, on the demand structures for all types of consumers and how they will evolve, and on the variables needed to implement income distribution goals, and if the regulator were insulated from interest group politics and trusted to pursue its statutory obligations fairly and efficiently. In that case the regulator could mechanically calculate the (second-best) optimal price levels and tariff structure for the regulated firm at every point in time and adjust them as

demand, cost, and income distribution conditions changed (given the balanced budget constraint). Elaborate administrative procedures would not be required because the regulator would be assumed to efficiently pursue well-defined public interest goals.

But this ideal situation never exists in reality, no matter how useful may be the theoretical benchmarks derived from models that assume that it does. Moreover, if one were to believe that these models were reasonable characterizations of reality, there would be no need to separate the regulator from the regulated firm, because the regulator would have all the information needed to efficiently produce and price the output of the regulated firm.

The fundamental problem that regulatory agencies must confront in designing regulatory mechanisms is that the regulated firm possesses better information about its production cost opportunities, the costs of its services, the operating characteristics of its network, the effort it expends to keep costs low, the attributes of its customers' demand patterns, the quality of its services, the costs of improving that quality, and the responsiveness of its customers to various tariff structures. In other words, there is an asymmetry of information between the regulatory agency and the regulated firm. This information asymmetry can be reduced, but not eliminated, by requiring the regulated firm to report to the regulator on its costs, prices, demand patterns, and technical operating characteristics and by providing for associated auditing authority. In addition, the regulator's capability to obtain good information about cost, demand, income distribution, and other relevant variables may improve as it gains experience and its auditing and analytical abilities improve.

Still, the information asymmetry can never be eliminated. Accordingly, the regulatory game is one in which the regulated firm will always know more about its economic environment than the regulator and will try to extract some rent from consumers as a result of its information advantage. In addition, both the regulated firm and the regulatory agency operate in a world where future cost opportunities, demand characteristics, and relevant income distribution variables are uncertain and where credible and complete contingent claims contracts are not viable. Finally, interest groups are unlikely to trust the regulatory agency to pursue reasonable public interest goals without regulatory procedures that are subject to public scrutiny and judicial review. The design of regulatory mechanisms and regulatory procedures should take all these considerations into account.

Two types of regulatory mechanisms provide useful benchmarks for thinking about the development of practical regulatory mechanisms, given a country's regulatory and legal capabilities and the initial conditions of the infrastructure sector. These are fixed price mechanisms and cost-plus mechanisms.

Fixed Price Mechanisms

Under a fixed price regulatory mechanism the regulator defines *ex ante* a set of prices (or a weighted average of prices for different services) that the regulated firm will be allowed to charge consumers for the services it provides. The regulatory

mechanism may also specify automatic price adjustment formulas tied to exogenous variables such as general inflation or productivity growth. The critical feature is that the prices are not tied directly to the regulated firm's realized costs or profits.

Fixed price mechanisms include pricing mechanisms that define specific prices for specific services going into the future, mechanisms that specify initial prices for each service and a formula for future adjustments, and price cap mechanisms that set initial prices for services and allow both adjustments in the average price level and a rebalancing of the structure of prices over time. A price cap formula is specified as a weighted average price for all regulated services supplied by the firm, and provisions are made for adjusting the weighted average price over time to reflect input price inflation and productivity growth norms. In all fixed price mechanisms the decoupling of prices from the costs and profits realized by the firm provides high-powered incentives for the firm to supply output efficiently. Depending on how the mechanism is designed, it may also give the regulated firm incentives to use its information about consumer demands to implement an (second-best) efficient price structure.

A number of potential problems must be confronted in designing a fixed price mechanism, however. The first is that without good information about the firm's cost opportunities and demand attributes, the regulator will be uncertain about the proper level at which to set initial prices and how these prices should be adjusted as economic conditions change over time. Prices that are set too high will shift income from consumers to the regulated firm and may conflict with the rent extraction goal. Prices that are set too low will not provide adequate incentives for new investment and maintenance of equipment, conflicting with the supply-side efficiency and firm viability goals. And without good information about marginal costs and demand patterns, it will be difficult to specify an efficient price structure (prices for different services or types of customers, nonlinear tariff structures, and so on) or set weights on a global price cap that would provide incentives for the regulated firm to establish the (second-best) optimal tariff structure.

Fixed price contracts can also create perverse incentives for the quality of service provided by a regulated monopoly. These mechanisms can create powerful incentives to keep costs low to maximize profits and inadequate incentives to invest in service quality. Accordingly, fixed price mechanisms should reflect the value of service quality either by tying prices directly to service quality or by specifying minimum quality standards and associated penalties if they are not achieved. Experiences in Peru and in England and Wales illustrate the benefits and drawbacks of fixed price mechanisms.

REGULATION OF TELECOMMUNICATIONS IN PERU. Telecommunications regulation in Peru provides an interesting example of both a sector reform model and the use of a fixed price regulatory mechanism applied to an incumbent privatized with a temporary monopoly concession (see OSIPTEL 1995, 1997; Telefónica del Perú 1996, 1997). In 1970 the Peruvian government nationalized the private firms that made up the telephone sector and created two state-owned companies to run the telephone system. Over the next 20 years, but especially as a consequence of the hyperinflation of the

late 1980s, the quality of telephone service deteriorated and the telephone system came close to collapsing. Both state-owned companies—Compañía Peruana de Teléfonos and Peru Entel—were essentially bankrupt. Price increases had lagged seriously behind cost increases, and prices were inadequate to cover operating costs or to support new investment. As in most other countries, prices for local service were kept very low while prices for intercity and international services were kept high to help cover the costs of the local network. There was low telephone penetration (by far the lowest in Latin America), huge excess demand for telephone service (with queues for legal connections as long as six years, along with the inevitable corruption associated with jumping the queue), poor service quality, and limited investment in new facilities.

In 1993 the government formed Telefónica del Perú (TDP) out of the existing state-owned enterprises and auctioned it off to the private sector using a fixed price regulatory contract. TDP was acquired by Spain's national telephone company. The basic reform model under which TDP was privatized envisioned giving it a monopoly over local and intercity services (except cellular) for a fixed period (roughly five years). At the end of the five years competitors would be allowed to enter the local, intercity, and international markets and compete with TDP. At that time a new regulatory mechanism would be applied to any services provided by TDP that required continued regulation. An independent regulatory agency was created to enforce TDP's obligations under the regulatory contract, develop policies to promote competition at the end of the exclusivity period, and regulate TDP after the initial five-year contract ran its course. The terms and conditions of the initial regulatory contract are enforceable under commercial law, giving the regulator little discretion during the exclusivity period.

There were three rationales for privatizing TDP as a monopoly with a fixed exclusivity period rather than immediately opening up intercity and long-distance markets to competition. First, attracting new investment to quickly improve telecommunications infrastructure was the top priority. Second, Peru needed to establish a reputation for credible regulatory rules to attract investment to the sector. Finally, introducing competition before significant tariff rebalancing would have led to cream-skimming and uneconomic bypass, undermining needed investment in basic infrastructure. Similarly, the decision to specify a regulatory contract enforceable under commercial law was motivated by the desire to make price commitments credible.

The regulatory contract that applies to TDP during the exclusivity period has two important features. First, it specifies maximum real prices for various service components for each year that the contract is in effect. The prices change over time so as to affect a tariff rebalancing, increasing local service prices and reducing intercity and international service prices. The idea was to bring intercity and international prices gradually in line with what they would become in a competitive market. Second, the contract specifies the minimum number of new lines that TDP must install each year and various minimum service quality improvement obligations, including obligations to install additional public telephones (important in remote rural areas). TDP files compliance reports with the regulator monthly and annually to demonstrate that it is fulfilling its contractual obligations.

This program appears to have been successful in a number of dimensions. TDP has generally exceeded all of its main investment and service improvement goals. Indeed, the installation of new lines and public telephones proceeded more quickly than required by the contract. Between 1993 and mid-1997 TDP increased the number of telephone lines in Peru by about 250 percent and the number of public phones by about 400 percent. In less than five years the telephone network went from being less than 40 percent digital to more than 90 percent digital. The average waiting time for a new line fell from 70 months in 1993 to 5 months in 1996. Over the same period TDP cut its work force in half. Finally, significant tariff rebalancing will have been completed by the end of 1998, setting the stage for competition to begin in intercity and international services.

Still, several concerns have been raised about this reform program and the associated regulatory contract. Some observers argue that the contract prices were too high and that TDP has been able to earn excess profits during the contract period. The perceived high prices have produced a vocal consumer backlash. Concerns also have been expressed that the reform program does not include low-use tariffs or direct subsidies to deal with the needs of the poor. In addition, there appears to be widespread concern that the five-year exclusivity period has given TDP an entrenched position that will make life difficult for future competitors. The entrenchment concerns are both economic and political.

These problems may be the inevitable outcome of the need to apply regulatory mechanisms that foster a reputation for making credible commitments to investors—especially given Peru's limited regulatory traditions and expertise—and the resulting reliance on simple rather than complex mechanisms in order to balance investment, rent extraction, and supply-side efficiency goals. Such tradeoffs are a fact of life in the development of practical and effective, but necessarily imperfect, regulatory mechanisms.

PRICE CAPS FOR ELECTRICITY DISTRIBUTION COMPANIES IN ENGLAND AND WALES. Much has been written about the experience with privatization, restructuring, and competition in the electricity sector in England and Wales (Newbery and Pollit 1997; Wolfram 1997; Armstrong, Cowan, and Vickers 1994). Here I focus on the regulatory mechanism used to regulate the monopoly distribution business.

The law restructuring the electricity sector in England and Wales created a competitive generation segment and separate regulated monopoly distribution (wires) and transmission (including network operations) segments. Large retail customers were initially given the right to purchase unbundled wires services from the distribution and transmission companies, giving these consumers direct access to the generation market to arrange for their service needs directly in that market or through "supply" intermediaries. The right to buy generation service directly in the wholesale market has gradually been extended to smaller consumers.

The restructuring program provided that the electricity distribution business would be regulated using a price cap mechanism. Basically, the initial distribution tariffs could be adjusted annually to reflect inflation and a prespecified productivity

factor (which in most cases initially assumed declining productivity because of the significant investments required to upgrade distribution systems). Distributors also were subject to a variety of service quality standards. Financial penalties were assessed if the standards were not met. The costs of generation services acquired by a distribution company and made available to retail customers who were not yet able to access the competitive generation market directly, or chose not to, were effectively passed through pound for pound in final prices. The parameters of the price cap mechanism and associated quality standards were subject to review and adjustment by the regulator every few years.

Thus the mechanism anticipated continued regulation of the distribution business using a “fixed price” price cap mechanism but built in a “ratchet” process that provided for regulatory review every few years. The review includes an assessment of the profitability (rate of return) of the distribution business, which in turn requires an operating and capital cost accounting system that can be used to make meaningful assessments of economic profits. In effect the regulatory mechanism is a price cap mechanism with a cost-of-service ratchet to the price level every few years.

The price cap mechanism has provided good incentives for reducing costs and maintaining service quality. Labor productivity has improved in the distribution companies and service quality deterioration has not been a significant problem. The main problem has been that distributors have achieved much larger productivity improvements than were anticipated by the productivity factors in the initial price caps, allowing them to earn high profits.⁶ High profits have led to public criticism of the system and enormous pressures on the regulator to reset prices and tighten productivity targets, which have been done. In short, it was difficult to find the right productivity adjustment factor *ex ante*, and this has conflicted with rent extraction goals. But the provisions built into the regulatory design for periodically resetting the price level based on actual performance and the productivity factor going forward made it possible to adjust price levels and their future trajectory as more information and experience were obtained by the regulator. Of course, the downward ratcheting of prices *ex post* dulls the regulated firm’s incentives to minimize supply costs, and there are clear tradeoffs among rent extraction, supply-side efficiency, and (if the productivity standard had been too tough rather than too soft) firm viability goals. Nevertheless, experience with price cap mechanisms around the world suggests that to be credible they should include a prespecified review and adjustment protocol from the outset, because excessive profits or non-compensatory prices will almost certainly create political pressure to reopen the mechanism down the road.

Both of these examples demonstrate the benefits and costs of fixed price regulatory mechanisms. It is important to recognize, however, that the balance of costs and benefits of alternative mechanisms is likely to differ significantly from country to country based on the initial economic conditions of the sector, the legal and political environment, and the information available to and technical capabilities of the regulator. For example, the balance of benefits and costs of privatizing an incumbent telephone company with a five-year monopoly, as in Peru, may be very different in

a country with a well-developed telephone network where performance is relatively good, there is no urgent need to attract large amounts of capital to the sector, and there are experienced regulators with good information and a reputation for protecting private property rights associated with investments in infrastructure. Moreover, Peru's fixed price contract might not have been feasible if macroeconomic reforms had not sufficiently stabilized the economy to make it feasible to define prices in terms of real values of the local currency. The absence of such macroeconomic stability creates enormous financial risks for investors—and makes it difficult to attract investments in long-lived immobile capital.

Cost-Plus Mechanisms

At the other end of the spectrum of regulatory mechanisms lies cost-plus regulation. Under cost-plus regulation the regulated firm effectively submits a bill for its operating expenses and capital costs, including an after-tax return on its investment that equals or exceeds its cost of capital (the "plus"). The regulator then passes these costs through in the prices charged to consumers.

There are a number of potential problems with pure cost-plus regulation when regulators have imperfect information about a firm's cost opportunities. At the extreme, unless the regulator has established a good operating and capital cost accounting system, can accurately measure the regulated firm's cost of capital, and can audit the firm's actual costs, the firm is in a position to misrepresent its costs, effectively passing itself off as the highest-cost firm that the regulator would find credible.⁷ As a result the regulator sets prices too high, which conflicts with the rent extraction goal. Even if the regulator has implemented good cost accounting, measurement, and auditing procedures, in the absence of good information about the firm's cost opportunities, managerial effort, and associated costs, pure cost-plus regulation dulls incentives for the regulated firm to minimize costs, causing the audited costs that the regulator observes and passes through to consumers to be too high. This outcome conflicts with both the rent extraction and supply-side efficiency goals (Schmalensee 1989).

The potential benefit of a credible cost-plus regulatory system is that it is likely to be very effective in attracting capital investment to the sector because investors are assured that they will recover their operating and investment costs. The capital mobilization (or firm viability) goal is easily satisfied, and this can be attractive to investors. Pure cost-plus systems shift a variety of firm-specific and market risks from the firm to consumers and reduce the regulated firm's financial risk and cost of capital. Infrastructure sectors in the United States and other industrial countries evolved under some variant of cost-plus regulation and, whatever the faults of this system, inadequate investment and shortages of capacity to meet demand are rarely among them. Accordingly, some form of cost-plus regulation is potentially attractive in a country where quickly attracting a large amount of capital investment is considered important, given the current state of the sector and the legal and regulatory environment and associated reputation for protecting private property rights and enforcing

contracts that the country has at the start of the reform program. Cost-plus regulation with accounting and auditing standards can also provide useful information to the regulator for designing tariff structures, although additional information about consumer demand is still necessary to design (second-best) optimal tariffs.

The history of cost-based regulation in the United States is replete with economic and regulatory debates over accounting and auditing procedures, capital valuation and depreciation rules, and methods for calculating firms' cost of capital. There is no reason for a developing country to replicate all this work (or to repeat mistakes), and there is much to learn from Canada and the United States, where cost-based regulatory principles historically have been applied in the most transparent way. Moreover, the typical theoretical characterization of a pure cost-plus regulatory system has little relationship to the way cost-based regulation works in practice.

The United States has the most experience with a form of cost-plus regulation called "rate of return" regulation, which has been applied to legal monopolies in the electricity, telephone, natural gas pipeline and distribution, and water distribution sectors for 75 years. In the United States prices for regulated services are not continuously tied to accounting costs. Rather, prices are set in public hearings based on an evaluation of capital and operating costs that conform to a specific accounting system and have been audited along with an estimate of the firm's financial cost of capital. Once these base prices have been set they are not adjusted (or are only partially adjusted) automatically for changes in costs over time. Prices are fixed until they are reviewed by the regulator again (Joskow 1974; Joskow and Schmalensee 1986). The period between regulatory reviews—known as regulatory lag—often lasts several years.

Regulatory lag effectively turns cost-plus regulation into a fixed price regulation system with a cost-based ratchet adjustment every few years. The two types of regulation also have similar incentive properties. This type of cost-plus regulation worked best in the United States when the lags were relatively long because nominal unit costs were falling due to productivity growth and low input price inflation, and formal regulatory reviews were few (Joskow 1974). Consumers left some rents on the table for the regulated firms but the system provided reasonably good efficiency incentives and high rates of productivity growth. In addition, not only are costs subject to elaborate accounting and auditing requirements in the United States, but regulators do not have to accept the costs that are presented to them if they determine that they are unreasonable or imprudent. Regulators can make these judgments through management audits or benchmarking of the firm's costs against comparable firms (or statistically comparable firms), effectively applying yardstick competition.

Managing a reasonably good cost-based regulatory system places an enormous information burden on the regulator, however. A good accounting and auditing system needs to be put in place. Methodologies for measuring the firm's cost of capital and evaluating its behavior through management audits and benchmarking studies need to be implemented. These tasks are especially challenging when technology is changing rapidly and good comparables for benchmarking are unavailable. At the very least, these are capabilities that a new regulatory agency is unlikely to be able to acquire instantly, and it may take years to develop the necessary capabilities

and experience (*The Economist*, 3 January 1998, 59–60). These considerations suggest that newly created regulatory agencies are unlikely to have the capability to implement (roughly) cost-plus systems effectively.

Hybrid Systems

Neither a pure fixed price system (including a price cap without a ratchet) nor a pure cost-plus system is likely to be the most desirable regulatory mechanism for an infrastructure sector expected to continue as a regulated monopoly for the foreseeable future (Schmalensee 1989).⁸ Moreover, the regulator's information and abilities are likely to improve over time if it is given adequate resources and independence. This evolution is likely to make more sophisticated mechanisms a viable option as the regulatory framework matures and experience with the industry increases. A variety of hybrid forms take some account of the firm's realized costs but also embody decentralized incentives that better align the objectives of the regulator and the regulated firm. The choice among them depends on the initial economic problems faced by the infrastructure sector, the nature and credibility of legal and regulatory institutions, the state of cost accounting and auditing capabilities, the information regulators have about cost opportunities and demand, and the human resources available to perform regulatory tasks.

I have already discussed price cap mechanisms that have a built-in mechanism to reset the cap and the adjustment factors that will be used going forward. The regulatory contract used for the telephone company in Peru has some of these features but also embodies a specific tariff rebalancing process to prepare for competition in intercity and international services. The regulatory mechanism that will replace the contract at the end of its term is not clearly specified, however, and the terms and conditions of follow-on regulatory arrangements will likely lead to significant disputes between TDP and its regulator.

Mechanisms that are more closely tied to audited costs but that build in incentives to reduce costs are also available. These include a variety of profit-sharing mechanisms that ensure that the benefits and burdens of cost savings are shared by the regulated firm and consumers (Joskow and Schmalensee 1986; Laffont and Tirole 1993). But these mechanisms are viable only if satisfactory cost accounting, reporting, and auditing capabilities exist. They too eventually require a ratchet to evaluate their performance and to reset the base price.

The choice of regulatory mechanism also depends on a country's macroeconomic stability. Cost-based systems are likely to work poorly in countries where rapid inflation is a possibility and interest rates and exchange rates fluctuate widely and are subject to government controls. This is because capital cost accounting systems and associated methods for calculating economic depreciation, costs of capital, and allowed rates of return do not work well in unstable macroeconomic environments. Regulatory lag can become a big obstacle to attracting investment in those environments as well. Price cap mechanisms may more easily incorporate adjustment factors to take these variations in economic conditions into account *ex ante*.

As noted, many developing countries face serious challenges in improving their infrastructure sectors. Well-established cost accounting and auditing protocols are lacking. Firms will likely resist detailed audits. Capital accounting and cost of capital measurement systems remain undeveloped. Tariffs need to be rebalanced in anticipation of competition. And macroeconomies are unstable. Given these shortcomings, a hybrid of a fixed price mechanism and a price cap mechanism—including quality of service and performance norms and prespecified provisions for evaluation and adjustment once every few years—will often provide the most effective regulatory approach.

Prices initially need to be set high enough so that the capital attraction (or firm viability) goal is likely to be achieved. It should be expected that consumers will leave some rent on the table during the initial years of the program. The loss of these rents should be more than compensated for by investment and supply-side efficiency incentives, which in the long run can be reflected in lower prices and higher service quality. For this approach to work well, however, there must be adequate legal protections through general contract law or specific regulatory law to ensure that the ratchet will not be used to hold up the firm. In addition, the regulator must use the fixed price period to develop the cost accounting, auditing, cost of capital measurement, benchmarking, and related information about costs and demand needed to do a good job evaluating the fixed price mechanism at its prespecified review.

Network Access and Competition

One of the most difficult, contentious, and important tasks for regulators is defining the terms and conditions under which suppliers of competitive services (intercity telephone, electricity generation, natural gas production) will have access to the regulated bottleneck monopoly network facilities that they need to provide competitive services to consumers. Without good access rules, efficient competition will be difficult to realize. The issues are complicated further and become even more contentious when the owner of the regulated monopoly network is also a supplier of services in the competitive segment—that is, when the owner both supplies network services to and competes with rivals in these segments. An obvious way to deal with these complications is to bar the network owner and operator from participating as a supplier in the competitive segments.⁹ This kind of restriction on the lines of business in which the network operator may participate raises difficult economic, technical, and political issues that vary by sector and by country, as I have briefly discussed. In what follows I simply assume that the bottleneck network operator also has an unregulated affiliate that competes in the competitive segments.

Most theoretical discussions of network access focus on the prices charged for access, as I do here. But while access pricing is important, network access issues encompass a much broader set of terms and conditions than pricing alone. These include cost separations between regulated and unregulated businesses to guard against cross-subsidization, the speed and quality of physical connections of com-

petitors to the network, billing arrangements, the provision of information about network characteristics and how they are likely to change over time, information transfers about customer characteristics by the vertically integrated network operator, and customer referrals by the network operator. Thus the full range of the terms and conditions of access and the possibility that the network operator will behave in ways that inhibit comparable access to the bottleneck facilities by its competitors are relevant and important policy considerations.

Following Laffont and Tirole (1996), I use a simple example to frame the access pricing problem and possible solutions to it. Consider a local telephone network N that is a regulated monopoly. The costs of the network that are recognized by the regulator as being recoverable from network customers consist of a fixed cost K_N (the joint cost of supplying local and long-distance service) and marginal costs c_N per unit of throughput on the network. The number of local calls is given by q_N . The network operator is also a competitor in the long-distance market, can supply long-distance service at a marginal cost of c_1 per unit of throughput, and supplies total throughput of long-distance traffic of q_{C1} . One or more unaffiliated competitive long-distance companies need to interconnect with the network to reach consumers. Assume that these competitors are symmetric and have constant marginal costs of c_2 per unit of throughput of long-distance service and total throughput over the network of q_{C2} . The costs for the local network (C_N), the network operator's long-distance affiliate (C_{C1}), and the competitive long-distance carriers (C_{C2}) are as follows:

$$\begin{aligned} C_N &= K_N + c_N (q_N + q_{C1} + q_{C2}) \\ C_{C1} &= c_1 q_{C1} \\ C_{C2} &= c_2 q_{C2}. \end{aligned}$$

The total number of calls is given by

$$Q = (q_N + q_{C1} + q_{C2}).$$

Prior to the introduction of competition the regulator sets prices that produce revenues equal to the incumbent's total costs (budget balance). Assume that prices for local and long-distance service include fully distributed fixed costs allocated in proportion to output (note that an infinite number of other joint cost allocation schemes satisfy the budget balance constraint as well). That is, the regulated price for local service (p_N) has been set at a level that covers a fraction of the fixed (or joint) costs of the local network (where q_{C2} is 0 before competition is introduced) plus the marginal cost of network service:

$$p_N = \frac{K_N}{Q} + c_N.$$

The regulated price for long-distance service (p_1) has been set at a level that covers a fraction of the fixed costs of the local network (where q_{C2} is 0 before long-

distance competition is permitted) plus the marginal cost of long-distance service plus the marginal cost of using the local network:

$$p_1 = \frac{K_N}{Q} + c_N + c_1.$$

Now assume that the regulator is contemplating allowing competitive (facility-based) long-distance carriers to enter the market. What access charge should these competitors pay to use the local network? One approach would be to require competitors to pay the marginal cost of using the local network (c_N) per unit of throughput. Assuming away any market power for the moment, the price that competitive long-distance carriers would charge in this case is given by

$$p_2 = c_N + c_2.$$

Setting the access price equal to the marginal cost of access creates problems, however. First, for every unit of long-distance service that competitors capture from the incumbent, the incumbent loses K_N/Q of revenue, which it historically has relied on to cover the fixed costs of the network. Where does the incumbent get the revenue to cover this "access deficit?" Second, the incumbent is implicitly paying a much higher access charge than its competitors in the long-distance market. The implicit access charge being paid by the incumbent in connection with supplying long-distance service is

$$a = \frac{K_N}{Q} + c_N = p_1 - c_1.$$

The incumbent will not be able to compete with equally efficient rivals unless it lowers its price, incurring a further loss of revenues (access deficit) needed to cover the fixed costs of the local network. Moreover, if it does not lower its prices it will encourage entry of inefficient rivals.

How should the regulator deal with this problem? One approach would be to apply a parity principle and require the competitive entrants to pay the same access charge as the incumbent implicitly pays. In the example above the access charge (a) is given by an allocation of the fixed costs of the network that is proportional to the competitors' throughput on the network (K_N/Q per unit of throughput) plus the marginal cost of using the network. With homogeneous products, symmetric firms, and no market power, the price charged by competitive firms subject to this access charge would be given by¹⁰

$$p_2 = a + c_2.$$

$$p_2 = \frac{K_N}{Q} + c_N + c_2.$$

In general, the access charge for implementing the parity principle would be given by $p_1 - c_1$. This more general formulation is referred to as the efficient component pricing rule (ECPR) and is attributed to Baumol and Willig (in Baumol and Sidak 1994). It can be extended directly to the case where the incumbent network operator supplies two or more competitive products with different prices and marginal costs. Variants on the ECPR have been adopted in a number of jurisdictions, and I believe it is a good practical benchmark approach for establishing access prices.

There are a number of theoretical and practical problems with both the parity principle and ECPR, which I discuss briefly below. First, however, it should be recognized that even in the simple world where the competitive segment involves homogeneous products, symmetric competitors, and no market power, the access price could still lead to costly inefficient behavior by competitors. If the access price (for example, $K_N/Q + c_N$) is higher than the cost of bypassing the local network, competitive suppliers may spend money to inefficiently bypass the local network. For example, a competitive supplier might build a separate access line so that a large user of the network can connect directly to the long-distance supplier, bypassing the local network and the associated access charges.

The problem in this case is that the allocation of the fixed costs of the network to competitive services is too high to be sustainable in the face of competition without creating inefficiencies. There are two solutions to this problem. One is to require competitive suppliers to contribute to the fixed costs of the local network—that is, to pay an access charge regardless of whether they use the local network or not. In effect, all long-distance calls terminating in a particular geographic area pay the access charge to the network operator or pay a tax (transferred to the network operator) equal to the contribution to the fixed cost of the network that the competitive supplier would make if it used the network.

An alternative solution is to first rebalance tariffs so that a smaller proportion of the costs of the local network are recovered from competitive services. This would mean increasing p_N and reducing p_1 (and p_2). So, for example, the regulator could shift responsibility for recovering the fixed costs of the network entirely to local calls, increasing p_N (or more likely the fixed monthly charge for a local line) so that local calls produced enough revenues to cover all of the fixed costs of the network as well as the marginal cost of local calling:

$$p_N = \frac{K_N}{q_N} + c_N.$$

The access charge for competitive suppliers would then simply be set equal to the marginal cost of using the local network for all competitive suppliers (c_N). The result is that local service pays the full “tax” necessary to cover the fixed costs of the local network, while long-distance service pays none of it. But there is no reason to believe that shifting the entire tax burden to local calling (or to a fixed charge for local access) is optimal unless the demand for local calling or local access is perfectly inelastic and income distribution effects are ignored. This is not likely

to be the case in many developing countries, where incomes and telephone penetration are low.

The possibility of tariff rebalancing and questions raised by the extreme case of rebalancing just examined call for a deeper exploration of the rationale for and efficiency of the regulated prices p_N and p_1 , as well as the associated allocation of the fixed costs of the network. The ECPR and related rules are essentially top-down rules that start with the prevailing regulated prices, implicitly assume that they are optimal, and derive access prices that implement the parity principle. An alternative approach is to calculate the optimal final goods prices (or optimal implicit allocation of joint costs) and then to specify access prices that will lead competition (including imperfect competition) to produce these optimal final goods prices. This type of analysis is presented very elegantly by Laffont and Tirole (1996).

As noted, a number of practical and theoretical objections have also been made to applying the ECPR. The first is that it requires the regulator to accurately measure the marginal (or avoided) cost that the incumbent incurs in supplying the competitive service. The incumbent in turn has the incentive to understate those costs to increase the access prices that its competitors must pay. Here the information disadvantage for developing country regulators may be especially severe. A related objection is that the incumbent can indirectly manipulate the access charge by requesting that the regulator adjust regulated prices (both levels and structure) for local and long-distance services. This makes the access charge a moving target for the incumbent, the outcome of a regulatory process that it may have difficulty monitoring and influencing.

More generally, competitors are concerned that the parity principle will be violated to their disadvantage as a consequence of the incumbent's ability to use the regulatory process to manipulate regulated final prices (and associated allocations of joint costs), which in turn determine the access price as a residual. These concerns can be addressed by building up the access charge (from the bottom up rather than from the top down) using clearly specified cost allocation rules (which could change in a predetermined way over time), unbundling the access fee from the incumbent's local and long-distance prices, and requiring that both the incumbent (if it chooses to compete in the competitive segments) and its competitors pay the same access price. That is, a cost-based access price is clearly defined that is consistent with the ECPR (as long as the components of the tariff add up to the bundled prices), and all competitors must pay the same price. The cost allocations can reflect those embodied in existing prices or in new prices that rebalance tariffs to reflect demand considerations (for example, Ramsey-Boiteux pricing) and income distribution goals. However, the cost allocations need to be visible and verifiable and to support unbundling of the incumbent's tariffs to include an access charge that is (roughly) a fixed point.

Laffont and Tirole (1996) argue that the ECPR may be inconsistent with optimal access pricing if competitive segments exhibit market power (so the access price gets marked up by competitors in those segments); competitors are supplying different products—reflecting brand loyalty, name recognition, or switching costs—and, as a result, have some market power; or the network operator has lower costs in the competitive segment than its rivals have. These considerations may imply access prices

that are higher or lower than the ECPR and implicitly different access charges for the incumbent and its competitors. The rationales developed by Laffont and Tirole (see also Economides and White 1995) for departing from the ECPR and the parity principle are clear and theoretically correct. As a practical matter, however, I believe that it would be a mistake to encourage regulators to set different access charges for different competitors as a way of dealing with market imperfections. Abandoning the parity principle or creating other rationalizations for manipulating access prices that do not follow some simple comparability rule invites rent seeking and regulatory mischief. Furthermore, new regulatory agencies are unlikely to have the information or expertise to effectively implement more complex access pricing rules. Finally, I do not believe that a network operator's competitors will ever find the global price cap proposed by Laffont and Tirole (1996) to be an acceptable approach, because they will be concerned that effectively allowing the incumbent network operator to set its own access fees will encourage predatory behavior. For similar reasons, market power problems should be directly addressed with structural and behavioral mitigation measures rather than through manipulation of access prices.

Conclusion

Supporting regulatory institutions should be created as an integral component of infrastructure sector reform, not as an afterthought. These institutions must be part of an economic environment for each infrastructure sector that protects consumers from abuses of market power. They must promote efficient supply behavior by firms providing residual monopoly services subject to public regulation. They must facilitate competition by implementing appropriate terms and conditions for access to bottleneck network facilities. And they must guard against cross-subsidization and unreasonable discrimination among competitors using these facilities. Investors should perceive regulatory rules and procedures as credible commitments on which they can rely when making investment decisions. Principles of independence, transparency, accountability, expertise, and credibility should guide the development of regulatory arrangements.

Effective regulatory institutions cannot be created overnight. Regulatory agencies may do best by starting with simple regulatory rules and procedures and refining them as they gain information and experience. In this way a balance can be struck between elaborating a complete regulatory framework at the outset and allowing regulatory procedures to evolve in response to sector developments and regulator abilities. There is a tradeoff between regulatory rigidity and flexibility. Going too far toward either extreme is not likely to yield regulatory institutions that perform well over long periods.

Notes

1. In many cases these sectors were once privately owned and were nationalized before or after World War II.

2. Distributors can, however, acquire a small amount of generating capacity for their own accounts.

3. In many countries, however, the customer with the worst payment record is the government itself.

4. Indeed, independent regulatory agencies in the United States have a somewhat ambiguous place in the country's constitutional structure, which is built on the separation of powers of the three branches of government: executive, legislative, and judicial. Some observers have argued that independent regulatory commissions at the federal level represent a fourth branch of government that is not provided for in the U.S. Constitution.

5. In what follows I assume that regulated firms must cover their costs from revenues generated by sales to consumers (Schmalensee 1989). Government transfer payments such as those that are central to the incentive mechanisms analyzed in Laffont and Tirole (1993) are assumed not to be available as components to regulatory incentive instruments.

6. The provisions for pass-through of generation-related costs may also have caused some distortions and induced distributors to enter into contracts with and take ownership positions in new generating plants that entered the market. However, the distributors are not allowed to own generating capacity in excess of 15 percent of the peak demand on their networks.

7. This is conceptually different from Laffont and Tirole (1993, ch. 1), where the regulated firm passes itself off as a high-cost firm by choosing a more lucrative regulatory contract than the one that merely satisfies the incentive compatibility constraint for the low-cost firm.

8. The considerations are different for a sector that is potentially competitive but where price regulation is thought to be required to respond to market power problems during a transition period until a sufficiently competitive market emerges.

9. Local Bell operating companies have not been permitted to participate in the inter-LATA (local area transport area) market in the United States since 1984. With the Telecommunications Policy Act of 1996, the Federal Communications Commission can lift these entry restrictions if a Bell operating company meets a long list of conditions to open up its local networks to competing suppliers of local and long-distance services.

10. With these assumptions there are either no active competitors other than the incumbent if $c_1 < c_2$ or the incumbent exits from the market if $c_2 < c_1$.

References

- Armstrong, Mark, Simon Cowan, and John Vickers. 1994. *Regulatory Reform: Economic Analysis and British Experience*. Cambridge, Mass.: MIT Press.
- Baron, David. 1989. "Design of Regulatory Mechanisms and Institutions." In Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*. Amsterdam: North-Holland.
- Baumol, William, and Gregory Sidak. 1994. *Toward Competition in Local Telephony*. Cambridge, Mass.: MIT Press.
- Boardman, Anthony, and Aidan Vining. 1989. "Ownership and Performance in a Competitive Environment: A Comparison of the Performance of Private, Mixed, and State-owned Enterprises." *Journal of Law and Economics* 32 (1): 1-33.
- Crampes, Claude, and Antonio Estache. 1997. "Regulating Water Concessions—Lessons from the Buenos Aires Concession." In World Bank, *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. Finance, Private Sector, and Infrastructure Network. Washington, D.C.
- Demsetz, Harold. 1968. "Why Regulate Utilities?" *Journal of Law and Economics* 11 (1): 55-66.

- Economides, Nicholas, and Lawrence White. 1995. "Access and Interconnection Pricing: How Efficient Is the 'Efficient Component Pricing Role'?" *Antitrust Bulletin* 40 (3): 557–79.
- Gilbert, Richard J., and David Newbery. 1994. "The Dynamic Efficiency of Regulatory Constitutions." *RAND Journal of Economics* 25 (4): 538–54.
- Goldberg, Victor. 1976. "Regulation and Administered Contracts." *Bell Journal of Economics* 7 (2): 426–52.
- Gray, Philip. 1997. "Colombia's Mixed Ownership Model for Privatizing Infrastructure." In World Bank, *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. Finance, Private Sector, and Infrastructure Network. Washington, D.C.
- Joskow, Paul L. 1974. "Inflation and Environmental Concern. Change in the Process of Public Utility Price Regulation." *Journal of Law and Economics* 17 (2): 291–327.
- . 1985. "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants." *Journal of Law, Economics and Organization* 1 (1): 33–79.
- . 1987. "Contract Duration and Relationship Specific Investment." *American Economic Review* 77 (1): 168–85.
- . 1998. "Electricity Sectors in Transition." *The Energy Journal* 19 (2): 25–52.
- Joskow, Paul L., and Richard Schmalensee. 1986. "Incentive Regulation for Electric Utilities." *Yale Journal on Regulation* 4 (1): 1–49.
- Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, Mass.: MIT Press.
- . 1996. "Creating Competition through Interconnection: Theory and Practice." *Journal of Regulatory Economics* 10 (3): 227–56.
- Levy, Brian, and Pablo Spiller. 1994. "The Institutional Foundations of Regulatory Commitment: A Comparative Analysis of Telecommunications Regulation." *Journal of Law, Economics, and Organization* 10 (2): 201–46.
- Lyon, Thomas. 1991. "Regulation with 20-20 Hindsight: Heads I Win, Tails You Lose." *RAND Journal of Economics* 22 (4): 581–95.
- Newbery, David, and Michael Pollit. 1997. "The Restructuring and Privatization of the CEGB: Was It Worth It." *Journal of Industrial Economics* 45 (3): 269–303.
- OSIPTEL (Organismo Supervisor de la Inversión Privada en Telecomunicaciones). 1995. *Regulación y mercado de las telecomunicaciones*. Lima.
- . 1997. *Boletín de Supervisión*. 1 (4). Lima.
- Posner, Richard. 1971. "Taxation by Regulation." *Bell Journal of Economics and Management Science* 2 (1): 22–55.
- Ruster, J. 1997. "A Retrospective on the Mexican Toll Road Program (1989–94)." In *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. World Bank, Finance, Private Sector, and Infrastructure Network. Washington, D.C.
- Schmalensee, Richard. 1989. "Good Regulatory Regimes." *RAND Journal of Economics* 20 (3): 417–36.
- Smith, Warrick. 1997a. "Utility Regulators—Decisionmaking Structures, Resources and Start-up Strategy." In *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. World Bank, Finance, Private Sector, and Infrastructure Network. Washington, D.C.
- . 1997b. "Utility Regulators—The Independence Debate." In *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. World Bank, Finance, Private Sector, and Infrastructure Network. Washington, D.C.
- . 1997c. "Utility Regulators—Role and Responsibilities." In *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. World Bank, Finance, Private Sector, and Infrastructure Network. Washington, D.C.

- Telefónica del Perú. 1996. *Prospectus* (May). Lima.
- . 1997. News Release (July 31). Lima.
- van de Berg, Caroline. 1997. "Water Privatization and Regulation in England and Wales." In *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. World Bank, Finance, Private Sector, and Infrastructure Network. Washington, D.C.
- Vickers, John, and George Yarrow. 1988. *Privatization: An Economic Analysis*. Cambridge, Mass.: MIT Press.
- Williamson, Oliver. 1976. "Franchise Bidding for Natural Monopolies—In General and with Respect to CATV." *Bell Journal of Economics* 7(1): 73–104.
- . 1983. *The Economic Institutions of Capitalism*. New York: The Free Press.
- Wolfram, Catherine. 1997. "Measuring Duopoly Power in the British Electricity Spot Market." Harvard University, Department of Economics, Cambridge, Mass.
- World Bank. 1997. *The Private Sector in Infrastructure: Strategy, Regulation, and Risk*. Finance, Private Sector, and Infrastructure Network. Washington, D.C.

Comment on “Regulatory Priorities for Infrastructure Reform in Developing Countries,”
by Paul L. Joskow

Isher Judge Ahluwalia

Paul L. Joskow’s article takes as its point of departure the basic reform model for infrastructure privatization, restructuring, and regulatory reform, emphasizing the enormous complexity of the process. Joskow rightly points out that introducing competition and deregulation in infrastructure sectors dominated by state-owned natural monopolies poses a challenge to a country’s politico-economic equilibrium. Thus any prescription for regulatory reform must take into account a country’s economic environment and its legal and political institutions.

I agree with Joskow that a well-defined privatization strategy should make clear the extent and pace of the privatization envisaged, identifying which sectors will remain regulated monopolies and how competition will be handled in areas where it is allowed. Such articulation will enable policymakers to identify possible problem areas and devise an appropriate regulatory regime. My comments are based on India’s experience with infrastructure reforms and how it relates to Joskow’s analysis.

In contrast to the coherent strategy spelled out in Joskow’s analysis, India has opted for a piecemeal process of infrastructure sector reform. The opening up of the infrastructure sector to private investment—including foreign investment—was initially based on the simple recognition that public resources could not cover the huge investments required in these sectors. Little attention was paid to creating the preconditions for attracting private investment. Given a tariff structure that would not ensure a remunerative return for investors, a lack of clear government policies on the interface between new private entrants and incumbent public enterprises, and the absence of independent regulatory agencies to ensure fair treatment for private investors, it was no surprise that private investment was slow to come. Although reform has moved slowly, it has moved in the right direction. I will highlight India’s experience with the ongoing reforms in the power and

Isher Judge Ahluwalia is director and chief executive at the Indian Council for Research on International Economic Relations, New Delhi.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

telecommunications sectors to drive home the importance of some of Joskow's points.

Power Sector Reform in India

Joskow's analysis of alternative pricing mechanisms leads him to conclude that prices initially need to be set high enough to attract investment capital. But in India's power sector, even when the government was willing to pay high prices for early projects, most projects had difficulty achieving financial closure because of uncovered risks perceived by lenders and investors.

The absence of an independent regulator, unsatisfactory fuel and transport arrangements, and the near-bankrupt status of state-owned power utilities acting as monopoly purchasers of power contributed to the heightened perception of risk, increasing the premium that had to be paid when determining initial prices. Such risks also needed to be brought within manageable proportions.

India's utilities pricing system used a cost-plus mechanism that guaranteed power stations a certain return as long as they operated above a minimum level of efficiency. The power purchase arrangements between public generating stations and public distributors were extremely vague, however, and did not make allowances for the risks involved. These include the risk of nonpayment by state-owned electricity boards, which have a monopoly on distribution; the risk of low offtake of power; and the risk of nonavailability of fuel when the power producer was required (under the agreement) to buy it from a specified location owned by a public monopoly. Private power producers wanted power purchase agreements with state electricity boards that allocated these types of risks within a commercial framework.

In the policy for private investment in power announced in 1992 private investors were offered a remunerative tariff based on a cost-plus formula with potentially attractive returns on equity linked to levels of capacity utilization. The new power purchase agreements were very different from existing arrangements in that the power tariff was expected to reflect explicit pricing of risks. This approach created problems, however: today public generating plants earn low returns based on the old pricing formula, while new private plants earn much higher returns and are protected against risks.

Power pricing policy for independent producers moved in the right direction in terms of attracting investment. But sharp public criticism has emerged because the process relied on negotiations rather than on competitive bidding to reach a mutually agreed cost-plus formula. The first new power purchase agreement, signed between the Maharashtra State Electricity Board and the Enron-sponsored power generation project at Dabhol, caused a political storm, raising questions about the reasonableness of capital costs and tariffs. Project construction was suspended for months, though the project has survived intense political scrutiny and public interest litigation. With some modifications to the original agreement, the project was restored and is due for commissioning. Recognizing the vulnerability of cost-based tariffs to public criticism and the need for transparency, the government has committed to using competitive bidding to reach such agreements in the future.

As for Joskow's hope that consumers will leave some rent on the table during the early years of the program, India's reform program has kept consumers in the dark by not spelling out the efficiency gains that are expected from competition. As a result there is a widespread perception that private producers are getting too good a deal—particularly relative to public producers, who are operating under suboptimal pricing.

Yet power prices to Indian consumers, especially rural and household consumers, are heavily subsidized. Notwithstanding heavy cross-subsidies from industrial and commercial consumers, the average revenues realized from all consumers in 1996–97 covered just 80 percent of the average cost of power. Such shortfalls further undermine the financial health of state electricity boards, which are already weighed down by their own inefficiency.

The financial nonviability of state electricity boards has implications for the financing of private power projects. These projects typically are financed on a non-recourse basis, and the risks associated with them are high because of their high capital intensity. The risks associated with the revenue streams from these projects are increased by the risk of nonpayment by the monopsonist buyer. The only solution to this problem is to fundamentally reform the state electricity boards and develop a tariff policy for power that moves away from subsidies.

The story of reform in the power sector in India has been one of learning by doing. Since the reform strategy has not been clearly thought out, solutions are sought at every stage as unanticipated problems arise, so the main concern is often viewed as one of slow implementation.

Telecommunications Reform in India

India's telecommunications sector demonstrates the importance of ensuring access on suitable terms by competitive suppliers to the regulated portion of a monopoly network—a point Joskow emphasizes. As in many countries, local call charges are kept low in India, and the local network is effectively subsidized by monopoly profits on long-distance calls. Private operators can compete with the public telephone system at the local and regional levels, but long-distance traffic remains a public monopoly, putting private operators at a disadvantage. If there is any justification for cross-subsidization, the same subsidy should be provided to private operators or the long-distance system should be privatized.

Although telecommunications was privatized through transparent and open bidding, colossal delays have occurred in closing deals because the basic framework was inadequately defined. Moreover, post-bid negotiations with successful bidders, which affect the terms of the initial bids, have been messy. This experience has demonstrated the crucial need to develop a basic policy and regulatory framework before inviting private entities.

A significant feature of telecommunications reform in India was the creation of an independent regulatory authority. The authority gained credibility through an early ruling in favor of private investors in a dispute between private investors and

the Department of Telecommunications. And unlike in the power sector, the tariff structure in telecommunications is financially viable. Although the public sector has exploited its monopoly position to overcharge high-income consumers, remunerative tariffs and rapid growth in demand have created a favorable environment for private investment. The main problem facing telecommunications reform is the resistance of the public monopoly, the Department of Telecommunications, to providing competitive access to its monopoly network for private competitors offering local and regional services.

Comment on “Regulatory Priorities for Infrastructure Reform in Developing Countries,” by Paul L. Joskow

Kotaro Suzumura

Paul L. Joskow’s article provides a careful analysis of regulatory reform of infrastructure sectors in developing countries—one that will prove useful for those in charge of designing and implementing such reform. Although I agree with most of Joskow’s points, I will qualify or even contradict some of his main assertions with alternative viewpoints based on my study of reform in Japan’s telecommunications sector (see Suzumura 1995, 1998). Obviously, Japan’s experience with telecommunications reform is not a universally applicable reform model for developing countries. Still, some lessons can be drawn from the country’s rocky reform experience.

Designing the Reform Model

To begin with, consider a network infrastructure industry with economies of vertical integration between natural monopoly segments and potentially competitive segments. A typical example is telecommunications, where local service is usually considered a natural monopoly and long-distance service is potentially competitive.

Now suppose that a vertically integrated monopoly initially prevails under public regulatory control, and consider the basic industry structure from the reform designer’s viewpoint. According to Joskow, there are two basic policy options.¹ The first is to allow the incumbent monopoly to remain vertically integrated but to encourage entry into potentially competitive segments by ensuring new competitors fair access to the incumbent’s network facilities. The second option is to divest the vertically integrated monopoly into natural monopoly segments that are kept under regulatory control and to facilitate entry into competitive segments by ensuring all competitors fair access to the natural monopoly network facilities.

Comparing these two reform options, Joskow rightly emphasizes that implementing the first requires considerable regulatory effort to ensure that the terms and conditions of access to network facilities do not discriminate against the network

Kotaro Suzumura is professor of public economics in the Institute of Economic Research at Hitotsubashi University.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

operator's competitors. But fair access to network facilities is no less important under the second option. In other words, fair network access rules are of vital importance whenever competition is introduced in potentially competitive segments, regardless of whether the incumbent monopoly is divested.

Joskow points out that most theoretical discussions of network access focus on the prices charged for access. He notes that while access pricing is important, network access issues encompass a much broader set of terms and conditions than pricing alone. Moreover, nonprice terms and conditions often involve much more subtle issues than do narrow access pricing issues. Yet Joskow's discussion of network access focuses almost exclusively on access pricing. Nonprice terms and conditions are justly emphasized, then left unanalyzed.

Given the importance of fair network access, systematic analysis is needed of price as well as nonprice network access terms and conditions. To illustrate the issues to be addressed in such an analysis, consider the institutional design and enforcement of fair network access rules. A strong case can be made for separating the roles of rule designer and referee on the one hand and regulator on the other. That is, one agency should be responsible for designing fair network access rules and settling disputes over them; another should handle public regulatory control of bottleneck natural monopolies. If these roles fall under the jurisdiction of the same regulatory agency, there is no systemic safeguard that network access rules will be tailored to the promotion of public welfare rather than to the promotion of the bureaucratic purpose of the agency. But even if separating these roles is considered faulty or infeasible, a general point still holds: designing fair network access rules and assigning enforcement and dispute settlement responsibilities are an integral part of infrastructure reform.

One additional point is worth noting. The border that separates natural monopoly segments and potentially competitive segments varies fluidly in response to changes in technology. In telecommunications, for example, serious competition in the local exchange market was almost inconceivable in 1985 or even in 1990. Yet competition is now becoming eminently feasible by virtue of wireless systems and integrated wired systems that can be digitized. Thus the design of the basic industry structure cannot be settled once and for all, as superficial readers of Joskow's article may presume. To the contrary, the task calls for continuous monitoring of reform implementation and innovative rethinking about additional reforms.

Implementing the Reform Model

Joskow identifies two prototypical approaches to implementing the chosen reform model. The first is the big bang approach, in which privatization, restructuring, and competition are introduced at one stroke. The second approach provides for a transition period during which the industrial organization and associated regulations evolve according to a planned program.

It seems to me that the second approach is almost a contradiction in terms. On the one hand, if the industrial organization and associated regulations are allowed to evolve spontaneously in privatized competitive segments, there is no guarantee

that this evolution will follow the structured transition program. On the other hand, if privatization, restructuring, and competition are introduced according to a premeditated reform plan, gradual changes in industrial organization and associated regulations are nothing more than the piecemeal materialization of an initially planned order. In this case competition cannot help but be straitjacketed by meticulous regulation to avoid divergence from the structured transition program.

To put this point somewhat differently, if competition is allowed to play the role of the Hayekian discovery procedure, thereby allowing spontaneous evolution of the industrial organization, the transitional approach to reform is necessarily unstructured gradualism.² This is in contrast to Joskow's clearly structured transition program, in which the initial design of the reform model must be carefully monitored and deliberately reoriented, reflecting the spontaneous evolution through market competition.

The unstructured gradual implementation of a reform plan, coupled with the flexible redesign of a reform model paying close attention to the spontaneous evolution of industrial organization through market competition, may precipitate opportunism as well as lack of commitment to the original reform plan. For this unstructured gradualist approach to have any chance of success in developing countries, two prerequisites must be satisfied. First, the regulatory mechanism should not handcuff competition in the liberalized segments; competition must be allowed to function as a discovery procedure. Second, the public decisionmaking mechanism must be capable of adjusting the reform plan in response to changing environments and of motivating agents to comply voluntarily with the shifting reform plans.

These are not simple prerequisites, but that should come as no surprise. After all, sensible regulation and public decisionmaking are an important—if intangible—component of a country's infrastructure. It is natural that the successful design and implementation of infrastructure reform hinge squarely on the provision of such intangible mechanisms.

A final remark is in order on the implementation of a chosen reform model. My emphasis on the provision of sensible regulation and public decisionmaking focused on the valuable role that these mechanisms can play in enabling competition to function as a discovery procedure or in adjusting the reform plan flexibly and motivating agents to comply with shifting plans. But these mechanisms may also have intrinsic values of their own (see Malinvand and others 1997 and Suzumura 1999). Paternalistic regulation that handcuffs competition may vitiate the value of freedom of choice, and public decisionmaking that imposes an inflexible reform plan at one stroke or in accordance with a clearly structured transition plan may run counter to the value of participatory democracy. During reform design and implementation care should be taken to preserve both the instrumental and intrinsic values of these mechanisms.

Japan's Telecommunications Reform: Teaching by Negative Example?

Japan's experience with telecommunications reform may lend concreteness to what I have said so far. Until 1985 Japan's telecommunications industry was legally

monopolized by a public corporation. The 1985 reform privatized this corporation, but the newly created private corporation, Nippon Telegraph and Telephone (NTT), remained vertically integrated. Potentially competitive segments of the industry were liberalized, and new entry of telecommunications carriers was strongly encouraged.

Two idiosyncrasies of this reform stand out. First, the Ministry of Posts and Telecommunications continued to heavily regulate both the vertically integrated carrier, NTT, and the new carriers in competitive segments. Second, the reform deferred the important public decision about whether NTT, which inherited the bottleneck network facilities from the public corporation, should be divested into natural monopoly segments and competitive segments. Instead, the reform plan merely obligated the government to review this issue in 1990. Thus Japan's 1985 reform followed Joskow's first model of industrial organization design, while implementation of the reform followed the model of unstructured gradualism.

Handcuffed competition is still competition, and after the reform the telecommunications industry saw a remarkable drop in charges in competitive segments. Charges in noncompetitive segments were kept basically unchanged by regulation. A devoted consequentialist, whose judgments on a policy are focused exclusively on outcomes, might claim that there is nothing wrong with handcuffed competition. However, not only were the regulatory criteria and procedures guiding handcuffed competition nontransparent, but paternalistic regulation by the Ministry of Posts and Telecommunications made competitors care more about the regulator's discretion than their rivals' competitive strategies—to the detriment of public welfare.

The programmed review of the 1985 reform was also problematic, as Japan's public decisionmaking mechanism was not up to the task of adjusting the reform plan in response to a changing environment or of motivating agents to comply voluntarily with public decisions. Indeed, the reviews in 1990 and 1995 resulted in no conclusive public decision, and only in 1997—12 years after the original reform—was agreement finally reached on the managerial form of NTT. Great uncertainty in Japan's telecommunications remained during this long and crucial period.

Reform of Japan's telecommunications sector is hardly a success story. It highlights the importance of reforming the intangible infrastructure of public regulation and decisionmaking before tangible infrastructure reform can be successfully designed and implemented.

Notes

1. Joskow also examines a third option of retaining the integrated regulated monopoly without competition in potentially competitive segments, thereby preserving economies of vertical integration. He claims that this option, combined with privatization and regulatory reform, is potentially viable in countries where appropriate regulatory and commercial arrangements can be introduced to make regulated monopoly infrastructure segments perform well. But I do not believe that this option qualifies as a viable reform model, because turning a state monopoly into a private monopoly is unlikely to help create a more dynamic market economy (Stiglitz, in this volume).

2. Hayek (1978) wisely observed that competition is a decentralized procedure for discovering, through successive trial and error, who is best suited to a specified task when no one has all the relevant information. This type of competition is more important the less transparent are the market contingencies due to a rapid and unforeseen technological progress, so that no one, including the government, can predict with any degree of confidence which way and how fast progress will be made.

References

- Hayek, Friedrich A. 1978. "Competition As a Discovery Procedure." In *New Studies in Philosophy, Politics, Economics, and the History of Ideas*. London: Routledge & Kegan Paul.
- Malinvaud, Edmond, Jean-Claude Milleron, Mustapha K. Nabli, Amartya K. Sen, Arjun Sengupta, Nicholas Stern, Joseph E. Stiglitz, and Kotaro Suzumura. 1997. *Development Strategy and Management of the Market Economy*. vol. 1. Oxford: Clarendon Press.
- Suzumura, Kotaro. 1995. *Competition, Commitment, and Welfare*. Oxford: Clarendon Press.
- . 1998. "Japan's Industrial Policy for Telecommunications: The 1985 Institutional Reform and Its Aftermath." In Konosuke Odaka and Juro Teranishi, eds., *Markets and Government: In Search of Better Coordination*. Tokyo: Maruzen.
- . 1999. "Consequences, Opportunities, and Procedures." *Social Choice and Welfare* 16 (1): 17–40.

Floor Discussion of “Regulatory Priorities for Infrastructure Reform in Developing Countries,” by Paul L. Joskow

Jean-Jacques Laffont (presenter in another session) noted that all the speakers had mentioned the importance of nonprice issues in determining competitors’ access to the regulated portion of a monopoly network. Laffont and Jean Tirole have developed a global price cap approach that helps to address those issues. By delegating pricing to monopolists, the global price cap decreases incentives to foreclose entry into the market. Moreover, fears that a global price cap can lead to predatory pricing can be eased by using the efficient component pricing rule. Finally, said Laffont, a global price cap makes access prices sensitive to demand—a task that is too complicated for regulators.

Paul L. Joskow (presenter) agreed that global price caps were an elegant solution to a complicated problem in the sense that they align the incentives of the incumbent monopolist so that if it is economical, the monopolist will want to provide access to its network. The problem Joskow had with global price caps is convincing people that it makes sense to allow the incumbent monopolist to set the access price. The general assumption is that the incumbent will try to hurt its competitors. In addition, it is difficult to set the global price cap at the right level and to deal with the predatory and other types of behavior in which an incumbent might engage. In the end, Joskow concluded, global price caps were a useful way of thinking about the problem but were difficult to implement in most countries.

A participant from the World Bank believed that Joskow’s presentation provided an excellent framework for the World Bank’s operational dialogue with its client countries—especially the idea that reform should be phased and anchored in the industrial organization. Still, the participant felt that the framework left some questions unanswered. First, reform priorities differ according to the extent of development of the sector—for example, whether 30 percent or 100 percent of the urban population receives water services. Second, Joskow’s framework failed to address the important issue of the supply of (and need for) entrepreneurialism, ability, and finance in a reforming sector.

This session was chaired by Lyn Squire, director, Development Policy, at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

Joskow replied that he hoped that he had made clear the need for reform to be sensitive not only to the starting point in the sector but also to the political and legal environment in the sector and economic conditions in the country. Joskow noted that he had offered three alternatives, one of which is simply a regulated monopoly. Though this form is often criticized, it has worked well in the United States—for example, electricity, water, and telephone services are almost universal. And while a regulated monopoly might not be the most efficient approach, it is by no means a disaster. Countries that have poor performance, low service penetration, high costs, and limited resources should consider relying on a regulated monopoly in a sector like water, where there is less experience with competitive systems. Similarly, some countries may be too small to have a competitive electricity sector, for example, so competition may not be the best approach. The solution has to match the problem, said Joskow, and certain features are essential to attract investment.

As for entrepreneurial ability, continued Joskow, the issue turns on whether foreign companies are allowed to enter the market. There is no shortage of entrepreneurial talent in infrastructure markets, and when the prices offered to private providers cover the costs of operating a system, many companies with good track records will offer bids. The issue becomes more complicated when infrastructure assets are required to have domestic owners, because then political considerations are at work. In Mexico, for example, the constitution requires that electricity be supplied by the state. Reforms in such an environment must be tailored to political realities—or the political environment must change.

Paul A. David (discussant in another session) asked Joskow to elaborate on non-price access issues in telecommunications in view of Joskow's endorsement of regulated monopolies. What should be done, David asked, when the incumbent resists an entrant that wants to upgrade standards to increase compatibility? For example, a cellular phone operator may want to add a service that requires upgrading the monopoly's network. If the incumbent is not interested in offering that service, or if doing so would require investment, the incumbent may simply opt not to do so. Should such conflicts be subject to arbitration, perhaps by international bodies or multilateral organizations, to ensure that developing countries served by a regulated monopoly have the same service options as countries with more competition?

It depends, said Joskow, on why the incumbent network operator is resistant to providing the new service. If it is because the incumbent wants to avoid competition, then that is a good reason to keep the company out of competitive lines of business. But if it is because the company is concerned about costs and lacks knowledge about the best approach to providing the new service, then those issues arise in all network industries.

One solution being tried in the United States offers promise, thought Joskow, at least for electric power. The state of California has introduced a competitive electricity market. An independent system operator with a professional management team operates the transmission network and controls the terms and conditions of access to it. The managers of the independent system operator have no financial stake in the generation or sale of electricity. Instead, their objectives are to ensure

that the electricity supply is reliable, to provide reasonable interconnection for competitive generators and distributors and a clearinghouse for marketers, and to evaluate the efficiency of the electricity market. The managers possess technical expertise as well as the authority to order owners of transmission distribution systems to add facilities and receive compensation for them. But there are no conflicts of interest.

A participant from Columbia University asked how government can maximize revenue when privatizing infrastructure services. In Joskow's view maximizing revenue is not a sensible goal for privatization. Still, he offered a possible approach: privatizing a provider with a permanent monopoly and no regulatory intervention. The firm would then figure out how to maximize its surplus, which could be factored into the purchase price.

A participant from the World Bank asked Joskow about his sanguine view of Peru's approach to privatizing telecommunications services. True, she said, the extended transition to widespread competition provides time for the regulator to establish itself and for the private investor to build up the system. But it also gives the monopolist time to build its political power base and bribe officials so that it can retain its monopoly. At the same time, the fact that consumers will "leave some rents on the table" during the early years of the program might make the regulatory pricing framework less credible to new entrants, who will fear political backlash against large rents being captured by the monopolist.

Joskow said that he had acknowledged that tradeoff. In his judgment the most pressing needs for Peru's telecommunications sector are investment capital, increased efficiency, and lower prices. The welfare losses from any abuse of market power are small relative to continued inadequate telephone coverage and extremely poor service. Moreover, the issue is not just economic, but also political. Peru has established an independent regulator, and the next few years will demonstrate the government's commitment to telecommunications reform. According to Joskow, a more sophisticated approach to competition and regulation would not have been feasible, at least not in the early 1990s. It was also important to implement a commercial contract that was enforceable by the depoliticized court system. There may be faults with that approach—particularly if a single regulated monopoly is considered an undesirable outcome—but given Peru's low starting point, the five-year concession seemed to Joskow like the right approach.

A participant from the University of Maryland asked Joskow about the sequencing of restructuring and privatization. In Norway electricity services were restructured but they were never privatized. Yet it had much better results, in terms of lower prices, than England and Wales, where restructuring was followed by privatization.

Not many other countries offer lessons, reported Joskow. Norway already had competition, and reforms introduced considerably more competition. In some sense, however, the reforms in England and Wales have been more successful, because the incumbent system was much less efficient than in Norway. As a result there have been significant cost savings and welfare gains in England and Wales. The main question is whether there are enough competitors—an issue that may also arise

in Norway as mergers continue. In fact, said Joskow, that issue will eventually arise in every infrastructure sector subject to reform, as countries try to strike a balance between increased efficiencies from mergers and competition. The United States, for example, confronted that issue in the long-distance telephone market. For many years American Telephone and Telegraph's (AT&T's) prices were capped because it was thought that the company had significant market power. Deregulation was incomplete because entrants had not achieved substantial market share.

Lyn Squire (chair) closed by complimenting Joskow's comprehensive approach and practical advice—particularly his emphasis on the importance of a country's institutional environment in shaping policy advice. According to Squire, Joskow's article laid the foundation for a potentially valuable empirical and analytical work program for the World Bank. The Bank should pursue such a program, said Squire, and invite Joskow back in a few years to assess the progress.

Competition, Information, and Development

Jean-Jacques Laffont

Developing countries are hamstrung by a wide range of weak institutions. Tax systems are inefficient. Auditing expertise is lacking. Technological knowledge is poor. Corruption is rampant. Financial systems are inefficient. Governments lack credibility. And politicians and bureaucrats are captured by special interests. This article asks whether and how each of these problems calls for a different view of competition, to what extent competition helps solve these problems, and to what extent competition can be implemented. Although competition generally has a positive effect, care must be taken in implementing it where institutions are weak. Thus it is crucial for donors to support institutional improvements and to condition aid on such improvements.

One way to provide more effective incentives, including enhanced consumer orientation, is to extend the scope for competition.

— Joseph E. Stiglitz (1997, p. 20)

Competition is an unambiguously good thing in the first-best world of economists. That world assumes large numbers of participants in all markets, no public goods, no externalities, no information asymmetries, no natural monopolies, complete markets, fully rational economic agents, a benevolent court system to enforce contracts, and a benevolent government providing lump sum transfers to achieve any desirable redistribution.

Because developing economies are so far from this ideal world, it is not always the case that competition should be encouraged in these countries. Economic theory—in particular, industrial organization theory—offers myriad examples where some

Jean-Jacques Laffont is director of Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France. The author is indebted to Jean-Paul Azam, Antonio Estache, and Patrick Rey for discussions on this article and to Timothy Besley, David Martimort, Paul Seabright, and Jean Tirole for comments on a previous draft.

Annual World Bank Conference on Development Economics 1998
©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

form of competition may be detrimental in industrial countries. Patent policies that limit competition to create incentives for innovation help deal with the public-good nature of discoveries (Aghion and Howitt 1997). More generally, restricted ex post competition is often the only way to encourage investment in specific nonobservable assets in a relationship or an organization. Natural monopolies require a small number of participants. In such markets competition may initially be organized (in the form of auctions), but once market participants are in place competition must be restricted to avoid a wasteful duplication of fixed costs. Similarly, recent literature on finance views competition as increasing the probability of bank runs and systemic crises (Caminal and Matutes 1997a, b; Mishkin 1997). More generally, market competition for inputs and outputs affects the efficiency of firms and organizations, with ambiguous effects on the level of competition.¹

The general philosophy behind these examples is that in the absence of one of the assumptions sufficient for establishing the fundamental welfare theorems, as second-best theory suggests, restricting competition may be beneficial. These examples must be evaluated with care, however. Competition should be restricted only if the impediment to the validity of one assumption is major and cannot be eliminated quickly enough. Otherwise, the solution is to remove the impediment and benefit from competition. The choice between the two options becomes a question of opportunity, which can lead to conflicting views according to subtle dynamic considerations. What is meant by quickly enough? How hard is it to remove the impediment? How irreversible will the limitation of competition be? Do we take a purely normative view or do we take into account political or administrative constraints? I have phrased the policy problem in terms of restricting competition. Even then the implementation of such a policy is not obvious, as black markets can spring into existence. The implementation problem appears even more clearly when the preferred policy is to encourage competition. In the standard general equilibrium model, competition is the result of a large number of participants, complete markets, and so on. How can competition be implemented in the absence of these conditions?

A simple case illustrates the methodology I use in my analysis of developing countries. Consider Buchanan's (1969) example of a polluting industry. In the absence of an externality tax, a monopoly organization of the industry (which contracts output) may coincide with the first-best production level and dominate a competitive industry that overproduces. Should competition be restricted in such a situation? Probably not. It seems more reasonable to establish a tax to internalize the externality and benefit from the dynamic virtues of competition. But what if the agency in charge of the environment is nonexistent, poorly staffed, or captured by the industry, or if the pollution is diffuse and cannot be measured at the individual level? A barrier to entry, such as a license to operate, may then be the only way to limit production and therefore pollution, at least if this policy can be implemented and is not a pretext for rent seeking. The right policy answer should take into consideration many aspects of the problem that are not easily measured or even modeled. Thus it is not surprising that the right answers may

differ by industry or country, particularly between industrial and developing countries.

One approach to studying competition in developing countries is to examine measures taken to encourage competition—that is, competition policy—in industrial countries and see how conclusions should be modified for developing countries. This is the path followed, for example, by Rey (1997). I take a different perspective. Considering the major structural problems of developing countries, I ask whether and how each of these problems calls for a different view of competition, to what extent competition helps solve these problems, and to what extent competition can be implemented. In so doing I take a broader perspective than is usual in discussions of competition policy.

The analysis of such a large question cannot be exhaustive. I try to find examples that suggest policy prescriptions specific to developing countries. I consider problems that are universal and yet are particularly acute in developing countries: inefficient tax systems, inadequate auditing expertise, limited technological knowledge, corrupt institutions, inefficient financial systems, lack of government credibility, and capture of politicians and bureaucrats. Drawing on industrial organization theory, I ask whether increasing competition helps solve these problems.

Inefficient Tax Systems

Under the weight of fiscal imbalances many developing countries are finding it increasingly difficult to invest in infrastructure and to protect real spending on human resources and antipoverty programs. And many are unable to control large-scale corruption, to introduce organizational reforms within the public sector, or to effectively regulate the private sector.

— Dilip Mookherjee (1998, p. 104)

In most industrial countries it costs between 0.3 and 0.5 to raise 1 unit of public funds. In developing countries the costs can run much higher: between 1.19 and 1.54 in Thailand, 1.20 in Malaysia, and 2.48 in the Philippines (Jones, Tandon, and Vogelsang 1990). Obviously many developing countries must make serious efforts to improve their tax systems.

Two types of reforms can be used to achieve that goal: those that improve the auditing technology of tax inspectors, increase the number of tax inspectors, and mitigate corruption and those that use available information to eliminate historical (often politically motivated) anomalies of the tax system (Ahmad and Stern 1991). Both types of reforms are difficult: the first type because developing countries often lack the resources—in particular, human resources—for such improvements, which are bound to be slow anyway, and the second type because it meets political resistance by powerful interest groups.

The Usual Role of Cross-Subsidies

The enormous differences between the tax systems of industrial and developing countries have strong implications for other areas of public policy. In infrastructure financing, for example, cross-subsidies are a common substitute for tax revenues. Cross-subsidies have a bad reputation among some economists, particularly at the World Bank. Furthermore, competition kills cross-subsidies through cream skimming. Still, it is possible to rehabilitate cross-subsidies in developing countries in a way that is compatible with some form of competition in infrastructure financing.

For example, Africans have extremely poor access to basic public services, with access to safe water averaging 45 percent, to electricity 30 percent, and to telecommunications mainlines 1 percent (Kerf and Smith 1996). Consider the problem of providing universal access to telecommunications (or electricity, water, or transportation) in an African country composed of a rich city and a poor rural area. The European or U.S. solution to such a problem would be to open the city to competition, define a level of service to be provided in the rural area, and organize an auction (financed by a general tax or a tax on telecommunications in the city) to determine the subsidies needed to convince an operator to become the universal service provider in the rural area.

But even in the United States this approach can be extremely costly. Hausman (1997) calculates the social deadweight loss associated with using long-distance telephone charges to subsidize schools and libraries and obtains a 0.86 cost of funds—compared with a general estimate of around 0.3 in the United States. This high cost is due to a sizable elasticity of demand for long-distance services (0.7), the existence of other taxes on long-distance services, and a high markup above marginal cost by long-distance carriers. Hausman argues that a monthly subscriber charge would have a much lower social cost.

Moreover, an incumbent's advantages in rural areas often undermine competition in the auction for universal service. This is one reason why the U.S. Federal Communications Commission constructs costly proxy models, so that it can impose (if credible) a reservation value for the subsidies and maintain competition both in the market (ex post competition) and for the market (ex ante competition).

This approach to providing universal service is probably not relevant for developing countries, where it would be even harder to implement. In light of these countries' inefficient tax systems, it is likely that once competition was established in the city, driving prices down to average costs, the tax authorities would be unable to deliver the tax money needed to subsidize operators in the rural area. Even if they did, it would be at a high social cost—perhaps three times the level of the subsidy—because of the high cost of public funds, the lack of expertise to evaluate the cost of universal service, and the lack of competition in the auction for universal service. In the end the network would not be developed in the rural area, as has been the case in Africa.

A Different Approach

Thus it is worth considering a very different approach. First, the notion of universal service must change. It must be enlarged to include delivery of a service, but not

necessarily connection to a network. Then the best use must be made of alternative technologies—solar energy for electricity, mobile phones for telecommunications, wells for water—and the network in the rural area must be carefully designed and appropriately sized. With these conditions in place an obvious alternative for financing network expansion is with cross-subsidies, within a franchised operator's accounts, between the rich customers in the city and the poor ones in the rural area.

Leaving the money with the operator avoids the inefficiency and corruption of the tax system, the more so if managers of the firm are sensitive to Hausman's arguments and are not prevented by regulators from using efficient pricing methods such as two-part tariffs. From a financing viewpoint this method is clearly more efficient (see Laffont and N'Gbo 1998 for more details). But such a method could not survive the liberalization of telecommunications in the city, since profits in the city would be skimmed off, leaving no money for the rural area.

The appropriate response is not to destroy cross-subsidies by liberalizing telecommunications in the city. Rather, it is to define several territories, each including a piece of the city and a piece of the rural area, and offer them to service providers through competitive bidding, with appropriate guidelines for service quality and network expansion in the rural area.² In addition to benefiting from *ex ante* competition, this approach offers possibilities of *ex post* yardstick competition between areas, at least at the next bidding stage. If competition is insufficient, as it might be in many countries, these franchising contracts could be designed by the regulatory authorities. The danger then is the capture of the regulatory agencies, and the empirical question remains whether a relatively sophisticated franchise can be designed and monitored more effectively than the tax system. That is where international aid could focus its attention, by providing noncorruptible expertise to help design these regulations and *ex ante* competition rules. Such a policy might be cheaper and easier to implement quickly than a reform of the tax system.

Gasmi, Laffont, and Sharkey (1998) provide an empirical method of exploring the relevance of cross-subsidies in a modern regulatory framework; I use that approach (developed in Gasmi, Laffont, and Sharkey 1997) here. An engineering model of the cost of local exchange telecommunications firms is extended to allow for information asymmetries between the regulator and the regulated firm.³ The model allows for two types of regulatory rules to be compared for an area composed of a central business district and a suburban area. The first involves the optimal regulation of a monopoly that has a universal service obligation; the second introduces competition in the business district, but the former monopoly still must satisfy the universal service obligation.

In both cases transfers from the regulator are possible, but they entail a social cost of public funds. Competition in the business district limits the information rent in that area but duplicates fixed costs and lowers revenues there. More public funds are then needed in the suburban area. As the cost of public funds increases, the first option improves and dominates beyond some value of this cost. As long as the tax system remains weak, this option is preferable to a form of competition that kills cross-subsidies, the only available financing method. Cross-subsidies should not be

demonized. They can play a useful role in network industries when the tax system is failing.

Lack of Expertise in Monitoring, Auditing, and Enforcement

My impression is that many former command economies and developing countries face a particular deficiency whose importance is easy to underestimate. This pertains to the critical shortage of what might be called “the human capital of capitalism”—legal, managerial, economic, accounting, statistical, etc., required to effectuate and operate a market economy and from a public sector perspective to regulate or otherwise address its dysfunctions and limitations effectively.

— Michael J. Trebilcock (1996, p. 42)

The lack of resources and human expertise for monitoring, auditing, and enforcement is pervasive in developing countries (see Gould and Amaro-Reyes 1983). But these shortcomings are usually underestimated, despite their dramatic implications for many areas of public policy. They are a major cause of the inefficient tax systems discussed above.⁴ And as analyzed below, they foster corruption.

One of the best ways to combat corruption is to establish sound financial management practices, including a timely and efficient accounting system and punctual, professional review by internal and external auditors (Wesberry 1996). Inadequate monitoring and auditing limit regulatory options in monopolistic or oligopolistic sectors, and poor information or bad auditing of costs makes regulation particularly inefficient in developing countries. Indeed, bad auditing forces regulators to use high-powered incentive schemes (such as fixed price contracts), thereby forgoing large information rents, encouraging corruption, and requiring (costly) public funds. What, then, can be expected from fostering competition?

A Principal-Agent Approach to Information

To explore this point in more detail, consider a simple agency relationship, such as Laffont and Tirole’s (1986) regulation model with cost observability. The regulated firm, or agent, has an unknown cost characteristic for realizing a project and can lower that cost by exerting effort. The regulatory authorities, or principal, are benevolent. The firm can be either efficient or inefficient, and this is its private knowledge. Optimal regulation entails a tradeoff between efficiency and rent extraction. The incentives for an inefficient type are reduced to decrease the information rent of an efficient type. The higher is the subjective probability that the firm is efficient, the more acceptable is distortion in the inefficient type’s incentives, because the higher is the expected cost of the rent.

How can we model the information effect of competition in a simple way? Suppose that the information provided by the competitive environment is correlated

with the type of firm facing the regulator. It is as if competition were providing a supervisory function in the agency model. Suppose that the supervisor discovers the firm's type with some probability or otherwise observes nothing (as in Tirole 1986). When the firm's type is discovered, the regulator can offer a fixed price contract that extracts all the rent. In this case the power of incentives is higher, as is the probability of such a discovery. But when the firm's type is not discovered, the optimal response of the regulator to better supervision technology depends precisely on the modeling of that technology.

In some cases better technology and low-powered incentives are substitute instruments to extract rents. Then an increase in competition increases incentives, the more so if a country has a high cost of public funds and poor supervision technology. Competition has an even greater effect on incentives in developing countries than in industrial countries and is even more desirable socially. In other cases better technology and low-powered incentives are complementary instruments to extract rents, and the results are reversed. Finally, there is a general equilibrium effect that favors stronger incentives because, as information improves, the rents to give up decrease and consequently the cost of public funds decreases. Taking into account all the effects, it can be shown that, in general, the expected power of incentives increases with this type of competition, more so in a developing country (see Boyer and Laffont 1998).

The information effect of competition can be modeled differently, however. By allowing the principal to observe contracts in place in similar agency relationships, competition provides a better evaluation of the distribution of costs facing the principal. This better information will induce a more dispersed offer of contracts. The principal now has a stronger belief that he or she is facing either an efficient type or an inefficient type. For each signal the optimal tradeoff is driven by the shape of the distribution of beliefs, and there is no simple result, even when the information structure becomes finer. However, Laffont and Tirole (1993) show that if a finer information structure corresponds to a partition of the support of beliefs in subintervals, then expected incentives will increase. And for the two-types case and small uncertainty, expected incentives are constant (as long as both types are kept) and they increase when the inefficient type is shut down. Furthermore, the general equilibrium effect always favors stronger incentives. Despite a number of ambiguities, we can conclude that in general the information effect of competition (which, in this case, is obviously favorable from a welfare point of view) will induce higher-powered incentive schemes, the more so the less developed the country.⁵

The Role of Increased Competition

The analysis remains cloudy because it has not explicitly modeled how increased competition improves the principal's information. So consider the more obvious case of increased product competition. A competitive sector produces a good that is a demand substitute for a monopoly's good. If the two goods are strategic substitutes, then increased competition in the form of a lower marginal cost for the com-

petitive sector decreases production and therefore incentives in the monopolistic sector. The opposite occurs if the goods are strategic complements (see Boyer and Laffont 1998). Results are mixed on the effect of greater external competition on incentives within organizations. More competition may be good, as in the examples above, without creating stronger incentives.

It is also possible to introduce competition in the public sector. Increasing competition—either *ex ante* (in the form of auctions) or *ex post* (with some duplication of fixed costs)—is even more profitable in developing countries than in industrial countries, where regulators can directly reduce the extent of asymmetric information and where the auditing of costs makes possible the use of less powerful incentives that lower information rents (see Laffont and Tirole 1993; Auriol and Laffont 1993; Dana and Spier 1994; and Anton and Yao 1987, 1989, 1992).

Consider a procurement auction where firms differ in efficiency and can affect their cost by effort—the setting of Laffont and Tirole (1993). We can calculate for the optimal auction the gain of having an auction. Take two situations in which the winner must bid against other bidders or be regulated as a monopoly. The gain in transfer from the auction increases with the size of the information asymmetry between regulators and firms. A poor auditing mechanism leaves this asymmetry large. Thus introducing competition generates a double gain in developing countries because it *shrinks this asymmetry and so lowers costly public transfers*. On the other hand, the cost of generating competitors is likely to be higher in developing countries for lack of expertise and for other reasons discussed below (corruption, lack of commitment, and so on). Competition is even more desirable than elsewhere but is more costly to implement.

Traditional competition policy is another area of concern for weak monitoring and auditing systems. After reviewing the theoretical and empirical literature on the effect of competition on growth in industrial countries, Rey (1997) analyzes traditional competition policy from the perspective of developing countries. He concludes that collusion is facilitated by entry barriers, market concentration, and capacity constraints and that these factors are more likely in developing countries (see Bernheim and Whinston 1992 and Compte, Jenny, and Rey 1997). Moreover, the transactions costs of collusion may be lower in developing countries because of an implicit lower risk aversion for economic agents who have little to lose (though this point is debatable) and because of less efficient monitoring technologies (see Faure-Grimaud, Laffont, and Martimort 1998). Similarly, predatory strategies may be particularly dangerous in developing countries because credit markets are weak. Rey (1997) argues that the high entry barriers often found in developing countries give more force to the market foreclosure argument when discussing the essential facility doctrine. He also recommends a more cautious attitude toward vertical restraints.

Although even more desirable in developing countries, U.S.-style competition policy—with its armada of lawyers and economists—is neither affordable nor implementable.⁶ But designing simple and transparent rules for developing countries, particularly to prevent horizontal collusion and abuse of dominant position, remains a

worthy task. Optimizing the use of scarce human and material resources for regulatory purposes is also crucial. Furthermore, a competition agency can play a valuable educational role in advocating the social benefits of fair competition.

Nevertheless, the benefits that can be expected from competition policy in very poor countries will be quite small for the foreseeable future, for several reasons. Complexities and ambiguities remain in the economic analysis of predatory behavior and vertical restraints. Emerging industries will necessarily be highly monopolistic, yet competition agencies lack expertise and information, and interest groups have considerable discretion and potential for interference.

Beyond institutional weaknesses, competition is weak in developing countries because transactions are localized as a result of poor communication systems and inefficient trading organizations. Focusing attention on these areas should be useful, but these problems call even more for investments in infrastructure than for better competition policy.

Limited Technological Knowledge

Less-developed nations must develop, which in the absence of rich physical resource endowments, means absorbing modern technology.

— Frederic Scherer (1997, p. 1)

The most compelling argument in favor of temporary protection is that development requires modern technology, which must be acquired and cultivated, and that learning by doing must occur. Examples abound of successful industrial policies, particularly in East Asia. For such policies to succeed, governments must be able to identify strategically important industries and firms that can act as “national champions” once the learning-by-doing phase has been carried out under appropriate funding and protection. If the country is large enough, this phase can be quite competitive among national firms.

The fact that no systematic positive relationship has been found between firm size and profit, export activity, or research and development (Khemani 1997) will not easily convince developing countries that such strategies are bound to fail.⁷ Nor will the fact that there are probably more examples of failed industrial policies than of successful ones. The question then becomes a tradeoff between, on the one hand, the (in)ability of governments to pick the right industries when economies of scale do not allow internal competition and, on the other hand, a form of external competition that may not set into motion technology transfers and learning by doing.

In developing countries lacking technical, management, and government expertise, it may be that only foreign investment can help development, because it brings new technologies and the credibility needed to borrow on international capital markets. Such investments should be welcomed if they help build local capacity. But when international cartels engage in anticompetitive practices, foreign investment

can undermine economic development in developing countries (see Jenny 1997). For this reason it is clearly desirable to make progress in global competition policy to discipline international cartels.

It is difficult to reach a clear conclusion on the desirability of industrial policy in developing countries. A smart industrial policy, with restrictions on competition that favor learning in a few infant industries, is probably a good step toward development. But such a policy is unlikely to emerge from the captured, corrupt, poorly staffed governments of many developing countries (see Ades and di Tella 1997). Perhaps competition policy would be easier to implement if it were accompanied by expert advice on the restricted industrial policies that could be used to shape a brighter future.

Widespread Corruption

In general, any reform that increases the competitiveness of the economy will reduce incentives for corrupt behavior.

— Susan Rose-Ackerman (1998, p. 46)

Rose-Ackerman's assertion may be overly optimistic. Despite fewer controls, lower entry barriers, and deregulation and privatization, corruption increased slightly in Africa during 1990–95 (Langseth and Bryane 1998). With enough money and political will, corruption can be eradicated. That does not mean that it is worth it to do so, and in fact all societies accept some corruption because eliminating it would be too costly.

If we assume a benevolent government, we must presume that it chooses an optimal policy to deter corruption. We can then ask what effect competition has on the “optimal” level of corruption. We must also presume that the government chooses an optimal competition policy, which together with the underlying technological, informational, and behavioral characteristics of the economy generates competitive pressure. In other words, competition is endogenous and the relevant question becomes, how will changes in the pro-competition characteristics of the economy affect actual competition and corruption? Changes such as the information effect of a more competitive environment, the greater substitutability of competitors' goods, and the lower costs of those competitors can be viewed as exogenous competitive pressures whose impact must be studied.

Alternatively, we can assume that the government maximizes the objectives of certain interest groups. This might entail competition policy that is too lax (compared with a socially optimal one) or tariffs that are too high. Then we can ask, in addition to the former questions, how would a change in these “bad” policies affect competition and corruption? However, the exercises are then normative and beg the question of their implementability. A political economy approach, for economists convinced of some welfare-improving economic measures, is then to propose a

reform path that attracts the interest of those in charge of the government. Institutions that have the means for direct intervention—like the World Bank—can condition their loans or aid on specific institutional reforms to obtain a political consensus.

There is very little literature on the relationship between competition and corruption. Rose-Ackerman (1978) started the analysis of the effect on corruption of competition within the bureaucracy. She concluded that a small number of honest bureaucrats can be very effective by allowing applicants to reapply for the service delivered by bureaucrats if they are asked for bribes. More generally, Rose-Ackerman argues (1998) that the role of competitive pressures in preventing corruption may be an important aspect of a strategy to deter bribery of low-level officials, but it requires a broad-based exploration of the impact of both organizational and market structure on the incentives for corruption facing both bureaucrats and their clients.

Shleifer and Vishny (1993) consider a situation in which government officials have discretion over the provision of some goods—say, permits—and can collect bribes from private agents. But they offer no model explaining why this discretion came about: they take the principal-agent problem as a given. As a result their analysis is preliminary because they cannot take into account how alternative structures of government affect the principal-agent problem and the regulatory response of the top level of government. They treat the officials providing complementary permits as sellers of complementary goods who ignore the externalities they create for one another. This leads to larger bribes but smaller corruption revenues. Even though the market structure approach is part of the problem, by ignoring the foundations of the principal-agent problem, Shleifer and Vishny cannot access the specifics of corruption that are closely tied to the information gaps of the principal and the side contracting between agents. The “market structure” of government officials influences both the allocation of information relating to the agent and the side contracts being written.

Laffont and Martimort (forthcoming) show that the inefficiency of bribe extractions by multiple bureaucrats can be exploited by the principal of the bureaucracy to construct less costly collusion-proof mechanisms. The principal can use traditional yardstick competition between bureaucrats by exploiting the correlation of their information. We also show that information increases the returns from side contracts and makes desirable the separation of bureaucrats, each endowed with his or her own information technology, even when their information signals are uncorrelated.

Ades and di Tella (forthcoming) find empirical evidence that exogenous increases in product market competition reduce corruption in the bureaucracy. They recognize the ambiguous effect of competition in theory. Less competition means that there are more rents to be protected by corruption, but there is also a stronger incentive for a regulatory response. Empirically, their most robust result is that competition, measured by the share of imports in gross domestic product (GDP), significantly decreases corruption. Better schooling, antitrust laws, and higher per capita growth also reduce corruption, but not always significantly.

Bliss and di Tella (1997) ask whether competition kills corruption. Theirs is more a model of gangster activity than of corruption. Corrupt officials demand money from firms, which either pay or quit. Hence corruption affects the number of firms in a free-entry equilibrium. The authors argue that competition is not necessarily exogenous and that corruption affects the extent of competition. They distinguish deep competition parameters—transport costs, uncertainty of cost distribution, overhead costs—which they vary exogenously, from the measure of competition, which is an outcome of the economic system just as the level of corruption is. Their results on the level of corruption and on welfare are quite ambiguous and depend heavily on the structure of uncertainty about costs facing corrupt officials. As the authors recognize, they take as a given the power enjoyed by corrupt agents and focus on the process of demanding bribes. There is no theory of why and how this power came about. This is the main weakness of this interesting paper, resulting from the feature well recognized by the authors that they are not modeling agency relationships.

Collusion can be modeled in the way advocated by Rose-Ackerman—that is, as the problem arising when a principal uses intermediaries as supervisors or monitors in an agency relationship (see Laffont and N'Guessan 1998). Using the Tirole hierarchical model with hard information, consider a principal regulating a firm. There are two types of supervisors, corruptible and less corruptible ones. Depending on the parameter values of the model, it may or may not be better for the principal to allow corruption. It is often not worth paying high salaries to intermediaries to avoid corruption, particularly in developing countries where the cost of funds is very high.

Now consider an increase in yardstick competition. Results again depend on whether low-powered incentive schemes and better supervision are complementary or substitutable. Suppose they are complementary. Increasing competition decreases incentives. In cases where corruption of the most corruptible supervisors is allowed, weaker incentives are called for because more rents are given up. When weak incentives and competition are complementary, competition is more effective in the case with corruption and therefore increases the probability that permitting corruption is good policy. The opposite results obtain in the case of substitutability.

A similar logic applies to the case of greater product market competition imposed on a monopolist. Again, when corruption is accepted, weaker incentives are desirable because the principal has a greater likelihood of facing an efficient type and wants to lower rents even more. Lower costs of competitors bring higher production of competitors. If the goods of the monopolist and competitors are demand substitutes and strategic substitutes, this implies less production for the monopolist. But less competition makes effort less effective and calls for weaker incentives.

Corruption may be particularly harmful when it is asymmetric. When a bureaucrat delivers a scarce resource, widespread corruption can lead to the allocation of goods to those who value them most. Efficiency is preserved at the cost of lost resources for the government. If the wages of the bureaucrats are reduced appropriately, the damage may be limited (Besley and McLaren 1993). Redistributive

programs are difficult to implement. But if only some agents are willing to pay bribes, corruption entails a favoritism that can be very costly because it induces an inefficient allocation of resources. An inefficient contractor may be selected simply because he or she is willing to pay bribes. Laffont and Tirole (1993, ch. 11) show how to fight corruption in this situation. It requires eliminating discretion in procurement rules—for example, by imposing the choice of the lowest offer without any consideration of the nonverifiable dimensions about which the bureaucrat might have information, such as the quality of the project or the financial viability of the applicant—or paying high rewards to the bureaucrat to induce proper behavior.

Another case occurs when one type of firm engages in corruption to evade taxation while others do not. Competition is then biased.⁸ Corruption here creates unfair and inefficient competition, starting a vicious circle through the higher cost of public funds it entails, because less money is then available to pay for incentives against corruption within the administration.

Controlling corruption is crucial for growth: 40 percent of the variance in growth rates in Africa between 1990 and 1996 can be explained by corruption (Laffont and Meleu 1998). The effects of competition on corruption are somewhat ambiguous from a static theoretical viewpoint, but to the extent that competition brings growth, it provides the instruments needed to fight corruption. However, corrupt governments may want to weaken competition (see the section below on governments with private agendas).

Inefficient Credit Markets

Although financial and banking crises have been costly to the economies of industrial countries ... the damage that these crises impose on developing countries seems to be far greater.

— Frederic S. Mishkin (1997, p. 29)

Following Mishkin, I note that it is widely understood that a crucial impediment to an efficient financial system is asymmetric information arising from adverse selection or moral hazard. Banks are particularly important in developing countries, where information about private firms is harder to collect than in industrial countries and, consequently, securities markets play a small role in financing the economy. But depositors who lack information about the quality of the loans made by banks need protection to avoid bank panics when shocks occur. A government safety net in the form of deposit insurance or less formal government protection weakens incentives to monitor banks and exacerbates excessive risk taking by bank managers. Because most depositors are unable to monitor banks, government regulation is needed to promote risk diversification, impose capital requirements tied to the amount of risk, control management to prevent fraud, and ensure broad disclosure

of information. Such regulation is obviously complex and may be hampered by insufficient resources and knowledge or by regulatory capture.

A financial crisis is a nonlinear disruption to financial markets in which adverse selection and moral hazard become much worse, so that financial markets are unable to efficiently channel funds to those who have the most productive investment opportunities (see Mishkin 1997). In addition to the shocks that promote financial crises in industrial countries (increases in interest rates, increases in uncertainty, stock market crashes, and so on), developing countries face such shocks as unanticipated depreciation of the local currency (because debt is often denominated in foreign currency) or terms of trade shocks in primary goods (for countries with little export diversification). A prompt response by regulators involving monetary expansion or greater availability of loans may prevent these shocks from degenerating into financial crises. However, Mishkin (1997) argues that because debt contracts are generally of very short duration (due to a bad record on inflation or poor enforcement of creditor rights) and because of a lack of credible commitment to low inflation, these brief expansionary policies only fuel inflation and currency depreciation. Speedy recoveries are not possible without foreign assistance.

Thus it is important for developing countries to have strong bank regulatory and supervisory systems. Achieving this goal requires providing sufficient resources to bank supervisors, implementing adequate accounting and disclosure requirements, encouraging bank supervisors to take prompt corrective action, and insulating bank supervision from the political process (Mishkin 1997). Moreover, “although [deregulation and liberalization] are desirable objectives ... [they] can be disastrous if not managed properly. If the proper bank regulatory and supervisory structure is not in place before liberalization, risk-taking behavior will not be adequately constrained” (Mishkin 1997, p. 57).

Mishkin reaches the same conclusion as I did above when considering the liberalization of natural monopolies: it is important to set up an appropriate regulatory framework before liberalizing (Laffont 1996). But this task is difficult, if not impossible, in developing countries that lack adequate human resources, auditing technologies, and political institutions. Thus “developing countries may need to move slowly in financial liberalization in order to keep a lending boom from getting out of hand” (Mishkin 1997, p. 57). I will reiterate that the most effective aid is aid that helps to design those regulatory institutions and that it is irresponsible to advocate liberalization without providing such aid.

The complementarity of general competition policy and good banking sector regulation is worth noting. When the banking sector is inefficient and makes borrowing costly or impossible, an effective competition policy may destroy the rents that allow firms to invest. It also increases the probability of bankruptcy and calls for efficient bankruptcy rules.

In addition to the concern with financial crises, the question of financial innovations adapted to developing countries is crucial. Some hope lies in developing collective borrowing such as that offered by Bangladesh’s Grameen Bank. So far

competition among traditional banking institutions has done little to increase lending to the poor.

Lack of Commitment Power

Low credibility of rules is associated with lower rates of investment and growth.

— Aymo Brunetti, Gregory Kisunko, and Beatrice Weder (1997, p. 3)

The lack of credibility associated with weak governments is a major problem for developing countries, one that makes the emergence of a competitive environment difficult and affects all areas of economic activity. Agents cannot write long-term contracts because the court system cannot be trusted to protect them. And agents do not find it worth investing because they fear that their gains will be stolen or expropriated, either in their market activities or in their contractual relationships with the government. This lack of protection for agents who are no longer residual claimants for the results of their efforts induces the Williamson underinvestment effect, particularly in noncontractible investments.

Consider a framework of repeated auctions for the management of a public utility. The winner of today's auction will restrict her nonverifiable investment because she knows with some probability that she will not win the next auction and will lose her investment, which cannot be compensated for because it is nonverifiable. The tougher the competition is tomorrow, the greater the underinvestment.⁹ When the government is strong, it can commit to bias the auction in favor of the incumbent to mitigate the underinvestment effect (see Laffont and Tirole 1993, ch. 8). But a weak government will not be able to commit credibly to a biased auction. Weakening competition today may be the only way to credibly restrict competition tomorrow.

Let us simplify this framework to a single firm regulated by an agency. The lack of regulator commitment also introduces the important ratchet effect (see Laffont and Tirole 1993, chs. 9 and 10). Specifically, when the regulated firm's information is correlated over time and the agency can commit only in the short run, the firm underperforms in the first period to hide its efficiency, because the regulator would make use of the information learned in the first period to eliminate any prospect of rent in the future. Olsen and Torsvik (1993), with linear contracts, and Martimort (1998), with nonlinear contracts, show that having multiple regulators—a type of competition—can be a way to commit to less rent extraction in the future and therefore to more revelation of information in the first period. The optimal design of institutions in developing countries calls for less competition in one case and more competition in the other to partly overcome the problem created by the lack of commitment.

But the lack of commitment power may itself make it difficult to implement competition. This is the case if the threats of competition policy are not credible because

judges can be corrupted or because the expertise to prove noncompetitive behavior is lacking.

It is difficult to attract foreign competition and foreign capital if the government does not have the credibility required to protect investments. After all, this is why many developing countries established public firms, because they could not attract private investment. Institutions such as the World Bank and the International Monetary Fund have enhanced the credibility of governments in protecting foreign investment (if only by the fear of large penalties). In this way they have contributed to increased competition.

Another example shows how the lack of government credibility undermines the institutions required for competition. Consider an auction of contracts for public projects. If competition is effective, the auction fulfills its objective of lowering the cost of the projects. But collusion is a potential problem, and I have already argued that the transactions costs of collusion are lower in developing countries. One way for government to fight collusion is to use a low reservation cost. But this instrument is effective only if it is credible—that is, only if government can credibly cancel the project if no offer is below the reservation cost.

The lack of commitment power may call for more or less competition. However, implementing competition requires commitment power and so is more difficult in developing countries.

Governments with Private Agendas

Governments of many developing countries are either captured by powerful interest groups—whether local elites or foreign investors—or better described as predatory than protective. In recognition of these shortcomings, a consensus is emerging that little can be done to reform institutions without local political support. This is particularly clear for competition policy. It is easy for authorities to extract bribes and favors when they control barriers to entry. An oligopolistic structure can be maintained under the pretext of regulation, generating rents that can be shared by those in power.¹⁰ Implementing effective competition policies is the best way to destroy the source of such rents. Still, no illusions should be entertained when competition laws proliferate in developing countries if strong political commitment is lacking (see Khemani and Dutz 1995). It is too easy to prevent effective competition (using non-tariff barriers) for such laws to be taken as credible commitments. Poorly designed and applied competition laws can even discourage trade and foreign investment.

It is difficult to see how institutions like the World Bank can promote competition in the face of political resistance without allowing (currently nonexistent) international competition authorities to intervene directly within the countries. There are too many nonverifiable instruments in the hands of authorities to monitor policies. To be effective, only penalties based on performance seem possible: provide funds only to countries that have achieved good results on the competition front.

The implementation of good competition policies requires compensating the decisionmakers who would benefit from rents generated by the lack of competition.

For example, it may make sense to dispense favors associated with investment projects that take place only if objective measures of competitiveness have improved, at least if corruption is centralized enough. Of course, it is difficult to openly advocate such policies.

What is happening with privatization bears some similarity to such a deal. For the rulers of a country and their political supporters, privatization often means the loss of numerous private benefits that are socially costly. For privatization to take place, alternative benefits are needed in the form of shares at discounted prices, direct kickbacks, or World Bank loans that provide indirect benefits. A satisfactory *quid pro quo* can be engineered because the privatization decision is easily verifiable (see Laffont and Meleu 1997). It seems more difficult to enforce competition in similar ways.¹¹

The difficulties encountered in implementing competition policy have implications for the optimal sequencing of reforms. Privatization and formal liberalization are likely to lead to private monopolies, which will generate resources for interest groups apt to resist further development of authentic competition. Efforts to impose these reforms before a credible set of institutions—for regulation, competition policy, financial regulation—has been designed will yield disappointing results. It seems easier to use privatization as an opportunity to compensate decisionmakers and other powerful political interest groups (for example, with shares) for setting up competitive conditions and good regulatory institutions. Taking this step, however, requires human and financial resources—hence the need to emphasize education and international aid in this area.

Conclusion

Stimulating competition is essential for development. But effective competition requires democratic institutions that are often complex and costly to set up in countries lacking financial and human resources, auditing expertise, and a reputation for commitment. Competition is not easily implemented, and captured governments may not wish to implement it. Because competition is not an automatic outcome of deregulation, simply conditioning loans on the existence of competition laws will not ensure the creation of proper institutions for effective competition. Only a strong state can implement competition.

Thus little progress on competition can be expected in countries where political will is lacking. In countries which have the political will, aid is essential to build the institutions needed for development. A strong state and competition are complementary in many vital economic sectors. Aid and conditionality should focus on fundamentals favoring both.

Notes

1. See Willig (1987), Martin (1993), Horn, Lang, and Lundgren (1994), Hart (1983), Scharfstein (1988), Hermalin (1992, 1994), and Schmidt (1997). A weakness of this literature is that it takes as exogenous the number of firms representing the level of competition.

2. In Argentina the tariff structure for water and sewerage allows the licensee to balance its economic returns through certain groups of users (see Chisari and Estache 1997). For telecommunications Argentina has been divided into two regions of roughly the same size, each with an urban area and a rural area. Each region has been allocated to a different operator.
3. This model was originally developed by Gabel and Kennet (1994).
4. See also Mookherjee (1998). Note that inefficient tax systems due to lack of resources for auditing and monitoring may be a deliberate choice by governments. In China, for example, the inefficiencies due to the decentralization of the tax system were accepted out of fear that recentralizing public resources would increase the power of those opposing reforms. Providing international aid to increase the efficiency of controls would then be pointless because such reforms would not be implemented.
5. If we model increased competition as an increase in competition for talent, we also obtain an increase in incentives, the more so for developing countries.
6. Not surprisingly, different schools of economic thought have strongly conflicting views on the relevance and content of competition policy in developing countries; see Khemani and Dutz (1995).
7. It is worth remembering that before World War II most countries treated monopolies and cartels, domestic or foreign, with benign neglect or, as in Germany and Japan, actively encouraged them (Scherer 1996).
8. From this perspective the large movement to eradicate bribing in U.S. and European companies active in developing countries may create inefficiencies in the short run.
9. Aghion, Dewatripont, and Rey (1995, 1997) develop the interesting idea that the main merit of competition may be to induce firms to adopt behavior that is closer to profit maximization. Profit maximization unambiguously induces those firms to innovate more, which here means to invest more.
10. Rowat (1995, p. 21) notes that “the political economy of reform needs to be examined carefully since vested interests that enjoyed the highly protectionist environments of the past usually are averse to succumbing to a new competitive regime ... Competition authorities need to make a special effort to ‘educate’ the public about the merits of a sensible competition policy.”
11. Note, however, that privatized firms often make enormous profits, which can be interpreted as the result of efficiency gains or of lax regulation.

References

- Ades, Alberto, and Rafael di Tella. 1997. “National Champions and Corruption: Some Unpleasant Interventionist Arithmetic.” *Economic Journal* 107 (July): 1,023–42.
- . Forthcoming. “Rents, Competition, and Corruption.” *American Economic Review*.
- Aghion, P., and P. Howitt. 1997. “A Schumpeterian Perspective on Growth and Competition.” In D. Kreps and K. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications*. Cambridge: Cambridge University Press.
- Aghion, P., M. Dewatripont, and P. Rey. 1995. “Competition, Financial Discipline and Growth.” Institut d’Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- . 1997. “Agency Costs, Firm Behavior and the Nature of Competition.” Institut d’Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- Ahmad, Ehtisham, and Nicholas Stern. 1991. *The Theory and Practice of Tax Reform in Developing Countries*. Cambridge, Mass.: Cambridge University Press.
- Anton, J., and D. Yao. 1987. “Second Sourcing and the Experience Curve: Price Competition in Defense Procurement.” *RAND Journal of Economics* 18 (1): 57–76.

- . 1989. "Split Awards, Procurement and Innovation." *RAND Journal of Economics* 20 (4): 538–52.
- . 1992. "Coordination in Split Award Auctions." *Quarterly Journal of Economics* 107 (2): 681–707.
- Auriol, E., and Jean-Jacques Laffont. 1992. "Regulation by Duopoly." *Journal of Economics and Management Strategy* 1 (3): 507–33.
- Banerjee, A. 1996. "Can Anything Be Done about Corruption?" In M. G. Qubria and J. M. Dowling, eds., *Current Issues in Economic Development: An Asian Perspective*. Oxford: Oxford University Press.
- Bardhan, Pranab. 1997. "Corruption and Development: A Review of Issues." *Journal of Economic Literature* 35 (3): 1320–46.
- Bernheim, B. D., and H. D. Winston. 1992. "Exclusive Dealing." Institute of Economic Research Working Paper 1622. Harvard University, Cambridge, Mass.
- Besley, Timothy, and John McLaren. 1993. "Taxes and Bribery: The Role of Wage Incentives." *Economic Journal* 103 (January): 119–41.
- Bliss, C., and Rafael di Tella. 1997. "Does Competition Kill Corruption?" *Journal of Political Economy* 105 (October): 1001–23.
- Boyer, M., and Jean-Jacques Laffont. 1998. "Competition and Incentives in the Public Sector." Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- Brunetti, Aymo, Gregory Kisunko, and Beatrice Weder. 1997. "Credibility of Rules and Economic Growth: Evidence from a Worldwide Survey of the Private Sector." Policy Research Working Paper 1760. World Bank, Office of the Chief Economist and Senior Vice President, Development Economics, Washington, D.C.
- Buchanan, J. M. 1969. "External Diseconomies, Corrective Taxes, and Market Structure." *American Economic Review* 59: 174–76.
- Caminal, R., and C. Matutes. 1997a. "Bank Solvency, Market Structure and Monitoring Incentives." Universitat Autònoma, Barcelona.
- . 1997b. "Can Competition in the Credit Market Be Excessive?" Centre for Economic Policy Research Paper 1725. London.
- Caprio, Gerard Jr., and Daniela Klingebiel. 1997. "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1996*. Washington, D.C.: World Bank.
- Chisari, O., and Antonio Estache. 1997. "The Needs of the Poor in Infrastructure Privatizations: The Role of Universal Service Obligation. The Case of Argentina." Centre de Estudios Economicos de la Regulacion, Buenos Aires.
- Compte, O., F. Jenny, and P. Rey. 1997. "Capacity Constraints, Mergers and Collusion." Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- Dana, J., and K. Spier. 1994. "Designing a Private Industry: Government Auctions with Endogenous Market Structure." *Journal of Public Economics* 53 (1): 127–47.
- Faure-Grimaud, A., Jean-Jacques Laffont, and D. Martimort. 1998. "A Theory of Supervision with Endogenous Transaction Costs." Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- Gabel, D., and M. Kennet. 1994. "Economies of Scope in the Local Exchange Market." *Journal of Regulatory Economics* 6 (4): 381–98.
- Gasmi, F., Jean-Jacques Laffont, and W. Sharkey. 1997. "Incentive Regulation and the Cost Structure of the Local Telephone Exchange Network." *Journal of Regulatory Economics* 12 (1): 5–25.
- . 1998. "Competition and Universal Service Obligation." Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France.

- Gould, D., and J. Amaro-Reyes. 1983. "The Effects of Corruption on Administrative Performance." World Bank, Washington D.C.
- Hart, O. 1983. "The Market Mechanism As an Incentive Scheme." *Bell Journal of Economics* 14 (2): 366–82.
- Hausman, J. 1997. "Taxation by Telecommunications Regulation." Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass.
- Hermalin, B. 1992. "Effects of Competition on Executive Behavior." *RAND Journal of Economics* 23 (3): 350–62.
- . 1994. "Heterogeneity in Organizational Form: Why Otherwise Identical Firms Choose Different Incentives for their Managers." *RAND Journal of Economics* 25 (4): 518–37.
- Horn, H., H. Lang, and S. Lundgren. 1994. "Competition, Long-Run Contracts and Internal Inefficiencies in Firms." *European Economic Review* 38 (2): 213–33.
- Jenny, F. 1997. "Competition Policy in a Global Economy." Ecole Supérieure de Sciences Economiques et de Gestion, Paris.
- Jones, L., P. Tandon, and I. Vogelsang. 1990. *Selling Public Enterprises: A Cost Benefit Methodology*. Cambridge, Mass.: MIT Press.
- Kerf, M., and W. Smith. 1996. *Privatizing Africa's Infrastructure: Promise and Challenge*. World Bank Technical Paper 337. Washington, D.C.
- Khemani, S. 1997. "Competition Policy and Economic Development." *Policy Options* (October): 23–27.
- Khemani, S., and Mark Dutz. 1995. "The Instruments of Competition Policy and their Relevance for Economic Development." In Claudio R. Frischtak, ed., *Regulatory Policies and Reform: A Comparative Perspective*. Washington, D.C.: World Bank.
- Laffont, Jean-Jacques. 1996. "Regulation, Privatization and Incentives in Developing Countries." In M. G. Quibira and J. Malcolm Dowling, eds., *Current Issues in Economic Development*. Oxford: Oxford University Press.
- Laffont, Jean-Jacques, and D. Martimort. Forthcoming. "Separation of Regulators against Collusive Behavior." *RAND Journal of Economics*.
- Laffont, Jean-Jacques, and M. Meleu. 1997. "A Positive Theory of Privatization for Sub-Saharan Africa." *Journal of African Economies*.
- . 1998. "Corruption and Privatization: An Empirical Evaluation for Sub-Saharan Africa." Centre de Recherches en Microéconomie du Développement, Abidjan.
- Laffont, Jean-Jacques, and A. N'Gbo. 1998. "Network Expansion in Developing Countries." Centre de Recherches en Microéconomie du Développement, Abidjan.
- Laffont, Jean-Jacques, and T. T. N'Guessan. 1998. "Competition and Corruption in an Agency Relationship." Centre de Recherches en Microéconomie du Développement, Abidjan.
- Laffont, Jean-Jacques, and J. Tirole. 1986. "Using Cost Observation to Regulate Firms." *Journal of Political Economy* 94 (3): 614–41.
- . 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, Mass.: MIT Press.
- Langseth, P., and M. Bryane. 1998. "Ministers and Parliamentarians Assess Corruption." *Governance* 1 (January): 53–60.
- Martimort, D. 1998. "Renegotiation Design with Multiple Regulators." Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- Martin, S. 1993. "Endogenous Firm Efficiency in a Cournot Principal-Agent Model." *Journal of Economic Theory* 59 (2): 445–50.

- Mishkin, Frederick S. 1997. "Understanding Financial Crises: A Developing Country Perspective." In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1996*. Washington, D.C.: World Bank.
- Mookherjee, Dilip. 1998. "Incentives Reforms in Developing Country Bureaucracies: Lessons from Tax Administration." In Boris Pleskovic and Joseph E. Stiglitz, eds., *Annual World Bank Conference on Development Economics 1997*. Washington, D.C.: World Bank.
- Olsen, T., and G. Torsvik. 1993. "The Ratchet Effect in Common Agency: Implications for Regulation and Privatization." *Journal of Law, Economics and Organization* 9 (1): 136–58.
- Rey, P. 1997. "Competition Policy and Development." Institut d'Economie Industrielle, Université des Sciences Sociales, Toulouse, France.
- Rose-Ackerman, Susan. 1978. *Corruption: A Study in Political Economy*. New York: Academic Press.
- . 1996. "Redesigning the State to Fight Corruption—Transparency, Competition, and Privatization." *Public Policy for the Private Sector* (World Bank, Washington, D.C.) (March): 49–52.
- . 1998. "Corruption and Development." In Boris Pleskovic and Joseph E. Stiglitz, eds., *Annual World Bank Conference on Development Economics 1997*. Washington, D.C.: World Bank.
- Rowat, Malcolm 1995. "Competition Policy in Latin America: Legal and Institutional Issues." Paper presented at the Seminar on Good Government and Law, sponsored by the British Council, March 27–28, London.
- Scharfstein, D. 1988. "Product-Market Competition and Managerial Slack." *RAND Journal of Economics* 19 (1): 147–55.
- Scherer, Frederic. 1996. "International Competition Policy and Economic Development." D P. 96–26. Zentrum für Europäische Wirtschaftsforschung, Mannheim.
- . 1997. "Competition Policy and Economic Development." Outline of World Trade Organization presentation, November 29, Harvard University, Department of Economics, Cambridge, Mass.
- Schmidt, K. 1997. "Managerial Incentives and Product Market Competition." *Review of Economic Studies* 64 (2): 191–213.
- Shleifer, Andrei, and Robert Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108 (3): 599–617.
- Stiglitz, Joseph E. 1997. "The Role of Government in Economic Development." In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1996*. Washington, D.C.: World Bank.
- Tirole, J. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, and Organization* 2 (2): 181–214.
- Trebilcock, Michael J. 1996. "What Makes Poor Countries Poor? The Role of Institutional Capital in Economic Development." University of Toronto, Law and Economics Program.
- Wesberry, J. 1996. "Good Financial Management versus Corruption." World Bank, Latin America and Caribbean Region, Washington, D.C.
- Willig, R. 1987. "Corporate Governance and Market Structure." In A. Razin and E. Sadka, eds., *Economic Policy in Trade and Practice*. London: Macmillan.

Comment on “Competition, Information, and Development,” by Jean-Jacques Laffont

Ana Julia Jatar

Jean-Jacques Laffont’s article offers an original framework for rethinking competition policy in developing countries. He accepts that, in the first-best world of economists, competition is a good thing. But he expresses increasing doubt that competition delivers benefits when those assumptions are relaxed—as they invariably must be in an imperfect world.

Laffont also analyzes market inefficiencies in developing countries, considering how structural problems affect the potential benefits of competition. As the former head of Venezuela’s competition agency, I believe that his conclusions have important implications for policymakers everywhere—but especially in liberalizing economies. Here I focus on some of the problems Laffont considers.

Inefficient Tax Systems and the High Cost of Public Funds

According to Laffont, distortions in the tax system make the cost of public funds many times higher in developing countries than in industrial countries. Thus policymakers should be aware of such costs when considering whether to use public funds to promote more or less competitive markets. Laffont argues that in developing countries monopolistic structures financed with cross-subsidies tend to be more efficient than more competitive structures financed with taxes. In other words, he argues that cross-subsidies are not naturally good or bad. Rather, they must be evaluated relative to alternative sources of financing.

Laffont concludes that cross-subsidies should not be dismissed as a source of financing—though it is common knowledge among economists that cross-subsidies introduce important distortions—because distortions in the tax system of developing countries may impose even higher costs. This is a valid argument, one that is particularly useful during privatization, especially for utilities, transportation companies, and the like. But a note of caution should be added from the competition policy perspective. It has to do with time.

Ana Julia Jatar is senior fellow at the Inter-American Dialogue.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

Service providers may become increasingly inefficient when a monopoly has been granted. In fact, prices could theoretically rise to the point where they equal the cost of the tax distortions they were designed to avoid. In Venezuela I participated in discussions between the recently privatized telecommunications company, CANTV, and its regulator. CANTV was granted a nine-year monopoly on international calls with the mandate of covering—using the profits from international services—the sizable shortfall in local telephone lines existing at the time of privatization. Imposing a time limit on monopolistic control of international calls provided the right incentive for the company, preparing it for foreign competition and guaranteeing consumers better prices in the long run.

Traditional Competition Policy

Laffont argues that competition policy is too complex and costly to be implemented in developing countries. I agree that it is important to design simple and transparent rules for developing countries, particularly to prevent horizontal collusion and abuse of market power by dominant firms. I also agree that copying complicated mechanisms—such as merger analysis—can be highly inefficient. Still, experience shows that competition rules are not only affordable but also extremely useful in recently liberalized economies. Competition policy can reinforce the institutional base in economies with highly concentrated and imperfect markets.

Recent developments should ease Laffont's concerns. Experiences in various countries indicate that competition policy has been useful. Still, some lessons should be taken into account to avoid inefficiency. According to specialists and practitioners from Africa, Eastern Europe, and Latin America, competition rules in recently liberalized economies must be simple and straightforward (Jatar and Tineo 1998). They should prohibit horizontal agreements among competitors, including price cartels. And they should establish clear differences between agreements among competitors that are designed to restrict competition and vertical restraints that are designed to increase efficiency (Rodríguez and Williams 1996).

In Latin America the enforcement of competition since liberalization has been instrumental in lowering entry barriers to foreign competition, particularly where dominant local firms have used vertical mechanisms to control distribution channels. In addition, many competition agencies have found it useful to give high priority to nontradable sectors and quasi-nontradable products such as perishables and those with low price-volume ratios (paints, beverages), since potential foreign competition will have little or no impact on local firm behavior.

Experience also shows that competition agencies must be empowered to perform their advocacy and deregulatory roles (Jatar 1993). A competition agency should have the legal authority to challenge other government agencies' decisions that conflict with competitive principles. In addition, a competition agency must enjoy complete independence from the rest of the administration (to ease political and interest group pressures) and must be able to appeal the decisions of administrative judges to the supreme court (to preserve a system of checks and balances).

Corruption

Laffont also considers the relationship between competition and corruption, concluding that in the short run competition may increase corruption in the public sector because it raises the stakes for collusion. This finding may hold for the specific case evaluated in the article. But I strongly believe that in some cases competition may decrease corruption. For example, the introduction of competitive mechanisms (such as multiple sourcing) has helped to control corruption in health and education services. When hospitals and schools are allowed to buy supplies from different sources that are publicly disclosed, the process is less susceptible to corruption than when a centralized agency (such as the ministry of health or education) buys directly from one supplier.

Two key elements for fighting corruption, particularly in judicial systems, are disclosure of information and an efficient system of checks and balances. In Venezuela the absence of these elements can cause high corruption in different courts.

Credit Markets

Laffont also evaluates the impact of competition on inefficient credit markets. He concludes that increased competition in the financial sector is desirable to diminish the negative impacts of asymmetric information. Still, Laffont argues—and I agree—that liberalization must be carried out slowly and should be complemented by a strong regulatory body.

I would like to introduce a new competition element related to the globalization of financial markets—the credibility of bank regulators. Increasing numbers of multinational banks are competing everywhere for the same deals. These banks are subject to home-country consolidated supervision. In other words, U.S. banks are supervised in a consolidated manner in the United States, Spanish banks in Spain, and Venezuelan banks in Venezuela. When these banks meet and compete for projects, the credibility of their regulators becomes increasingly crucial. The better is the regulator, the better are a bank's chances of securing a deal. Thus the internationalization of banking institutions is affecting competition among regulators and changing the political economy of regulation (Bavin and Hausmann 1997). In today's world, banks need a credible regulator.

Credibility and Governments

Finally, Laffont addresses the anticompetitive implications of the lack of credibility associated with governments in developing countries. The argument is simple but forceful: lack of credibility results in low investment and slow growth. With low investment, markets will be structurally monopolistic. Furthermore, according to Laffont, the lack of credibility affects all areas of economic activity. Agents cannot write long-term contracts because the court system is inefficient, and agents do not find it worth investing because they fear that their returns will be expropriated.

To decrease the incentives for expropriation by any government, ownership of a particular asset should be widely dispersed. With the dispersal of ownership, governments will have fewer political incentives to expropriate assets or rents. This issue becomes relevant for privatization when institutions such as pension funds provide more economically efficient and politically acceptable alternatives to privatized industries.

References

- Gavin, Michael, and Ricardo Hausmann. 1997. "Make or Buy Approaches to Financial Market Integration." In Liliana Rojas-Suarez, ed. *Safe and Sound Financial Systems: What Works for Latin America*. Washington, D.C.: Inter-American Development Bank.
- Jatar, Ana Julia. 1993. "Implementing Competition Policy in Recently Liberalized Economies: The Case of Venezuela." Fordham Institute of Law, New York.
- Jatar, Ana Julia, and Luis Tineo. 1998. "Five Years of Competition Policy in Peru: Challenges in the Transition to a Market Economy." National Institute for the Defense of Competition and Protection of Intellectual Property, Lima, Peru.
- Rodriguez, A.E., and M. Williams. 1996. "Limits to Antitrust Policy for Reforming Economies." *Houston Journal of International Law* 18(2).

Comment on "Competition, Information, and Development," by Jean-Jacques Laffont

Richard Schmalensee

Jean-Jacques Laffont serves as a capable guide in an interesting region of the second-best jungle, artfully employing the compass of sound judgment and the machete of economic theory—or the reverse, if you prefer. He does not, however, reveal easy routes to clear policy prescriptions. Rather, he persuades me that the terrain is difficult, no matter how much underbrush is cleared away. Laffont's item-by-item analysis of the implications of increased competition for various structural problems in developing countries is complex enough. The interaction effects that arise when several of these problems coexist add another layer of complexity.

I will not extend Laffont's analysis in that direction, at least not in any systematic way. Similarly, I will not try to expose any deep flaws in his analysis, because I do not think that there are any. Instead, I will take the path of least resistance and offer a few supplementary observations suggested by his provocative analysis.

Transparency and Architectural Inertia

Laffont's first example is the problem of subsidizing rural telephone service. This analysis assumes that universal service is so desirable that it is not optimal to follow the otherwise efficient policy of setting higher prices in rural areas because costs are higher. Formally it is clear that if a country's tax system is sufficiently inefficient, it is better in theory to finance universal service with a cross-subsidy from urban to rural customers than with tax revenues. It is important to recognize, however, that once this choice has been made, it may be very difficult to reverse when the tax system improves.

The history of U.S. telecommunications illustrates this well. Under regulation in place before World War I, the Bell telephone companies in each state were generally required to charge uniform rates and to provide service to anyone in their territories who requested it. These requirements built in a cross-subsidy from urban to rural customers—a cross-subsidy later supplemented by a cross-subsidy from long-distance to local service.

Richard Schmalensee is professor and dean of the Sloan School of Management at the Massachusetts Institute of Technology.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

The U.S. tax system may have been inefficient before World War I, but it was pretty good by the late 1960s, when pressures to introduce competition in telecommunications began to mount. And because telephone costs had fallen and incomes had risen, the subsidies required to make telephone service affordable to most rural residents had shrunk substantially. Still, despite 30 years of movement toward more competition in telecommunications and the clear undesirability (from a Ramsey perspective) of using long-distance to subsidize local service, the old structure of cross-subsidies has proven quite durable. An important source of its durability is the general lack of transparency in cross-subsidization: able and disinterested economists long had a hard time agreeing on even the order of magnitude of the cross-subsidy in U.S. telephony.

This is not to say that the United States was wrong to rely on cross-subsidies in this instance. Nor do I disagree with Laffont's conclusion that this approach should not be demonized. The point is that the basic structures of policy responses to particular problems, policy architectures if you will, tend to endure even as implementation details change. Once it is decided that cross-subsidies are to be relied on in telecommunications, for instance, it may take many decades and a great deal of political capital to implement an alternative architecture—one based on tax revenues, for instance. Once a policy architecture has been chosen, various investments, including investments in careers, are made that depend for their profitability on the continuation of that basic regime. And when a regime that lacks transparency has been put into place, removing it is particularly difficult.

Industrial policy raises these same difficulties, along with others. Laffont argues that a smart industrial policy can aid development, and by definition this is surely true. As a skeptic, however, I doubt that it is true in any nontautological sense. Over the years we have seen plenty of dumb industrial policies even in industrial countries. Moreover, many of the structural problems that Laffont considers—in particular, scarcity of expertise, pervasiveness of corruption, and absence of government commitment—make it especially unlikely that developing countries will devise and implement policies that will later prove to have been “smart.” Smart or dumb, though, industrial subsidies often lack transparency and tend to persist. Infant industries are notorious for their inability to grow up.

These sorts of architectural inertia should be taken into account when deciding how to respond to structural problems. In particular, if a second-best policy architecture is put in place because of a specific structural problem, such as an inefficient tax system, one should be fairly confident that the solution will not long outlive the problem that gave rise to it. The relevant cost-benefit analysis should have at least an implicit dynamic dimension; short-lived problems should not be treated with policy architectures that will become long-lived burdens.

Information, Expertise, and Discretion

In his discussion of the implications of lack of expertise in monitoring and auditing, Laffont considers fully rational, welfare-maximizing regulators operating with noisy

data. He argues that in developing countries the large information rents forgone by high-powered incentive schemes are especially costly given the high cost of public funds. I generally agree with his conclusion that competition can serve as a valuable source of information under these conditions and that it can make higher-powered regulatory incentives optimal. But I would reach that conclusion by a slightly different path.

Where monitoring and auditing expertise is scarce, one cannot safely assume that regulators will have the expertise and staff support needed to make fully rational decisions, and the assumption that they are interested only in social welfare is rarely a safe one. Under these conditions it is important to consider limiting regulators' discretion by putting in place simple but robust policy regimes that reduce opportunities for massive errors and massive corruption, generally at some cost in theoretically attainable efficiency. This approach tends to favor, for instance, relatively simple price-cap regimes, perhaps with negotiated initial prices, over more complex rate-of-return or related systems. The contractual solution that was adopted in Peru and on which Paul L. Joskow reports (see elsewhere in this volume) also seems consistent with this approach.

What about the fact that price-cap regimes tend to yield substantial information rents? When such rents are paid out of tax receipts, as in government procurement, and the tax system is inefficient, this is a potentially important point. In the case of regulated monopolies, however, rents generally arise because prices are allowed to exceed costs. (In practice as well as in theory, it is generally recognized that when monitoring and auditing are difficult, it is unwise to allow regulators to make large payments to the firms they regulate.) Under regulation where tax receipts are not used to subsidize, the inefficiency of the tax system does not translate directly into higher costs of information rents.

Moreover, when regulators are badly informed or corrupt, the rents received by regulated firms may not be terribly sensitive to regime choice, while incentives to invest and to reduce costs may be. A clever firm can extract rents from a poorly informed regulator even in a cost-plus regime. In the extreme case, there are worse things than unregulated, profit-seeking monopolies—and regulators have historically been the ones who create those worse things.

Competition Policy

I agree with Laffont's observation that developing countries cannot implement anything like U.S. antitrust policy and should not try. Some parts of U.S. policy arguably reduce competition by protecting inefficient competitors, and other parts have administrative and other costs that outweigh any pro-competition benefits, even in a country where legal and economic expertise are comparatively abundant. And while the United States can afford broad private enforcement of its antitrust laws, it is hard to imagine such a regime being beneficial in developing economies.

But I disagree with Laffont's contention that the expected benefits from competition policy in developing countries are likely to be quite small for some time. In

1890, when the United States embarked on the antitrust experiment, it was no wealthier than many of today's developing countries. There were few telephones, no automobiles or televisions, and not a single well-trained industrial organization economist—largely because the field of industrial organization did not yet exist. The first constructive principle that was established in this field was that price fixing and other forms of cartel behavior were illegal. The second principle was that mergers that created monopolies would not be allowed. The problem of abuse of a dominant position was also considered early on, but no useful simple rules were devised then, and none has emerged since. Few argue that this early antitrust policy served the United States badly.

I think this lesson generalizes. I contend that a competition policy focused on blatant cartel behavior and mergers to monopoly would be relatively cheap and could have substantial benefits in developing countries. These are analytically the simplest cases, and they are the ones most likely to benefit consumers. There is also a significant potential benefit from institutionalizing the role of advocating competition. The Antitrust Division of the U.S. Department of Justice has often been an important advocate for competition and consumer welfare.

Corruption, Commitment, and Capture

Corruption is a slippery subject to analyze. It is not clear whether equity or efficiency effects are the fundamental concern, nor is it clear to whom a normative analysis should be addressed. Issues of causality seem difficult. There is not much to be gained by telling corrupt officials that increasing competition will reduce their incomes. The strongly negative correlation between indexes of corruption and rates of economic growth does not help much, because corruption tends to be correlated with a lot of other things that affect growth, including the extent of competition.

This is perhaps the problem for which Laffont's item-by-item analysis is least satisfactory. All other things being equal, it is hard to know how to think of an increase in competition in a corrupt economy, or of a decrease in corruption in an uncompetitive economy. Putting aside formal models of organizations and incentives, it seems to me that a move to more competition is likely to be part of a general increase in transparency and information that will make corruption more difficult.

Similar issues are raised by commitment and capture by special interests. Laffont correctly emphasizes that implementing competition requires government commitment. Except in narrow problem areas, however, it is not very interesting to treat competition and ability to commit as independent variables. And economics has little to say about creating the political and institutional conditions that produce the power to commit.

Similarly, if governments are captured by special interests and are not interested in the welfare of their citizens, it is unlikely that they will favor competition in any arena. Monopoly rents translate too easily into homes in the countryside or bank accounts in Switzerland. I agree with Laffont's somewhat pessimistic conclusion that international financial institutions generally have little leverage under these condi-

tions. Moreover, lack of competition may be the least of the problems associated with predatory governments.

Credit Markets

I agree with Laffont that it can be dangerous to liberalize banking markets without an adequate regulatory and supervisory system. Still, I doubt that restricting competition in the banking sector is a desirable long-run substitute for such a system. In the United States we learned that deposit insurance coupled with the threat of bank failure tends to induce risk-seeking behavior. But bad loans do not result only from risk seeking—they may also reflect the laziness and cronyism that can arise in uncompetitive financial systems. With or without liberalization, adequate regulation and supervision of financial institutions is an important virtue. With adequate regulation and supervision, monopoly remains a serious vice.

Conclusion

Throughout his article Laffont suggests an important general point that he makes explicit at the end: competition is not easily created or sustained in many settings. I could not agree more. Reading most economic principles texts, one comes away with the impression that most markets are naturally competitive and that only a small dose of antitrust policy is needed to keep them that way. In fact, observing developing and transition economies makes it crystal clear that what is natural is rent seeking, not competition based on merit, and that a host of cultural, legal, political, and institutional elements must be in place for competition to arise and survive. To the extent that those elements are endogenous, the benefits of creating competition in particular settings must be weighed against the tangible and intangible costs involved. Or, to paraphrase Voltaire, one should not let the best be the enemy of the good.

Floor Discussion of “Competition, Information, and Development,” by Jean-Jacques Laffont

A participant from the European Commission asked about the relationship between regionalization and competition policy in developing countries. If a cartel is operating in, say, Brazil, it is an internal business between the government and the cartel. But when the cartel prevents an Argentine competitor from entering the Brazilian market—despite the common market agreement among Southern Cone countries (Mercosur)—it distorts competition for the entire region. Thus Argentina has an incentive to encourage effective competition policy in Brazil.

One reason why such problems occur, replied Richard Schmalensee (discussant), is that competition authorities have a narrow geographic scope—so, for example, U.S. states tend to treat native firms well. If Washington State had a competition authority, it would not be suing Microsoft (as the U.S. government is). Competition law offers a political advantage when it is broadly administered and does not impinge on local firms.

Ana Julia Jatar (discussant) added that the tendency has been to maintain competition policies at the national level. But recently there has been increasing discussion of the need for a supranational agreement to guard against specific behaviors, such as collusive practices. And in the context of Mercosur and the Andean Pact there has been discussion of replacing antidumping laws with competition laws, especially to deal with predatory pricing. Jatar believed that a similar approach would be beneficial for the European Commission.

A participant asked Jean-Jacques Laffont (presenter) about his support for cross-subsidies as a means of financing infrastructure services in countries with ineffective tax systems. If it is more efficient to raise revenues through the tariff structure of a regulated firm, asked the participant, why stop at infrastructure tariffs? Why not use the tariff structure to raise revenue for public schools and for other public goods? Because in many cases cross-subsidies are not an efficient way to raise revenue, said the participant. Too often, cross-subsidies are used as a nontransparent way of raising and spending revenue off-budget. Politicians bury cross-subsidies in incompre-

This session was chaired by Rughvir K. Khemani, manager of the Business Environment Unit in the Private Sector Development Department at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

hensible, regulated tariffs to achieve what they cannot achieve through regular tax, budget, and spending programs. Such practices are particularly common in countries that have imposed limits on politicians' ability to tax. In the United States, for example, recent electricity legislation contains extensive subsidies for things that are arguably related to electricity. They appear in the legislation only because Congress would not have been able to raise such revenue in any other way.

Laffont agreed that such practices are inadvisable. Still, he said, cross-subsidies can achieve certain goals—particularly in countries where the state is not strong enough to have a good tax system. Say that the goal is to ensure that within five years, all the poor people in a country have access to safe water. That goal is much more likely to be achieved by contracting a supplier and building cross-subsidies into the tariff structure than by demanding that the tax system be reformed. Laffont recognized, however, that cross-subsidies raise complicated issues, including the danger of nontransparent behavior. You do not want to go beyond the internal financing of the sector, he said, because the tax system is ineffective—by which he meant both inefficient and corrupt. Thus the tax money disappears (partly) when you want to use it.

Schmalensee added that it is not clear that cross-subsidies are efficient even by Laffont's standards. To determine whether they make sense in telecommunications, for example, one would have to calculate the deadweight loss from charging high prices for international calls—demand for which is quite elastic—and compare it with the efficiency of raising the revenue through the tax system. Developing countries might rely on cross-subsidies not because they are a more efficient way of raising money, but because they are a more practical way.

Laffont acknowledged that such calculations had to be made. But, he said, an efficient international telecommunications firm can effectively redistribute money between the rich and poor areas of a country, especially if it uses two-part tariffs to minimize the deadweight loss. And outcomes will certainly be better than under a tax system that spends two or three units to raise one unit of public funds because of its corruption.

Joseph E. Stiglitz (presenter of the opening address) noted that when antitrust laws were passed in the United States there was a feeling that the government would be unable to enforce them. As a result the United States has a system of private sector enforcement in which any damages awarded are tripled. He asked the speakers to comment on the potential for private enforcement in developing countries.

Schmalensee first pointed out that in the United States private enforcement accounts for the bulk of antitrust cases—as well as for the bulk of meritless antitrust cases. The lure of triple damages encourages many antitrust claims that may be relatively harmless but are not socially productive. If punishable offenses in developing countries were limited to, say, clear cases of price fixing, then private enforcement might be worthwhile. For the most part, however, private enforcement—as well as public enforcement—in developing countries should be limited to areas that can be handled with limited expertise.

Laffont added that France does not have private enforcement because it can easily be abused. But the idea behind private enforcement, it seemed to Laffont, is to

teach people how to complain. If private enforcement increased public complaints, then he fully supported it.

Jatar said that she believed that private enforcement would be extremely hard to implement in developing countries. Judicial systems lack expertise about antitrust matters and are often corrupt. Moreover, companies might use private enforcement as a weapon against one another. There would be few criteria for selecting important cases. Similarly, added Jatar, in their early years competition agencies would be more effective if they selected strategically important cases rather than accepting all the cases presented to them. Being more selective would improve jurisprudence and educate companies and consumers.

Rughvir K. Khemani (chair) agreed that it was important to limit the circumstances under which private enforcement is permitted. Still, he agreed with a model used in Australia. In that system complainants have the right to launch private action if the Competition and Consumer Commission refuses to take action on the cases brought before it. In a developing country such a model could serve as a useful check on an inefficient, incompetent, or corrupt antitrust agency.

*Financial Market
Liberalization in
Developing
Countries:
Theory and
Practice*

International Adjustment in the Face of Imperfect Financial Markets

Bruce Greenwald

Traditional open economy models assuming efficient global financial markets cannot explain the incidence or severity of sudden crises such as those recently experienced in Mexico and East Asia. Thus this article presents a new model—one that can account for many characteristics of the recent crises. The model focuses on the information asymmetry between well-informed managers of local firms and banks and less-informed outside investors and shows that imperfect information severely hinders financial market transactions, with adverse effects on long-term economic development. The qualitative response of the model to financial and exchange rate shocks is analyzed, and the response is compared with that of a perfect information model to see how well the new model accounts for aspects of the recent crises that are difficult to explain with traditional models. The model responds to a shock with a large initial downturn followed by a slow recovery. The development of effective financial intermediaries accelerates recovery, provides valuable institutional conduits for external financial support, and offers an effective means of deploying capital to its highest-return applications in support of long-term growth. Other interventions—most notably aggressively restrictive monetary policies designed to raise domestic interest rates—do more harm than good.

Effective international economic policy has always depended on a clear understanding of the market imperfections that justify public intervention. This applies equally to efforts to promote long-term growth and to those to mitigate short-term crises. Information failures in financial markets have long been recognized as a serious barrier to productivity improvements in developing countries. As a result many development efforts have tried to nurture efficient financial institutions to alleviate such problems (see Akerlof 1970 and Caprio 1992).

Efforts to understand the recent crises in Mexico and East Asia have focused on the role of imperfect financial market information about the investment activities of

Bruce Greenwald is Robert Heilbrunn Professor of Finance and Asset Management at Columbia Business School.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

financial intermediaries (Krugman 1998). Two aspects of these endeavors are noteworthy. First, the models used to examine the East Asian and Mexican crises bear little or no relationship to the well-developed models of imperfect financial market information that have been used to analyze longer-term growth problems. Because the problems appear to have common roots, common models have much to recommend them. Second, some policy prescriptions offered in response to the recent crises are entirely unrelated to the market failures that have been identified as putative causes. Indeed, in the context of models of financial market imperfections, some policy proposals appear more likely to exacerbate than to alleviate the adverse consequences of these crises.

This article presents a simple model of one type of financial market imperfection—the information asymmetry between well-informed managers of local firms and banks and less-informed outside investors—that has been used to address issues of long-term growth (see Greenwald, Kohn, and Stiglitz 1990). The qualitative response of the model to financial and exchange rate shocks is analyzed and compared with the responses of a perfect information model to see to what extent this type of imperfect information can account for aspects of the recent crises that are difficult to explain with traditional models. The model is then used to analyze policy interventions in the kind of financial crises seen in East Asia and Mexico and to arrive at specific policy conclusions.

A Benchmark Model

As a first step in examining the kinds of forces that might be at work in a sudden financial crisis, it is useful to look at a simple, full-information, efficient-markets model of global finance. This model is intended to capture what it means for a small open economy to operate in an integrated world of global output and financial markets in the absence of information imperfections.

Specifications

The first element in the model is an export sector, which is characterized by a single representative firm selecting a level of output in each period. Formally, the firm is assumed in period t to

$$(1) \quad \text{Max } E_t \left[\tilde{p}_{t+1}^I \tilde{e}_{t+1} f(l_t^I) - (1 + \tilde{r}_t^I) w_t l_t^I \right]$$

where E_t is an expectation operator evaluated using period t probability distributions, \tilde{p}_{t+1}^I is the dollar price of exports in the international market in period $t+1$ (tildes indicate random variables), \tilde{e}_{t+1} is the exchange rate in local currency per dollar in period $t+1$, f is the production function of the usual sort, l_t^I is the employment of labor in the export sector in period t , \tilde{r}_t^I is the interest rate at time t for export sector firms (for loans paid off in period $t+1$), and w_t is the wage in local currency units.

There is a single direct factor of production—labor. However, labor must be paid and employed one period before output is available for sale. Because this is a standard circulating capital model, there is a second “capital” input with an associated price r_t^I . The assumption that interest rates are sector specific while wages are not reflects the fact that capital costs usually include a sector-specific risk premium. The crucial information assumption here is that the firm maximizes expected profits. Because unsystematic production risks are fully diversified through global equity markets and systematic risks (that is, betas) do not depend on the level of production, the particulars of the distribution of returns to investors do not matter to the firm’s decisionmakers. The demand part of this export sector is captured by the global output price (p_{t+1}^I), which is assumed to be given exogenously to firms in a small open economy independent of their output decision. Finally, the interest rate (\tilde{r}_t^I) is represented as a random variable because of the possibility of default, so actual debt repayments may be below contractual levels.

To simplify the analysis of this model, I do not consider the transitory dynamics of exchange rate evolution. Instead I look at the comparative statics of a change from a precrisis steady-state level of exchange and interest rates. I also assume, for simplicity’s sake, that looking forward from period t , the distributions of exchange rates and world prices are mutually independent. Under these conditions equation 1 becomes

$$(1') \quad \text{Max } p_{t+1}^I f(l_t^I) - \left(1 + r_t^I\right) \left(\frac{w_t}{e_{t+1}}\right) l_t^I$$

where p_{t+1}^I is the expected value of dollar prices in period $t+1$ (looking forward from period t), r_t^I is the expected cost of interest paid in period $t+1$ (based on period t interest rates), and $1/e_{t+1}$ is the expected exchange rate in period $t+1$.

The maximization (rescaled so that it is expressed in dollars rather than local currency) yields both the demand for export sector labor and the level of export sales and output as a function of factor prices (w_t, r_t^I), the expected exchange rate (e_{t+1}), and expected international prices. For a small open economy this supply is always equal to international demand, since it can be sold in any amount on the international market without affecting international prices.

The second element of the model is a domestic output sector. Again, it is characterized by a representative firm that

$$(2) \quad \text{Max } E_t \left[\tilde{p}_{t+1}^D g(l_t^D) - \left(1 + \tilde{r}_t^D\right) w_t l_t^D \right]$$

where \tilde{p}_{t+1}^D is the local currency price of domestic production in period $t+1$, g is the production function of the usual sort, l_t^D is the employment of labor in the domestic sector in period t , and \tilde{r}_t^D is the interest rate at time t for domestic sector firms (for loans to be paid off in period $t+1$). This equation is completely analogous

to the specification of the export sector except that quantities are calculated in domestic currency. Taking expectations as in the export sector, the firm's problem becomes

$$\text{Max } p_{t+1}^D g(l_t^D) - (1 + r_t^D) w_t l_t^D$$

where p_{t+1}^D is the expected domestic price level in period $t+1$ and r_t^D is the expected return on funds in the domestic sector. A specification of the demand for domestic output is required to complete the description of this domestic sector equilibrium, but I defer providing it for the moment.

The third element of the model is the determination of the firm's cost of capital. I assume that there are no barriers to capital mobility; that global capital markets are efficient, with investors optimally diversifying their portfolios across national markets; and that there is no trend in exchange rates (and implicitly in local price levels) that would create a wedge between local and global nominal interest rates. The last of these assumptions is made purely for expositional convenience so that we do not have to distinguish continually between real and nominal costs of capital. Under these conditions local and national costs of capital will be equalized, and cost of capital determination takes the form

$$(3a) \quad r_t^I = r_{ot} + \beta^I (RP)_t$$

and

$$(3b) \quad r_t^D = r_{ot} + \beta^D (RP)_t$$

where r_{ot} is the global risk-free return on capital, RP_t is a global risk premium, and β^I and β^D are the betas appropriate for investment in the export and domestic sectors of the small open economy.

The fourth element of the model is the description of labor supply and the determination of the wage level (w_t). I use a very simple specification

$$(4) \quad w_t = ap_t^D + bp_t^I e_t$$

where a and b are constants. The easiest way to interpret this equation is as a constant real reservation wage determined by alternative employment in a large subsistence sector, by local custom, or by Keynesian-like real rigidities. It could also be interpreted as an efficiency wage determined by any of several shirking, turnover, or worker effort considerations. The choice of this specification greatly simplifies the exposition and does not significantly affect the broad implications of the model. To the extent that it does bias matters, it is in the direction of larger rather than smaller output fluctuations, since it rules out changes in real wages

that might dampen the impact of output price fluctuations on output levels. Finally, it may actually be a reasonably appropriate description of wage level determination in developing economies with large agricultural labor pools (Stiglitz 1972).

The fifth element of the model is the determination of the domestic price level (p_t^D). To keep matters as simple as possible I assume that

$$(5) \quad p_t^D = kM_t$$

where k is a constant and M_t is the money supply in period t . This is a simple quantity theory specification without interest rate or output elasticity to the demand for money. The absence of interest rate elasticity has few substantive implications because interest rates are fixed exogenously by the free global flow of investment funds. The independence of money demand from output levels is slightly more significant but not seriously so. With a fixed money supply, a decline in output will lead to a decline in the demand for money and, in a more traditional quantity theory model, to a rise in the price level. In many models this will tend to mitigate the original decline in output (although not when the classical dichotomy holds). Thus not including output in equation 5 intensifies output fluctuations in this model relative to models with a more comprehensive specification of money demand. However, I am mainly concerned with analyzing situations in which, with suitable adjustment in M_t , the monetary authorities effectively control the price level. This is fully captured by equation 5.

The final elements of the model are the specification of output demand in the market for domestic goods, the market clearance condition for the goods market, and the condition that the balance of payments be in balance. It is not necessary to account for any of these in detail. The remaining undetermined variable is the exchange rate, e_t . The model as specified so far is not easily compatible with fixed exchange rates (although it can be made to be). Consequently, the exchange rate (e_t) is a variable rather than a parameter of the model. Given a monetarily determined price level (p_t^D), international prices (p_t^I), international financial market conditions (determining r_t^I and r_t^D), and a level of reservation wages (w_t), the exchange rate will be determined by the condition of equilibrium in the domestic goods market.

If that market is unbalanced—that is, if savings are greater than investment because of, say, an increase in the propensity to save—there will be a net flow of funds overseas and a rise in the exchange rate (in other words, a devaluation). The falling value of the domestic currency will reduce imports and increase the domestic value of export-related factor income. Together these forces will increase demand for domestic output until the domestic goods market clears and the excess outflow to foreign markets ceases. A detailed description of how this process will work is necessarily complicated. However, we can greatly simplify matters by treating the exchange rate as a variable under the control of the fiscal authorities, just as we treated the price level as under the control of the monetary authorities.¹ Given this approach the full model is now completely specified, with the exchange rate (e_t) implicitly treated as a parameter.

For the purposes of analysis the model now consists of two basic equations describing output levels in the export and domestic sectors. These have been obtained simply by substituting out the wage level (w_t) and imposing the steady-state conditions that $p_{t+1}^I = p_t^I$ and $p_{t+1}^D = p_t^D$. Equations 1 and 2 become

$$(6) \quad \text{Max } p_t^I f(l_t^I) - (1 + r_t^I) \left(a \frac{p_t^D}{e_t p_t^I} + b \right) p_t^I l_t^I$$

and

$$(7) \quad \text{Max } p_t^D f(l_t^D) - (1 + r_t^D) \left(a + b \frac{p_t^I e_t}{p_t^D} \right) p_t^D l_t^D$$

where $p_t^I e_t / p_t^D$ is the price of foreign goods relative to domestic goods. Together with the level of interest rates (equations 3a and 3b), equations 6 and 7 describe how various levels of domestic prices (monetary policy) and the exchange rate (fiscal policy) will affect the level of economic activity in a small open economy.

Observations

The first important observation that can be made about this model is that as a policy goal "driving up interest rates" is simply not possible. Nominal interest rates can be driven up through a policy of domestic price inflation (and accompanying devaluations), but real interest rates are fixed by global financial market conditions (equations 3a and 3b). Furthermore, it is not clear that financial crises are possible in this context. Any sudden withdrawal of funds by one group of investors should be met, with only negligible movements in returns, by the portfolio adjustment of masses of other global investors.

Costs of capital can indeed rise with increases in the perceived riskiness of a local economy. But the appropriate measure of risk is not the level of local uncertainty but the covariability of that uncertainty with global financial market conditions. To the extent that financial crises are the unpredictable result of local economic conditions, they represent unsystematic risks from the perspective of global investors, and fully diversified investors should be only minimally concerned about them. From this perspective what is significant about the Mexican and East Asian crises is that they were not the result of a global economic downturn. Thus, in theory at least, these crises should not have been accompanied by any significant increase in the cost of capital to domestic or export producers.

Clearly the model is at fault here. Crises do occur, and costs of capital do rise as a consequence. Moreover, global investors are not fully diversified. Home country investment preferences are a well-established fact of international financial life (see

Feldstein and Horioka 1980). The failure of the model has important consequences for thinking about actual policy formulations. This failure means that imperfect information and irrational investor behavior play an important role in real economies, making it inadvisable to formulate policy without thinking explicitly about the nature of these deviations from efficiency.

The second observation about the role of capital costs in this model is less immediately obvious. The short-term impact on output of increasing capital costs should be relatively small. Consider a firm in either the export or domestic sector whose production cycle is about two months. An increase in real interest rates from 6 percent to 24 percent a year will increase the cost of output from 1.01 times the cost of wages to just 1.04 times the cost of wages. It would not take much of a decline in wages to offset this increase.

Of course, this result is partly an artifact of the absence of fixed capital from the model. But the model can easily be extended to include fixed capital simply by having future output depend not just on I_t^D but on a whole sequence of past labor inputs ($I_{t-1}^D, I_{t-2}^D, \dots$) that are embodied in plant and machinery at time t . This adjustment does not significantly increase the response of output to period t interest rates. Earlier labor inputs are fixed costs by the beginning of period t , and the associated debt costs (and equity costs) are sunk. Changes in the current and future cost of capital will affect output only slowly as capital stocks slowly adjust to the higher marginal cost of capital. This is a familiar aspect of supply-based models and should apply generally to a small open economy, especially to its export sector, which sells its output in a large international market.

A third characteristic of the model is that devaluations associated with financial crises should not have significantly adverse impacts. For the export sector this is immediately obvious from an inspection of the maximization of equation 6, which is cast in dollar terms. A devaluation lowers the cost (in dollars) of wages in the export sector. Thus a devaluation will expand export supply and, at fixed international prices, export earnings. This effect is partly offset when the domestic sector theoretically experiences an increase in real wages in local currency, which happens when workers increase their local currency reservation wage to compensate for their declining ability to purchase imports. But in practice this effect is likely to be small because the values of leisure, nonmarket employment, and other alternatives underlying reservation wages are likely to depend more on local opportunities than on access to imports. (In the language of the model, a in the reservation wage equation 4 is likely to be much larger than b .)² On balance, therefore, the effect of even large devaluations is likely to be positive and certainly not overwhelmingly negative.³ This characteristic of the model also appears to conflict substantially with observed reality.⁴

A fourth point that follows from the model is that the local stock market and other fixed asset prices play no significant role in the level of economic activity. A collapse in asset prices does not immediately affect supply conditions (equations 6 and 7) or the cost of capital (equations 3a and 3b). It may affect local demand for goods, both imported and locally produced. But with flexible exchange rates, the decline in demand leads to a decline in the value of the local currency, which equi-

librates local supply and demand. Because local supply is largely unaffected (except marginally through the wage effects of devaluation), the overall impact on domestic output should be slight and the impact on exports should be positive. Again, this theoretical result appears far removed from actual experience.

A final point follows generally from this discussion of devaluations and asset price deflation but not explicitly from the model itself. Suppose that we concede that interest rates are not completely determined by optimally diversified global investors and that concerns about the financial viability of local firms in the face of declining currency and stock market prices may drive up the cost of capital. This effect should be limited to the domestic sector of the economy, since the prospects of the export sector improve unambiguously. Moreover, with home country investment preferences and no other information imperfection, capital flight from the domestic sector should be predominantly to the export sector, rather than overseas. Thus the cost of capital to the export sector might actually decline. This was not the case in East Asia or Mexico.

These contradictions suggest that even reasonable modifications of traditional small open economy models to accommodate more realistic financial market assumptions cannot account for the observed nature of financial crises. The information imperfections involved appear pervasive and deeply embedded in the structures of the economies in question. Thus it is important to develop careful, detailed models of the effect such information imperfections have on these economies to ensure that corrective policy prescriptions are properly conceived.

A Model of Imperfect Information

Models of imperfect information fall into two broad categories: moral hazard and adverse selection. This article investigates the impact of adverse selection in financial markets. The underlying assumption is that firms are better informed about their prospects than are investors at large (for empirical support for this proposition see Mayer 1990; Asquith and Mullins 1986; Ritter 1991; Robinson 1997).

The consequences of this information asymmetry for firms' access to equity capital is relatively straightforward. If at a given market price—say, \$10 a share—a firm's management is eager to sell stock, then it is unlikely that well-informed managers will regard the firm's future prospects as justifying any current stock price significantly above \$10 a share. If the firm's prospective value were greater than \$10 a share, the firm would issue debt in the short term (which a firm with good prospects should easily be able to service). Stock would be issued only after the firm's good prospects were reflected in its share price. In contrast, if well-informed managers saw future income as worth less than \$10 a share, they would sell stock at \$10 a share.

Outside investors, aware of the incentives facing managers, should react negatively to a firm's decision to sell stock, and new equity sales should not be easily achieved. This intuition is confirmed empirically by the overwhelming reliance of firms on retained earnings rather than sales of shares as a source of equity funding, by the fact that firms that issue equity experience significant stock price declines on

announcement of their intention to do so, and by the fact that equity-issuing firms produce returns for their stockholders substantially below those of the stock market as a whole in the years following equity issues. For simplicity I assume that firms cannot issue equity.

A similar, if slightly more intricate, phenomenon arises in debt markets when borrowers are better informed than lenders about future risks and prospects. As the observed financial condition of firms deteriorates and they become riskier borrowers (with higher default probabilities), the traditional full-information assumption is that higher contractual interest rates will offset the rise in defaults. But with asymmetric information, safe borrowers respond to higher interest rates by dropping out of the loan pool, and the remaining borrowers compensate for higher interest rates by undertaking riskier projects with still greater default probabilities. The net result is that expected returns to lenders may fall as contractual interest rates rise beyond a certain level. Thus there is a ceiling on contractual interest rates at which loan demand must be controlled through credit rationing. As in the equity example, the net result is quantitative restrictions on firms' access to external finance (Stiglitz and Weiss 1981). Although I do not model these borrowing restrictions explicitly, I discuss their qualitative impact below.

The consequences of the equity issue restriction on firms are twofold. First, if firms cannot sell equity freely, their operating risks cannot be fully diversified among economic agents. Thus firm managers will no longer maximize the expected value of future profits but will instead maximize the expected utility to be obtained from the uncertain level of future profits with which the firm must live. This strategy is obvious for entrepreneurial and family firms (Leland and Pyle 1977). If for information reasons entrepreneurs or family owners must hold a disproportionate share of their wealth (relative to complete diversification) in the family firm, they will be sensitive to variations in their returns rather than the level of their returns alone.

Professionally managed firms will respond to information failures of financial markets in a similar way. Indeed, from an agency perspective the difference in behavior between owner-managed and professionally managed firms may not be all that significant. If professional managers enjoy the benefit of private information and by using it may appropriate for themselves some fraction of their firm's equity value, they are effectively equity owners with a concentrated interest in the firm similar to that of an owner-manager. Furthermore, if the firm is large, this claim will necessarily constitute a significant fraction of the managers' net worth, and they will accordingly be sensitive to risk as well as expected returns when making operating decisions. Professional managers who are subject to compensation schedules that are linear in the end-of-period values of their firms will, in maximizing the utility of this compensation, be concerned with both the risks and the expected value of their firm's performance. Finally, bankruptcy-averse managers will, under reasonable conditions, act very much like owner-managers who maximize an expected utility rather than the expected value of profits (Greenwald and Stiglitz 1993).

The second important consequence of the equity restriction is that the equity position of the firm at the beginning of any period is an inherited state variable, not a matter of management choice. This means that when managers examine the utility implication of end-of-period equity value positions, the firm's initial equity position will play a critical role, entirely analogous to that of wealth for a household maximizing the utility of uncertain returns. Indeed, if the utility function is characterized by constant relative risk aversion, then the scale of risk-taking activities will be linear in the initial equity position of the firm, just as the investment levels of such a household are linear in its initial wealth position. Because a firm's risk-taking variables include output, hiring, investment, and, for financial firms, lending levels, the operating decisions of firms in imperfect information environments are subject to "wealth" effects in addition to the price effects of the traditional theory of firm behavior.

Specifications

These points can be summarized in a model of firms operating in an equity-constrained environment similar in all other respects to that of the traditional open economy model described above. Again the starting point will be the export sector. Now firms will

$$(8) \quad \text{Max } Eu \left[\tilde{p}_{t+1}^I f(l_t^I) - (1 + \tilde{r}_t^I) \left(\frac{w_t l_t^I}{e_t} - a_t^I \right) \right]$$

where a_t^I is the equity level (in dollars) of export firms at the beginning of period t :

$$a_t^I = p_t^I f(l_t^I) - (1 + r_{t-1}^I) \left(\frac{w_{t-1} l_{t-1}^I}{e_{t-1}} - a_{t-1}^I \right) - d_t^I$$

where d_t^I is export sector dividends and the other variables are defined in exactly the same way as in the traditional model. The difference between the firm's equity (a_t^I) and its circulating capital investment ($w_t l_t^I / e_t$) is the amount it must borrow at the beginning of period t .

One significant simplification has been made for expositional convenience. Export firms are assumed to be financed entirely by foreign sources and to have no domestic operations whose value fluctuates with the exchange rate. Under such circumstances the export sector should not be adversely affected by a crisis in local financial markets. In practice, of course, these conditions do not hold. Local crises spill over to the export sector through local sources of financing and the operating performance of local divisions of exporting firms. For expositional reasons these issues are deferred until later.

The impact of imperfect financial market information in the domestic sector is similar to that in the export sector. Formally, domestic firms

$$(9) \quad \text{Max } Eu \left[\bar{p}_{t+1}^D f(l_t^D) - (1 + \bar{r}_t^D) (w_t l_t^D - a_t^D) \right]$$

where a_t^D is the equity level (in domestic currency) of domestic firms at the beginning of period t , and the other variables are defined as before. Formally,

$$a_t^D = \bar{p}_t^D f(l_{t-1}^D) - (1 + r_{t-1}^D) (w_{t-1} l_{t-1}^D - a_{t-1}^D) - d_t^D$$

where d_t^D is domestic dividends. As with the export sector, domestic borrowing at the beginning of the period is investment ($w_t l_t^D$) minus firm equity (a_t^D). The labor market condition need not be changed (that is, w_t is a reservation wage determined by both import and domestic prices) from that of the traditional model. Also for simplicity, I continue to assume that the domestic price level is determined by the money supply level. However, the description of the financial sector differs appreciably from that of a world of perfect information.

First, I assume that loans are supplied by a banking sector described by a single representative bank. This bank, like any firm, has limited access to equity markets and therefore maximizes not expected profits but the expected utility of its terminal value position. Formally, this can be written as

$$(10) \quad \text{Max } Eu \left[\left((1 + \bar{r}_t^D) (w_{t-1} l_t^D - a_t^D) - (1 + \bar{r}_t^I) \bar{e}_{t+1} F_t^B - (1 + r_t^o) \right) \right. \\ \left. \left(w_t l_t^D - a_t^D - a_t^B - M_t - e_t F_t^B \right) \right]$$

where F_t^B is foreign borrowing (in dollars) by banks in period t , r_t^o is the domestic interest rate on safe domestic borrowings by banks in period t , M_t is the level of noninterest-bearing demand deposits in local currency (money) held by banks in period t , and a_t^B is bank equity (in local currency) at the beginning of period t :

$$a_t^B \equiv (1 + r_{t-1}^D) (w_{t-1} l_{t-1}^D - a_{t-1}^D) - (1 + r_{t-1}^I) e_{t-1} F_{t-1}^B - (1 + r_{t-1}^o) \\ \left(a_{t-1}^B + M_{t-1} - (w_{t-1} l_{t-1}^D - a_{t-1}^D) + e_{t-1} F_{t-1}^B \right) - M_{t-1} - d_t^B$$

where d_t^B is bank dividends in period t .

The last term in the maximization represents the bank's payments on its interest-bearing net debt (assumed to be positive), which is the difference between other available funds ($a_t^B + M_t + e_t F_t^B$) and total industrial lending ($w_t l_t^D - a_t^B$). In addition to bank behavior, financial sector equilibrium requires the determination of r_t^o , which I assume is now under the control of the monetary and fiscal managers of the small open economy.

In analyzing the impact of crises and possibly corrective policy action in this context, I continue to assume that, except with respect to labor supply, the export sector is independent of the rest of the economy. I also assume perfect information flow between banks and domestic producers so that they are, in effect, one integrated industrial group. This assumption is justified because it captures the reality of many crisis-prone developing economies in which banks and industrial firms are closely integrated. Moreover, in the present context, which focuses on the impact of information-related impediments to the flow of funds, this assumption minimizes the impact of various shocks. As the full-information model showed, free financial (and trade) flows tend to eliminate or at least strongly attenuate the impact of shocks. To the extent that shocks maldistribute funds between financial and industrial firms, the assumption of joint maximization (or free flow of funds) eliminates these effects. Thus in a fully general model, fund distribution among firms will play an important additional role in intensifying and prolonging the adverse impact of shocks (Greenwald and Stiglitz 1993).

Combining the banking and domestic production sectors yields a combined decision problem, as follows:

$$(11) \text{ Max } Eu \left[\tilde{p}_{t+1}^D f(l_t^D) - (1 + r_t^o) (w_t l_t^D - a_t^D - a_t^B - M_t - e_t F_t^B) - (1 + \tilde{r}_t^I) \tilde{e}_{t+1} F_t^B \right]$$

where a_t^D and a_t^B evolve over time with realized profits less dividends.

Observations

The responses of the domestic sector of the small open economy can be analyzed from equation 11. They are very different from those of the full-information economy. First, the effects of higher interest rates on output are negative, as in the traditional model, but are larger and longer lived. There is, to begin with, the traditional supply effect as labor inputs one period in advance become more expensive (that is, the coefficient of l_t^D rises). Beyond this, as long as the banking-industrial sector is a net domestic debtor ($w_t l_t^D - a_t^D - a_t^B - M_t - e_t F_t^B > 0$), an increase in interest rates shifts wealth and equity from this sector to the domestic household sector. This, in turn, reduces output to a degree that can (in a world of constant relative risk aversion) be far greater than the traditional supply effect.

Next, as both profits (at a given level of output) and the level of output decline, the time path of equity formation ($a_{t+1}^D + a_{t+1}^B$) shifts downward, reducing future firm wealth and equity and further depressing output. Furthermore, if the rise in interest rates is accompanied by a decline in the (real) money supply, the wealth and equity of the business sector is further impaired as its funding shifts from zero-interest demand deposits or notes (assuming that interest rates paid on money are controlled at zero) to higher-interest sources. Finally, if banking-industrial firms own significant amounts of market securities directly and the effects described so far sig-

nificantly impair the value of those securities (by reducing a_t^D and a_t^B and future output and equity), then the business sector's wealth and equity position will be further impaired by a further decline in output, and so on.

A second category of effects concerns uncertainties in prices and exchange rates. In the traditional model uncertainty does not matter because it is diversified away. With imperfect information this is no longer the case. An increase in uncertainty about domestic price levels (\bar{p}_{t+1}^D) or exchange rates (\bar{e}_{t+1}^D) will generally reduce output among risk-averse firms both directly and through the impact on reservation wages. Thus, once in motion, a crisis of unknown dimension can seriously reduce output because of rising uncertainty. These effects will be reinforced by the wealth and equity effects of lower output and lower profits.

Finally, devaluations in this model have a significant negative impact on output beyond their effect on real wages and the labor supply. A rise in the expected exchange rate (such as a devaluation) leads to a net transfer to foreigners of wealth and equity from the domestic banking-industrial sector. The result is an immediate reduction in domestic output and further reduced profits in the future.

Until now I have assumed that the export sector operates independently of the domestic business and financial sectors. Under these circumstances the only negative impact of a crisis is that of rising exchange rate uncertainty, which may partly offset the positive effect of real wage reductions resulting from a currency devaluation. But if the export sector is connected to the domestic sector through the banking system or through the domestic production divisions of the exporting firms, the financial effects outlined above will spill over into the export sector and counteract the positive effects of a currency devaluation on the level of output.

The simplest way to model this would be to aggregate all three sectors—banking, export, and domestic production—into a single firm, just as I did for banking and domestic production. The effect is to introduce negative wealth and equity effects into the output levels of export operations because of rising domestic interest rates, declining domestic asset values, an increasing domestic burden of overseas debt, and reductions in the level and profitability of domestic operations that affect future firm equity levels. As a result, in a world of imperfect financial market information an international crisis may lead to an impairment of export operations to an extent that is not theoretically possible in a more traditional open economy model.

A final element of the crisis story that I have not modeled explicitly is the effect of a crisis on debt markets. As firms' equity levels decline, loans become more risky. Banks that are more risk averse because of their impaired equity positions may try to raise contractual interest rates. But as noted, higher interest rates cause a further deterioration in the quality of the loan population, as better borrowers drop out of the loan pool and the remaining borrowers adopt riskier postures (especially once equity positions become negative). This effect makes it much less desirable to raise contractual interest rates. As a result overall loan levels may fall (at optimal contractual interest rate levels) and credit may be severely rationed. Thus just when their equity funding is impaired by losses, firms in a crisis economy may find their access to loans severely restricted.

Taken as a whole, this basic imperfect information model generates a picture of a financial crisis in a small open economy that is far closer to reality than the traditional full-information model described earlier. An initial loss of confidence in financial and foreign exchange markets impairs the equity positions of firms, which leads to a decline in economic activity and profitability. These changes lead to a further decline in asset and foreign exchange markets and to an increase in the perceived uncertainty in these markets, setting off another round of output and profit reductions. At the same time, loan markets become more risky and credit rationing more severe, with still more adverse effects on output and firm profitability. At this point many firms may have no access to financial resources, enter bankruptcy, and shut down production altogether. Market imperfections create negative externalities that intensify rather than attenuate any initial disturbances.

The economy will ultimately recover if wages are flexible. Declining real wages lead to higher profitability and growing equity levels in firms and banks. These changes induce higher output and profit levels that continue the process. But recovery generally takes time, because equity levels are restored only slowly through greater retained earnings. Total welfare losses may be correspondingly large. In contrast, in fully efficient financial markets any equity losses through declining operating performance are made good with new market sales of equity, and increasing uncertainty is fully diversified among global investors. In that case wages should suffer only minimally, if at all. Experience appears to argue fairly strongly in favor of imperfect information.

Policy Implications

The policy implications of the imperfect information model are relatively straightforward. A fundamental difficulty underlying the cumulative force of any crisis is the inability of firms and the banking system to limit the financial effects of losses by attracting the right kind of new investment capital. Thus solutions will likely take the form of improving access to external finance—particularly equity finance—or, perhaps less easily, immunizing economies against systematic equity losses.

Economies can be immunized against systematic equity losses in several ways. First, if net foreign borrowing is limited to the export sector, an initial devaluation need not have an adverse financial impact on domestic firms or financial institutions. The level of foreign funding appropriate to support the export sector may not be immediately calculable because many export firms may be locally financed. But export levels by industry multiplied by industry capital-output ratios should provide an approximate magnitude, which can then be subtracted from net foreign debt to estimate domestic sector foreign debt. If developing countries were to limit their domestic sector foreign debt to zero, in theory it would eliminate the redistributive consequences of a sudden currency devaluation. If domestic sector foreign debt is not zero, the domestic equity lost in a devaluation is equal to the percentage (unexpected) devaluation times the domestic sector (including banking) foreign debt. In principle, a transfer from international sources equal to this amount will restore the

aggregate equity level of an economy experiencing a devaluation. However, redistributive consequences (among firms) of a currency shock may still have a negative impact on aggregate domestic output.⁵

Price and exchange rate stabilization also have an important role to play in minimizing risks prospectively and unexpected capital gains and losses retrospectively. The emphasis on avoiding uncontrolled domestic inflation in the aftermath of a financial crisis is fully appropriate. Higher ex ante variance in the price level (\bar{p}_{t+1} , looking forward from period t) will reduce output as firms reduce their exposure to the related risks. Different ex post realizations of the price level have implications for the distribution of equity between households and firms and overall output levels. There is a tendency in the imperfect information model for a crisis to be self-reinforcing in the short run. Exchange rates may overshoot the appropriate long-run devaluation level. For some purposes, such as to discourage herding, this overshooting may be salutary. But in the long run exchange rate stabilization should help minimize any reduction in real output and help return firms quickly to noncrisis output levels by reducing nondiversifiable uncertainty. Like price stabilization, exchange rate stabilization is an appropriate policy initiative.

In searching for financial inflows to replace any net overseas losses due to capital flight or devaluation-related increases in overseas debt burdens, it is important to remember that the form of these inflows is as important as their level. Unless properly deployed, public capital inflows will not replace private capital losses. Debt inflows with strict covenants will not replace equity losses as a means of absorbing and spreading the risks of firm operations. Private equity funds will ultimately be required to restore private output. Thus private debt-equity swaps or private equity-public debt swaps may be at least as valuable in stimulating economic activity as large financial inflows from abroad to public debt markets. The information asymmetries that create barriers to financial flows do not disappear, however, and attempting to deploy equity capital is not the same thing as deploying it successfully.

The crucial element here is an effective infrastructure of financial institutions. These are as important to crisis responses as to long-term growth in the face of information imperfections. Thus a long-term strategy of concentrating development aid on the creation of efficient banking systems has much to recommend it. Recovery from the East Asian or Mexican crises would have been much faster had banking systems been able to make loans and had more flexible equity-like investments been available to replace the crisis-related funding losses of healthy firms. To perform this role effectively, banking institutions must be able and willing to distinguish healthy from unhealthy firms.

To cite one example, it has been suggested that an international bankruptcy workout agency could mitigate the consequences of financial crises. But successful private bankruptcy reorganizations in the United States still depend on the active intervention of well-informed financial intermediaries that have an appropriate degree of control, proper institutional incentives, and access to additional financial resources. A system of well-functioning banks is capable of filling this role. In its absence it is not at all clear how such public workouts could be managed.

A comprehensive bank development program would involve the establishment of effective organizational structures for asset deployment (such as loans) and technical support on transaction processing and retail distribution of banking services. Properly instituted, such a program would also deal with the moral hazard problems that have also been blamed for recent crises. By reducing moral hazard, such a program would also have a substantial economywide value quite apart from any role it might play in crisis prevention.

Finally, it is worth identifying one popular policy—raising interest rates—that probably exacerbates rather than mitigates the effects of a financial crisis. The information failures at the heart of such a crisis are likely to raise interest rates above levels that are appropriate in the long run. This is unavoidable in any given state of imperfect information, and it is counterproductive to try to prevent interest rates from rising along a post-crisis supply-of-funds curve. But a conscious effort to restrict supply further (through tight monetary policies, for example) and thus raise interest rates to attract foreign capital will, in the context of the models developed here, lead to greater output and welfare losses than would otherwise be the case. In draining valuable equity funds from an already depressed economy, such a policy would only make matters worse.

Notes

1. Here the exchange rate rather than the interest rate is determined by the fiscal authorities (as is the case in more traditional macroeconomic models) because interest rates are fixed by international financial markets.
2. Empirically, financial crises are almost invariably associated with declining rather than increasing real wages. This is a further shortcoming of the model (although, given the conditions of the labor supply specifications, not a surprising one).
3. It might seem that the effects of devaluations on export demand and import substitution have been ignored in this discussion. They have not. One effect of devaluation in the model is to equate the demand for domestic goods to the supply level determined by equation 7. The assumption is that government policy or falling domestic investment has led to a devaluation in the equilibrium exchange rate. In focusing attention on the supply relationship of equation 7, I am assuming that import substitution on the demand side is already taking place. Treating it as a separate factor would be double counting.
4. It should be noted that industries for which imported components constitute an important input will suffer significant declines in output. However, these industries should be separately identifiable. Anecdotal evidence suggests that all industries, not just those using imported components, suffer in a financial crisis.
5. The impact of a devaluation on the prices of imported intermediate goods should not automatically be counterbalanced by capital inflows. Where foreign debt is concerned, a sunk cost is being changed; without information imperfections this would have no impact on output—hence the logic of immunization. Real resource allocation is involved with intermediate goods, and some adjustment in the levels of domestic output in affected industries is appropriate.

References

- Akerlof, George. 1970. "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 74 (July): 288–300.

- Asquith, Paul, and David Mullins. 1986. "Equity Issues and Stock Price Dilution." *Journal of Financial Economics* 13: 296–320.
- Caprio, Gerard Jr. 1992. "Policy Uncertainty, Asymmetric Information, and Financial Intermediation." Policy Research Discussion Paper 853. World Bank, Washington, D.C.
- Feldstein, Martin, and Charles Horioka. 1980. "Domestic Savings and International Capital Flows." *Economic Journal* 90 (June): 314–29.
- Greenwald, Bruce, and Joseph E. Stiglitz. 1993. "Financial Market Imperfections and Business Cycles." *Quarterly Journal of Economics* 108 (February): 77–114.
- Greenwald, Bruce, M. Kohn, and Joseph E. Stiglitz. 1990. "Financial Market Imperfection and Productivity Growth." *Journal of Economic Behavior and Organization* 14: 117–28.
- Krugman, Paul. 1998. "What Happened to Asia?"
- Leland and Pyle. 1977. "Information Asymmetries, Financial Structures and Financial Intermediaries." *Journal of Finance* 32 (May): 371–87.
- Mayer, Colin P. 1990. "Financial Systems, Corporate Finance, and Economic Development." In R. G. Hubbard, ed., *Asymmetric Information, Corporate Finance, and Investment*. Chicago: The University of Chicago Press.
- Ritter, S. 1991. "The Long Run Performance of Initial Public Offerings." *Journal of Finance* 66 (January): 3–27.
- Robinson, Laura O. 1997. "The Underperformance of Public Firms with ESOPs." Milken Institute Working Paper 42. Santa Monica, Calif.
- Stiglitz, Joseph E. 1972. "Alternative Theories of Wage Determination and Unemployment in L.D.C.'s: The Labor Turnover Model." *Quarterly Journal of Economics* 87 (April): 194–227.
- Stiglitz, Joseph E., and A. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (September): 393–440.

Comment on “International Adjustment in the Face of Imperfect Financial Markets,” by Bruce Greenwald

Gerard Caprio Jr.

Bruce Greenwald provides a highly instructive framework for understanding and analyzing the role of information asymmetries in financial crises of the type seen in East Asia. He begins by reminding us that the traditional open economy, full-information model cannot explain such crises. Although this point may seem obvious, it bears repeating: economists using such a framework in the summer of 1997 projected only a mild slowdown in Thailand as a result of the financial crisis there. Greenwald next shows how imperfections arising from information problems explain various features of these crises. Finally, he derives policy recommendations.

I will comment briefly on some of the features of his framework and then his policy recommendations. I agree fully with his conclusions on the need to develop effective financial intermediaries and the role of improved infrastructure, and I note that the credit implications of those recommendations support his conclusion on the stance of macroeconomic policies—perhaps even more than he acknowledges.

Extending the Framework

One quibble with the article is that the assumption of perfect information between firms and banks goes further than I would prefer. To be sure, modeling choices are always tricky, given the inevitable tradeoffs between simplicity and realism. Still, in this case I would gladly accept any loss of tractability, because in my view the norm in developing countries is a looser relationship between banks and firms. Indeed, one reason for a post-crisis credit crunch is that information capital has been wiped out, either through the failure of banks (as occurred during the U.S. Great Depression) or the failure—or at least loss of attractiveness—of firms on which banks had information (Caprio 1992; Bernanke 1983).

To the extent that the firms that will benefit from an exchange rate depreciation are new or young firms that did not have close ties to banks, any credit crunch will

Gerard Caprio Jr. is research manager in the Development Research Group at the World Bank and director of the Financial Economics Group.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

then be longer and deeper. And it will be difficult to determine whether insufficient demand or insufficient credit is the problem: firm surveys may reveal no concern about credit but indicate that lack of demand is a problem. Of course, if there were more credit, demand would be greater, and if demand increased (say, from abroad), credit might be revealed to be a more explicit bottleneck. Investment forecasts based on standard models will miss this effect, and it is hard to analyze it if banks and firms are lumped together. This point is valid regardless of the openness of the capital account or how the crisis was triggered.

If banks and firms were separated, one could include credit as an input in the supply of exports (or output more generally), with the supply of credit in turn a function of variables affecting the profitability and solvency of banks (and, of course, of monetary policy). This approach should help capture the loss of information associated with a crisis and facilitate the link to discussions about policy stance.

Policy Implications

I agree with most of Greenwald's policy conclusions. Indeed, I could not agree more with the recommendation to improve financial sector infrastructure in response to crisis—especially since I think doing so is also important before a crisis. The only problem with Greenwald's prescription is that it will take considerable time to effect. Often the focus during the immediate aftermath of a crisis is on some quick response. Thus while such improvements are needed, they will disappoint many policymakers.

More significantly, I would tease out a few additional points. First, even for my taste, Greenwald overplays the role of information asymmetries in limiting equity finance. I understand the argument but believe that these imperfections can be mitigated. Even without well-developed disclosure rules, relatively poor U.S. residents held more wealth in stocks than wealthier U.K. residents for most of the 19th century, in part because U.S. banks were limited in branching nationwide. As firms' financing needs outgrew local banks' capacity, they tended to access national securities markets. By the turn of the century the U.K. market overtook the U.S. market because of better disclosure, only to be displaced after the United States adopted Securities and Exchange Commission regulations (Sylla 1997). Improving the legal system and the rights of shareholders can facilitate equity market development and promote faster growth (Levine, Loayza, and Beck 1998).

What is the relevance of this point? Throughout history, most crises have featured high debt-equity ratios. Beyond Greenwald's recommendation of limiting nonexporting firms from borrowing abroad—which may be difficult to enforce when foreign interest rates are low—better contract enforcement and superior shareholder rights will help tip the scale in favor of equity finance. Of course, limiting the deductibility of interest and not double taxing dividends would increase debt finance. Although industrial countries, especially the United States, may be able to afford such policy biases, they are less affordable in developing countries, where systematic risk is greater and high debt-equity ratios are riskier.

Second, I disaggregate financial infrastructure improvements into two parts: the information system (meaning accounting, auditing, disclosure requirements, and the legal system) and the incentive system, or what is usually referred to as the regulatory framework, because it determines how information is used. The question I would pose is, how much post-crisis improvement in financial infrastructure is desirable, and what does it imply for macroeconomic policy? Let me be clear: better information and incentives are needed before or after a crisis. But improving disclosure might amplify or mitigate post-crisis recessions, depending on how the disclosed information compared with expectations. If I thought that the true state of the banking system was significantly weaker than was generally appreciated and that disclosure of its true status would produce a contraction, I would want to be ready for this eventuality by easing macroeconomic policies.

Tighter financial sector regulation is even more likely to require offsetting efforts. Raising capital or liquidity requirements or limiting entry to raise franchise value will make bankers less inclined to lend and represents a tightening of credit policy with the same reserve or interest rate position. Recent research at the U.S. Federal Reserve demonstrates the existence of a separate credit channel, indicating that this point is empirically significant (Peek and Rosengren 1997).

I would end with an amplification of James Tobin's story about the abandonment of the gold standard by the United States and the beginning of recovery during the Great Depression (see elsewhere in this volume). It is true that there was some link between the two events. But two other events also occurred. First, on March 6, 1933, a nationwide bank holiday was declared, at the end of which many banks remained closed. The closure of the weakest banks, together with some government commitment that the reopened banks were solvent (and an implicit government guarantee), ended the bank runs (before an explicit deposit insurance system entered into force). Second, although the Federal Reserve's tight monetary policy was blamed for deepening and extending the Depression, it went in the other direction on bank regulation, calling on supervisors to adopt more lenient standards. Presidents Hoover and Roosevelt, along with Treasury Secretary Morgenthau, all urged regulatory forbearance in the 1930s and appear to have been effective (Grant 1996).

Closing banks sends signals that the incentive system is being made tougher, though this, as seen recently, can be dangerous when not done on a comprehensive basis. If other improvements are made to incentives—such as raising capital requirements, adopting a tougher collection policy on written-off loans, or implementing more stringent accounting standards—it would seem likely that some easing of macroeconomic policy would be required. Indeed, adopting tough regulatory measures early might be a way to win confidence in government policy and end capital flight. If expansionary macroeconomic policy steps cannot be taken, then a longer phase-in period for the improved incentives would be suggested.

Because Joseph E. Stiglitz has urged greater honesty in policy advice (see elsewhere in this volume), I have to conclude by saying that we do not know the answers here, but some principles stand out. Bruce Greenwald focuses the policy debate in the right direction, and for that he is to be commended.

References

- Bernanke, Ben. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73 (3): 257–76.
- Caprio, Gerard Jr. 1992. "Policy Uncertainty, Information Asymmetries, and Financial Intermediation." Policy Research Working Paper 853. World Bank, Development Research Group, Washington, D.C.
- Grant, James. 1996. *The Trouble with Prosperity: A Contrarian's Tale of Boom, Bust, and Speculation*. New York: Random House.
- Levine, Ross, Norman Loayza, and Thorsten Beck. 1998. "Financial Intermediation and Growth: Causality and Causes." World Bank, Development Research Group, Washington, D.C.
- Peek, Joe, and Eric Rosengren. 1997. "Collateral Damage: Effects of the Japanese Real Estate Collapse on Credit Availability and Real Activity in the United States." Federal Reserve Bank of Boston.
- Sylla, Richard. 1997. "The Rise of Securities Markets: What Can Governments Do?" In Dimitri Vitas, ed., *Reforming Financial Systems: Historical Implications for Policy*. New York: Cambridge University Press.

Comment on “International Adjustment in the Face of Imperfect Financial Markets,” by Bruce Greenwald

Ronald McKinnon

Although Bruce Greenwald’s article is purely theoretical, his analytical objective is admirable. What happens to the theory of international adjustment if the theorist imposes serious capital constraints on enterprises and banks? Does flexibility in interest and exchange rates still aid the adjustment to macroeconomic shocks in the same way as traditional macroeconomic models, without such capital constraints, would have it?

The Traditional Model

I hesitate to apply the adjective *traditional*, as Greenwald does, to macroeconomic models of open economies incorporating stationary expectations for exchange and interest rates. Even apart from stationarity, this traditional view depends on assumptions that are more restrictive than most economists realize. Greenwald first tries to show why the traditional model imposes no capital constraints on individual banks and enterprises. He stresses the model’s strong implicit assumption of perfect capital markets, as if the asset portfolios of firms and households were fully diversified internationally to eliminate both systematic risk and any problem that firms might have in raising risk capital. (Although he does not say so, hedging against specific foreign exchange risks becomes unnecessary when asset portfolios are fully diversified.) Thus in this full-information traditional model, the equity positions of banks and enterprises do not constrain their production of goods and services. (Of course, at the macroeconomic level the economy’s aggregate stocks of land, labor, and capital could still be binding constraints on output.)

This absence of equity constraints on firms enhances the potential usefulness of exchange rate and interest rate flexibility in dampening fluctuations in demand and output. But to simplify further, Greenwald also assumes—if only implicitly—that in the traditional model exchange rate expectations are stationary. Only then does a floating exchange rate become well behaved in the sense of passively adjusting to bal-

Ronald McKinnon is William D. Eberle Professor of International Economics at Stanford University.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

ance the current account: net export flows adjust to “warranted” net capital flows. In effect, Greenwald assumes (somewhat surprisingly) that these capital flows and the exchange rate are somehow under the control of the fiscal authority. Exchange rate and interest rate flexibility greatly enhance the stability of the traditional model in response to macroeconomic shocks. For example, a drop in demand for a country’s exports can easily be offset by devaluation, without unwanted side effects on firms’ equity positions.

In portraying the “traditional” approach to exchange rate determination, Greenwald chooses to ignore the more modern view that the exchange rate is a forward-looking asset price (Mussa 1979; Frenkel 1981). A change in expectations about, say, the future course of monetary policy in country A relative to country B—or whether a country can meet its collective debt denominated in foreign currencies—can lead to a sudden shift in international portfolio (currency) preferences and high volatility, with overshooting in the exchange rate. Instead of being a passively adjusting variable, an untethered exchange rate could then significantly disturb the macroeconomy, traditional or not.

Thus it seems that Greenwald’s choice of modeling assumptions for his traditional model are inconsistent. On the one hand, he assumes perfect international capital mobility: economic agents fully diversify their asset portfolios to eliminate unsystematic commercial risks and currency risks. On the other hand, he eschews the forward-looking asset market approach to exchange rate determination, which is the natural consequence of capital mobility. He cannot have it both ways.

Asymmetric Information, Adverse Selection, and Capital Constraints

But Greenwald’s traditional model is just a foil. The main contribution of his article is to present an alternative macroeconomic model in which risk-induced capital constraints on individual enterprises are important. Following Greenwald’s co-authored work with Joseph E. Stiglitz (1998) and what is now a venerable line of articles in the finance literature on asymmetric information, he posits that a representative firm’s managers know more about its future prospects than do its outside creditors, including outside shareholders.

A firm cannot easily raise new equity capital simply by selling shares at the prevailing price—say, \$10 a share. If the future profitability of the firm warranted a price greater than \$10, then outside shareholders know that—if managers were acting in their interests—the firm would use fixed-term debt to finance new investment. Conversely, if managers valued the firm’s future prospects at less than \$10 a share, selling more shares now could be advantageous to them but not to the shareholders.

As a result potential outside shareholders generally view new issues of shares with suspicion: a negative signal because of potential adverse selection among managers with inside information. Thus, other than through the flow of retained earnings, there is a natural market-based constraint on raising new equity capital. In mature economies new net issues of equity account for a very small share of the external funds raised by corporations.

A firm's ability to borrow in fixed-term debt contracts, such as bonds or bank loans, is also constrained by its historical equity position. Outside lenders worry about rising debt-equity ratios. But as long as retained earnings continue to grow, this constraint on issuing new debt is naturally more elastic than the constraint on issuing new equity, because overall debt-equity ratios need not rise. (In hindsight, East Asian banks and other lenders did not worry enough about rising debt-equity ratios in industrial corporations such as the Republic of Korea's *chaebol*.) How, then, do these capital constraints affect the ability of exchange rate and interest rate changes to dampen macroeconomic shocks?

A Macroeconomic Model Based on Adverse Selection

Greenwald chooses to ignore the problem of moral hazard arising from deposit insurance and of asymmetric information that leads to excessive risk taking by banks and the firms to which they lend (McKinnon and Pill 1996, 1997, and 1998)—an important omission to which I shall return. Instead he focuses on how adverse selection constrains the ability of any firm to issue equity and diversify risks domestically or internationally. Because the firm's and their own portfolios are imperfectly diversified, managers (perhaps also family owners) behave as if they were extremely risk averse in response to variations in the firm's income. An increase in perceived risk restrains output. In addition, because the firm's equity position is historically given, any shock that suddenly reduces the value of that equity constrains supply. If replicated across many firms, such shocks could lead to a downward spiral in real output. But how exchange and interest rate fluctuations affect the equity positions of firms depends heavily on the other simplifying assumptions that Greenwald builds into his model.

First, he assumes that the export sector finances itself with foreign exchange and is immune to exchange rate and domestic interest rate risk. Second, for the rest of the economy he consolidates firms and banks into big industrial groups as if there were a perfect flow of information between them. (No information asymmetry here.) In important measure, these industrial-banking groups finance themselves by incurring foreign currency debts at terms shorter than the time horizon over which they make their optimizing decisions. (Informally, Greenwald invites the reader to consider the export sector to be connected to this industrial-banking group.) In this case a devaluation impairs the group's equity by raising the cost of servicing foreign debt. Even the fear of exchange rate volatility could constrain output. Thus instead of stimulating output in general and exports in particular as the traditional model would have it, a devaluation could so impair the capital positions of firms that their output declines: the supply-side effect dominates the traditional demand-side effect.

I sympathize with the empirical importance of this possibility, particularly in light of the recent downward spiral in East Asian economies associated with deep (over)devaluations of their currencies. But as a purely theoretical matter, Greenwald's modeling strategy is inconsistent in focusing just on adverse selection. Within the model, if the avoidance of risk taking is paramount in the eyes of imper-

fectly diversified owner-managers, why would they leave themselves so exposed to short-term foreign exchange risk? Greenwald introduces foreign exchange liabilities into his model in a purely ad hoc fashion. Yet if bankers and industrialists were really risk averse, they should go (or in East Asia, should have gone) out of their way to hedge their foreign exchange liabilities.

The banking-industrial firm's vulnerability to interest rate increases is also modeled in an ad hoc fashion. The desire to avoid risk should naturally limit the willingness of banks to accept short-term liquid deposits on which to make longer-term illiquid loans—and limit the willingness of enterprises to rely on finance of a shorter term than their investment decision horizon. But Greenwald discusses neither term structure nor the problem of hedging against interest rate risk. When building his formal macroeconomic model, which is ostensibly centered on asymmetric information and risk avoidance, he simply assumes that industrial-banking groups leave themselves vulnerable to increases in interest rates or to currency devaluations.

In the article's final section, on policy implications, Greenwald (like everyone else) worries about the undue foreign exchange and interest rate exposure in East Asian and possibly other developing economies. But his own model throws no light on why that should be the case. His macroeconomic model based on adverse risk selection should lead agents to hedge as best they can—perhaps naturally resulting in no net foreign exchange and interest rate exposure. To be sure, in his informal comments, he strongly suggests that governments should regulate banks and firms to reduce their risk exposure. But within his formal model, where risk avoidance is paramount, such regulation would be redundant. What, then, is the key ingredient missing from his model?

Moral Hazard

Moral hazard in banks, and in the enterprises to which they made loans, was the prime cause of excessive risk taking in East Asia's crisis economies. Before the crisis, firms left uncovered their massive financial exposure, foreign currency indebtedness, and short-term borrowing. Lacking proper capital constraints, they aimed to increase current profits at the risk of some potential future catastrophe—albeit one with seemingly low probability. Rather than behaving as if they were risk averse, as Greenwald's model would have them do, in the precrisis Asian economies agents such as the Korean chaebol were aggressive risk takers—without being sufficiently constrained by their equity positions.

The reason for this moral hazard in the banks was clear enough. Blanket, if implicit, official deposit insurance on almost all classes of financial intermediaries gave them incentives to lend aggressively and to finance themselves in risky ways. Without depositors becoming nervous, banks could ignore potentially bad macroeconomic outcomes in the longer term, because they would be covered by the deposit insurance agency. Instead banks sought myopically to maximize short-term profits. This moral hazard was not offset by sufficiently tight or careful official regulation to force banks to be adequately capitalized in general and to hedge important risks, such as foreign exchange and interest rate risks, in particular.

Because deposit-insured banks were often part of industrial groups, moral hazard was easily transmitted to encourage banks' industrial borrowers to become risk takers. Huw Pill and I have stressed how an upsurge in lending by the banking system is often taken as a signal by nonbanks that the macroeconomic progress of the economy and its potential for future growth are (falsely) higher than they had previously thought. Similarly, suppose that before a crash occurs, the capital account of the balance of payments remains open. Then moral hazard leads to aggregate overborrowing, which shows up as overinvestment or overconsumption for the economy as a whole (McKinnon and Pill 1996, 1997, and 1998).

Resolving the Paradox between Adverse Selection and Moral Hazard in Macroeconomic Models

Efforts to develop a suitable macroeconomic model run into a paradox. Greenwald rightly wants to reconstruct the "traditional" model of an open economy to show that in certain circumstances exchange rate and interest rate changes might impede an economy's adjustment to macroeconomic shocks. To do this, he needs economic agents—banks and industrial enterprises—to be capital constrained according to the theory of adverse risk selection, a theory he uses intensively for modeling purposes. But to show that devaluations or increases in interest rates could be damaging, he also needs the agents in his theoretical model to be financially exposed, so that a devaluation or increase in interest rates would destroy equity and tighten capital constraints, as in the East Asian crisis.

But in Greenwald's pure risk aversion model without moral hazard, agents tend to behave conservatively and would not leave themselves financially exposed. To get the result that he wants, Greenwald imposes exposure in an ad hoc way from outside the model. He simply assumes that agents carry uncovered foreign currency indebtedness and that they are vulnerable to increases in interest rates. Thus there is a contradiction between his risk aversion approach based on adverse selection and his ad hoc assumptions. Can the paradox be resolved and the contradiction overcome? Consider the risk aversion and moral hazard modeling approaches separately.

Suppose that before the crisis, moral hazard is rampant. Because of government bailout provisions such as deposit insurance, agents take on much more foreign exchange and interest rate risk than their capital positions would normally warrant. Foreign lenders are much too profligate because they believe that they too will be bailed out if necessary. In these precrisis circumstances the principle of moral hazard leading to excessive risk taking, with an absence of suitable capital constraints, should dominate the building of the macroeconomic model.

Now model the system after the crash, when foreign lines of credit are suddenly withdrawn and domestic firms are left financially exposed. The government, perhaps with some prodding from the International Monetary Fund, now starts talking about enforcing capital requirements in banks and reducing debt-equity ratios in industry. Suddenly Greenwald's adverse selection model with risk aversion becomes appropriate. Domestic agents find themselves both capital constrained and finan-

cially exposed. Moreover, a devaluation of the national currency or an increase in interest rates worsens their financial positions—and could be counterproductive in helping the economy to adjust in the postcrisis situation.

I conclude that Greenwald's open economy macroeconomic model based on adverse selection does not really apply to the precrisis Asian economies, as he seems to intend. But with suitable modifications, it could help us understand what is going on after the crash.

References

- Frenkel, Jacob A.. 1981. "Flexible Exchange Rates, Prices, and the Role of News: Lessons from the 1970s." *Journal of Political Economy* 89 (4): 665–705.
- Greenwald, Bruce, and Joseph E. Stiglitz. 1998. "Imperfect Information, Finance Constraints, and Business Fluctuations." *Quarterly Journal of Economics* 102.
- McKinnon, Ronald, and Huw Pill. 1996. "Credible Liberalizations and International Capital Flows: The Overborrowing Syndrome." In Takatoshi Ito and Anne O. Krueger, eds., *Financial Integration and Deregulation in East Asia*. Chicago, Ill.: University of Chicago Press.
- . 1997. "Credible Economic Liberalizations and Overborrowing." *American Economic Review* 87 (2): 189–203.
- . 1998. "International Overborrowing: A Decomposition of Credit and Currency Risks." *World Development* 26 (7): 1267–82.
- Mussa, Michael. 1979. "Our Recent Experience with Fixed and Floating Exchange Rates." *Carnegie Rochester Supplement* 3: 1–50.

Floor Discussion of “International Adjustment in the Face of Imperfect Financial Markets,” by Bruce Greenwald

Bruce Greenwald (presenter) opened the session by responding to objections made by Ronald McKinnon and Gerard Caprio Jr. (discussants). The reason he lumped banks and firms together in his model, said Greenwald, is because it is much harder to disentangle the effects of higher interest rates when the two sectors are separate. Higher interest rates are good for banks but bad for firms; thus it is hard to assess their net effect on wealth. But when banks and firms are considered together, the effects of higher interest rates are much clearer. For example, if firms owe money to households, then higher interest rates transfer funds from the business sector (which needs the money to absorb risk) to the household sector (which cannot effectively redeploy the money back to the business sector). Greenwald conceded that his approach had implications for other elements of his model, but the larger point is that driving up interest rates unnecessarily is destructive to real business activity.

A participant from the Carnegie Endowment for International Peace noted that in Russia banks are too weak to provide significant funding. As a result issues of eurobonds and domestic bonds are on the rise. How, the participant asked Greenwald, do those kinds of bonds fit into the model?

Greenwald said that bonds do not help firms absorb risk if the firms intend to repay the bonds—as they do in Russia. With fixed obligations a firm that has \$400 million in outstanding bonds and faces significant sanctions will have to live with the accompanying risk. That risk influences the firm’s decision to build a \$400 million plant. In that situation bonds are just like any other factor of production.

Greenwald was missing the point, said McKinnon. Banks do not benefit from higher interest rates. If a bank’s liabilities have shorter maturities than its assets—as is usually the case—then higher interest rates weaken banks. But a highly developed bond market can move banks away from this maturity mismatch and make them less vulnerable to changes in interest rates. The Republic of Korea would do well to take such an approach, added McKinnon.

This session was chaired by John Williamson, chief economist, South Asia Region, at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

A participant from the London School of Economics asked Greenwald to elaborate on instances where authorities had raised interest rates unnecessarily. In Greenwald's view one of the best examples is the United States during the latter part of the Great Depression. During 1930–32 Australia, Germany, and the United States had similar unemployment levels. By 1936 unemployment in most other countries had returned to levels of the late 1920s. But in the United States unemployment was still close to 15 percent in 1940. Why? The actions of U.S. Federal Reserve Chairman Mariner Eckles, which led to a recession in 1937–38. In 1936 Eckles was squeezing speculative excess out of the U.S. economy by aggressively restricting monetary policy. His efforts were meant to restore confidence, but instead he lost control. And in Greenwald's view unnecessarily high interest rates are associated with restrictive monetary policy in times of crisis.

Caprio hypothesized that when governments fail to quickly address problems in the financial sector—as the U.S. government failed to do during the Depression—they end up having to adopt tighter monetary policy than would otherwise have been necessary. By moving quickly, governments might be able to stimulate a reflow of deposits by domestic and possibly even foreign residents.

McKinnon added that no one supports unnecessary increases in interest rates. But some increases are unavoidable. East Asian countries recently had to choose between raising interest rates (which devastates capital in Greenwald's model) and allowing an enormous devaluation (which can devastate capital even more in Greenwald's model). What was missing from the rescue programs offered by the International Monetary Fund and the World Bank, said McKinnon, was a future target for the exchange rate. If the domestic currency is extremely undervalued and currency holders are pessimistic about the chances for recovery, then interest rates have to be raised quite high because of the belief that the exchange rate could sink even further. But higher interest rates can be moderated by specifying a target for the future exchange rate. Moreover, if currencies in neighboring crisis countries can recover together, then interest rates can be much lower.

A participant from the U.S. Agency for International Development asked Greenwald about information asymmetry. Would it not be better for borrowers and issuers of equity to disclose their situations honestly so that financing can be secured on the most favorable terms? And if this level of disclosure is one of the big differences between emerging and advanced economies, what can the World Bank and country policymakers do to move markets toward more symmetric information?

The disclosures that matter, said Greenwald, are not those that make bad information publicly available. Rather, the disclosures that matter are those that reduce the information asymmetry between firms and outside investors—not anonymous investors, but investors that are in daily contact with the firms. Thus in many countries (other than the United States) banks make conditional investments in firms because the firms give the banks the information they want. Another outcome of the Great Depression for the United States, added Greenwald, was a weak banking system. France, Germany, and Italy, for example, are much less reliant on public financial markets. As a result it is easier for information to be disseminated in those countries.

Financial Liberalization and Financial Fragility

Aslı Demirgüç-Kunt and Enrica Detragiache

Banking crises are more common in countries that have liberalized their financial systems. But financial liberalization is less risky when a country's institutional environment is strong. High respect for the rule of law, low corruption, and good contract enforcement are especially relevant institutional characteristics. Examination of the behavior of bank franchise values after liberalization and of the relationships among financial liberalization, banking crises, financial development, and economic growth supports the view that financial liberalization should be approached cautiously where the institutions need to ensure good law and contract enforcement and effective prudential regulation and supervision are not fully developed—even if macroeconomic stabilization has been achieved.

Over the past 30 years many industrial and developing countries have liberalized their financial systems. Banks' interest rate ceilings have been eased or lifted. Compulsory reserve requirements and entry barriers have been lowered. Government interference in the allocation of credit has been scaled back. And banks and insurance companies have been privatized. In addition, some countries have actively promoted the development of local stock markets and encouraged the entry of foreign financial intermediaries.

The trend toward financial liberalization is part of a broader trend toward reduced direct intervention by the state in the economy. But in many developing countries financial liberalization is also a deliberate attempt to move away from using financial repression to fund fiscal imbalances and subsidize priority sectors—a move strongly advocated by the influential work of McKinnon (1973) and Shaw (1973). According to McKinnon and Shaw, by forcing financial institutions to pay

Aslı Demirgüç-Kunt is principal economist in the Development Research Group at the World Bank. Enrica Detragiache is an economist in the Research Department at the International Monetary Fund. The authors are grateful to Gerard Caprio Jr., Stijn Claessens, George Clarke, Philip Keefer, Ross Levine, Miguel Savastano, and Peter Wickham for helpful comments and to Thorsten Beck and Anqing Shi for excellent research assistance.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

low and often negative real interest rates, financial repression reduces private financial savings, limiting the resources available to finance capital accumulation. From this perspective, financial liberalization enables developing countries to stimulate domestic savings and growth and to reduce excessive dependence on foreign capital flows.¹

The work of McKinnon and Shaw stimulated a rapidly growing strand of research that analyzes how financial development can boost economic growth by accelerating productivity growth and mobilizing savings (see Levine 1997 for a survey).² This research includes a number of empirical studies on the relationship between financial development and growth. Most of these studies find that various measures of financial development are positively correlated with both contemporaneous and future growth rates of GDP, suggesting that by fostering financial development, financial liberalization can increase the long-run growth rate of the economy (King and Levine 1993).

This positive view of financial liberalization has been clouded by the marked increase in financial fragility experienced by both industrial and developing countries since the early 1980s. Banking sectors around the world have confronted a remarkable number of problems, some of which have erupted into full-fledged systemic crises (as documented in the extensive studies of Caprio and Kliegebiel 1996 and Lindgren, Garcia, and Saal 1996). In many cases banking problems emerged shortly after the financial sector was deregulated.³ These experiences suggest that the benefits of financial liberalization may have to be weighed against the costs of increased financial fragility, and some have prompted prominent voices in the policy debate to adopt the view that some degree of financial regulation is preferable to premature liberalization in developing countries (Caprio and Summers 1993; Stiglitz 1994).

Does Liberalization Lead to Fragility?

While the link between financial development and economic growth has been documented through careful empirical studies, the connection between financial liberalization and financial fragility has not yet been subjected to systematic econometric investigation. This article attempts to fill that gap. Building on our previous research on the determinants of banking crises (Demirgüç-Kunt and Detragiache 1997), we construct a financial liberalization dummy variable for 53 industrial and developing countries covering 1980–95.

The deregulation of bank interest rates is used as the observable policy change indicating liberalization, because case studies suggest that it is often the centerpiece of the liberalization process. The dataset encompasses countries that liberalized their financial markets well before the 1980s as well as countries that liberalized at other times during the sample period. Using a multivariate logit framework and controlling for other factors that may increase the probability of a crisis, we test whether banking crises are more likely in liberalized financial systems. The set of control variables includes macroeconomic variables, banking sector characteristics, and

institutional variables. We also test whether crises are more likely during the transition to a less controlled financial system or whether fragility is a permanent feature of liberalization.

Another issue often raised in the debate over financial liberalization is whether the dangers of liberalization are greater in countries where the institutions needed to support the efficient functioning of financial markets are not well developed. Such institutions include effective prudential regulation and supervision of financial intermediaries and organized security exchanges and a well-functioning mechanism for enforcing contracts and regulations. We investigate this issue by testing whether the relationship between banking crises and liberalization is stronger in countries with weaker institutional environments, as proxied by per capita GDP and various indexes of institutional quality. Finally, we subject our results to a variety of checks for robustness.

Our main finding is that banking crises are more likely in countries that have liberalized their financial sectors, even when other factors (including the real interest rate) are controlled for. Moreover, banking sector fragility generally does not increase until a few years after liberalization begins. The data also support the conjecture that liberalization is more likely to lead to a banking crisis when the institutional environment is weak. Financial liberalization tends to make banking crises especially likely in countries where the rule of law is weak, corruption is widespread, the bureaucracy is inefficient, and contract enforcement is ineffective. Thus there is clear evidence that financial liberalization increases financial fragility—especially in developing countries, where the institutions needed to support a well-functioning financial system are generally not well established.

To explore a possible channel through which liberalization may affect bank fragility, we use bank-level data to examine the correlation between variables proxying bank franchise values and the financial liberalization dummy variable. We find evidence that franchise values tend to be lower after financial markets have been liberalized, possibly because banks' monopolistic power has eroded. This finding suggests that the increased moral hazard often attributed to low bank franchise values may help explain why financial liberalization tends to make banking crises more likely (Caprio and Summers 1993; Hellman, Murdock, and Stiglitz 1998).

These findings raise the question of whether the many benefits of financial liberalization highlighted in the literature are offset by the costs in terms of greater vulnerability to banking crises. A rigorous answer to this complex question is beyond the scope of this article. Nonetheless, using our dataset we attempt to throw some light on one aspect of the issue—namely, the effect of financial liberalization and banking crises on financial development and growth. First, we show that financial development is positively correlated with output growth in our sample, confirming the results of King and Levine (1993). Second, we find that when there is no banking crisis, countries and time periods in which financial markets have been liberalized have higher financial development than countries and time periods in which markets are controlled. But in countries and time periods with both financial liber-

alization and a banking crisis the level of financial development is about the same as in countries and time periods with neither, so the net effect on growth through financial development is not significantly different from zero.

To explore this issue further, we split the sample between countries that were financially repressed (defined as having negative real interest rates) at the time of liberalization and countries that were financially restrained (defined as having positive real interest rates). The same tests described above were performed for the two subsamples. For the restrained group the results resemble those for the complete sample. For the repressed group financial liberalization is accompanied by higher financial development even if a banking crisis also takes place. These findings suggest that financial liberalization is likely to have a positive effect on growth through financial development in countries with a prior policy of financial repression, even if it increases financial fragility.

Theoretical Basis for Vulnerability to Banking Crises

To put the empirical results in perspective, it is useful to review some of the theoretical reasons why a liberalized banking system might be more vulnerable to crisis. In tightly controlled financial systems, bank lending rates are usually subject to ceilings, which make it impossible for banks to charge high-risk premiums. As a result loans to high-risk customers cannot be profitable. Once ceilings are lifted during financial liberalization, banks can finance riskier ventures in exchange for higher promised returns. Indeed, one of the benefits of financial liberalization is that socially desirable high-risk, high-return projects will find the necessary financing.⁴

If loan-specific risk is hedged with a well-diversified portfolio, riskier loans need not increase the risk of bank insolvency or, at an aggregate level, the risk of a systemic banking crisis. But even well-diversified portfolios with risky loans are vulnerable to economywide adverse shocks (such as a recession). Moreover, managing the risk of a bank loan portfolio is a complex task, and bank staff trained in a tightly regulated financial system may lack the necessary skills and experience. Evaluating risky investment projects and monitoring the borrower during the life of the loan also require skills that may be in short supply in a banking system where lending to government and collateral-based private lending were the primary activities for many years. Such skills may also be difficult to import from abroad.

Nominal interest rates are likely to be more variable in a liberalized financial system where interest rates are market-determined than in a controlled system (although real rates may not be).⁵ Since one of the functions of banks is to transform short-term liabilities (deposits) into long-term assets (business and consumer loans), banks may become more vulnerable in an environment where interest rates are more volatile. When liberalization takes place before a well-functioning interbank market develops, banks may also find it difficult to deal with temporary liquidity shortages unless the central bank is ready to step in. Liquidity problems at one bank may spread to others and cause a panic when agents are imperfectly informed (Chari and Jagannathan 1988).

Because liberalization increases the opportunity for banks to take on risk, any mechanism that discourages bank managers from appropriately evaluating the downside risk of their lending decisions becomes especially dangerous. Limited liability is one such mechanism. Implicit or explicit government guarantees to depositors or other bank claimholders make moral hazard even more dangerous. Another factor that may contribute to moral hazard is the erosion of bank franchise value once ceilings are lifted on deposit interest rates and entry barriers are reduced (Caprio and Summers 1993; Hellmann, Murdock, and Stiglitz 1998). Because increased bank competition causes monopolistic profits to disappear, the costs of losing a banking license when the bank becomes insolvent decrease and incentives to take on riskier loans increase. Unless these perverse incentives are controlled through effective prudential regulation and supervision, increased risk taking due to moral hazard can become a powerful source of financial fragility, as demonstrated in numerous banking crises.

Many countries have combined financial liberalization with the reduction or removal of controls on international capital movements. This approach opens the door for newly liberalized financial intermediaries to take on yet another type of risk, foreign exchange risk, by raising foreign currency funds on international markets and lending them to local borrowers. Prudential limits on foreign currency exposure have often been circumvented, or currency risk transformed into credit risk by lending in foreign currency to unhedged domestic borrowers. Not surprisingly, currency crises have often preceded or accompanied banking crises (Kaminsky and Reinhart 1996).

Thus by giving banks and other financial intermediaries more freedom of action, financial liberalization increases opportunities to take on risk. Higher risk tends to increase financial fragility, but it is not necessarily bad for the economy because high-risk, high-return investments may outnumber low-risk, low-return ventures. Still, limited liability and other forms of implicit and explicit guarantees can make bankers' appetite for risk far greater than is socially desirable. If prudential regulation and supervision are ineffective at controlling bank behavior and realigning incentives, liberalization may increase financial fragility beyond socially desirable limits. Banks in newly liberalized systems are likely to be particularly vulnerable because the skills needed to screen and monitor risky borrowers, manage risky loan portfolios, and perform efficient supervision can be acquired only gradually through learning by doing. Other things being equal, the risk of bank insolvency and, more generally, of systemic banking crises may thus be greater in liberalized financial systems. The following sections present the results of econometric tests of various aspects of this linkage.

Data and Methodology

In setting up the panel, we began with all the countries covered by the International Monetary Fund's (IMF) *International Financial Statistics*, excluding only centrally planned and transition economies. To obtain a sufficiently large number of time

series, the study covers 1980–95. This period includes a substantial number of banking crises and episodes of financial liberalization, so the dataset is sufficiently rich for the purposes of our investigation.⁶

Some countries had to be eliminated because of missing data or because we could not find sufficient information on financial liberalization. A few countries were left out because their banking systems were in chronic distress during the period under consideration, making it impossible to pinpoint a specific subperiod as a banking crisis period. Two countries (Argentina and Bolivia) were excluded because they were outliers for two of the regressors that we used (inflation and the real interest rate).⁷ This process of elimination left 53 countries in the baseline specification (table 1).

A Multivariate Logit Model

To gauge the effect of financial liberalization on financial fragility, we estimated the probability of a banking crisis using a multivariate logit model and tested the hypothesis that a dummy variable capturing whether the financial system is liberalized significantly increases the probability of a crisis when other factors are controlled for. Accordingly, our dependent variable, the banking crisis dummy, equals 0 if there is no banking crisis and 1 if there is a crisis. The probability that a crisis will occur at a particular time in a particular country is hypothesized to be a function of a vector of n variables $\mathbf{X}(i, t)$, including the financial liberalization dummy variable and $n-1$ control variables. Let $P(i, t)$ denote a dummy variable that takes the value of 1 when a banking crisis occurs in country i and time t and a value of 0 otherwise. β is a vector of n unknown coefficients and $F[\beta'\mathbf{X}(i, t)]$ is the cumulative probability distribution function evaluated at $\beta'\mathbf{X}(i, t)$. Then the log likelihood function of the model is:

$$\ln L = \sum_{t=1}^T \sum_{i=1}^n (P(i, t) \ln \{F[\beta'\mathbf{X}(i, t)]\} + [1 - P(i, t)] \ln \{1 - F[\beta'\mathbf{X}(i, t)]\}).$$

To model the probability distribution function F we used the logistic functional form. Thus the estimated coefficients do not indicate the increase in the probability of a crisis given a one-unit increase in the corresponding explanatory variables as in standard linear regression models. Instead, the coefficients capture the effect of a change in an explanatory variable on $\ln [P(i, t)/(1 - P(i, t))]$. Thus while the sign of the coefficient does indicate the direction of the change, the magnitude depends on the slope of the cumulative distribution function at $\beta'\mathbf{X}(i, t)$.

After the onset of a banking crisis the behavior of some of the explanatory variables is likely to be affected by the crisis itself. Because these feedback effects would muddle the estimation, years in which banking crises were under way were eliminated from the panel (alternative approaches are discussed in the section below on sensitivity analysis). The probability of a crisis occurring in a country that has experienced a banking crisis is likely to differ from that for a country that has never experienced one. To take this difference into account, we included different additional regressors in the estimated equations, such as duration of last crisis, time since last crisis, and number of past crises.

Table 1. Years of Interest Rate Liberalization and Banking Crises, 1980–95, Various Countries

<i>Country</i>	<i>Interest rate liberalization</i>	<i>Banking crisis</i>
Australia	1981–95	
Austria	1980–95	
Belgium	1986–95	
Canada	1980–95	
Chile	1980–95	1981–87
Colombia	1980–95	1982–85
Congo, Dem. Rep ^a	1980–95	
Denmark	1981–95	
Ecuador	1986–87, 1992–95	
Egypt	1991–95	
El Salvador	1991–95	1989
Finland	1986–95	1991–94
France	1980–95	
Germany	1980–95	
Greece	1980–95	
Guatemala	1989–95	
Guyana	1991–95	1993–95
Honduras	1990–95	
India	1991–95	1991–94
Indonesia	1983–95	1992–94
Ireland	1985–95	
Israel	1990–95	1983–84
Italy	1980–95	1990–94
Jamaica	1991–95	
Japan	1985–95	1992–94
Jordan	1988–95	1989–90
Kenya	1991–95	1993
Korea, Rep. of	1984–88, 1991–95	
Malaysia	1980–95	1985–88
Mali		1987–89
Mexico	1989–95	1982, 1994–95
Netherlands	1980–95	
New Zealand	1980, 1984–95	
Nigeria	1990–93	1991–95
Norway	1985–95	1987–93
Papua New Guinea	1980–95	1989–95
Paraguay	1990–95	1995
Peru	1980–84, 1990–95	1983–90
Philippines	1981–95	1981–87
Portugal	1994–95	1986–89
Sri Lanka	1980–95	1989–93
Sweden	1980–95	1990–93
Switzerland	1989–95	
Syria		
Tanzania	1993–95	1988–95

(Table continues on next page)

Table 1. Years of Interest Rate Liberalization and Banking Crises, 1980–95, Various Countries (continued)

Country	Interest rate liberalization	Banking crisis
Thailand	1989–95	1983–87
Togo	1993–95	
Turkey ^a	1980–82, 1984–95	1991, 1994–95
Uganda ^a	1991–95	
United States	1980–95	1980–92
Uruguay	1980–95	1981–85
Venezuela	1989–95	1993–95
Zambia ^a	1992–95	

a Banking crises other than those indicated occurred between 1980 and 1995 but were not included in the panel because of missing data.

Source: Caprio and Klingebiel 1996; Lindgren, Garcia, and Saal 1996

The Banking Crisis Variable

In constructing a banking crisis dummy variable, we identified and dated episodes of banking sector distress during 1980–95 using primarily two recent studies, Caprio and Klingebiel (1996) and Lindgren, Garcia, and Saal (1996). For an episode of distress to qualify as a full-fledged crisis, we established—somewhat arbitrarily—that at least one of the following conditions must apply: nonperforming assets represented at least 10 percent of banking system assets, the cost of the rescue operation was at least 2 percent of GDP, banking problems resulted in a large-scale nationalization of banks, or extensive bank runs occurred or emergency measures (such as deposit freezes, prolonged bank holidays, or generalized deposit guarantees) were enacted by the government in response to the crisis. (Below we explore the sensitivity of the results to the definition of a crisis.) For the length of the crisis, we relied solely on the dates provided in the case studies (see table 1).

The Financial Liberalization Variable

Empirical studies of financial liberalization often use the real interest rate as a proxy for financial liberalization (Fry 1997; Bandiera and others 1997). Especially when measured ex post, however, real interest rates are likely to be affected by a variety of factors that have little to do with changes in the regulatory framework of financial markets. This problem may be limited in a cross-country study in which interest rates are averaged over long periods. But in a panel study like ours with an important time-series dimension, using the real interest rate as a proxy for financial liberalization could be misleading. A positive correlation between real interest rates and the probability of a banking crisis, for instance, may simply reflect the fact that both variables tend to be high during cyclical economic downturns. Financial liberalization may play no role.

To avoid this problem, we constructed a financial liberalization variable based on observed policy changes. This strategy is not without its difficulties, however. No

available database records such policy changes, forcing us to rely on case studies, IMF country reports, and other sources. Furthermore, financial liberalization has taken many forms. Not all countries have eliminated the same restrictions in the same order. Some countries (Greece, Japan) have opted for a gradual approach, while others (Egypt, Mexico) have switched regimes quite rapidly. And temporary reversals have occurred in some countries.

Because the removal of interest rate controls was the centerpiece of liberalization in most countries, we chose this policy change as the indicator of financial liberalization. This approach left us with the need to identify a beginning year in countries where the process was gradual. Lacking a good theoretical ground for preferring one option over another, we chose as the beginning year the first year when some interest rates were liberalized, because it was easiest to identify. For some countries we identified two years because liberalization was temporarily reversed (see table 1).

Control Variables

The control variables are taken from our previous study of banking crises (Demirgüç-Kunt and Detragiache 1997) and reflect both the theory of the determinants of banking crises and data availability.⁸ (See the appendix for a list of the variables and their sources.) The first group of control variables captures macroeconomic developments that affect bank performance, especially through the level of nonperforming loans. This group includes the growth rate of real GDP, the change in terms of trade, and inflation. The real short-term interest rate is also introduced as a control variable because, regardless of whether financial markets are liberalized, banking problems are more likely if real interest rates are high.⁹

The second set of control variables includes characteristics of the banking system, such as vulnerability to sudden capital outflows (measured by the ratio of M2 to foreign exchange reserves, as suggested by Calvo 1996), exposure to the private sector (measured by the ratio of loans to the private sector to total loans), liquidity (measured by the ratio of bank cash and reserves to bank assets), and lagged credit growth. Lagged credit growth was introduced because high rates of credit expansion may finance an asset price bubble that can burst and cause a banking crisis. Finally, per capita GDP was used to control for a country's level of development.

Measures of Institutional Quality

Because the quality of institutions may affect the extent to which financial liberalization increases the probability of a banking crisis, in alternative specifications we interacted proxies of institutional quality with the liberalization dummy variables and introduced the interaction term as a separate variable in the regression. We experimented with six measures of institutional quality: per capita GDP and indexes of the respect for the rule of law ("law and order"), the extent of bureaucratic delay, the quality of contract enforcement, the quality of the bureau-

cracy, and the degree of corruption. These indexes increase with the quality of institutions.

Empirical Results

The results of the logit regressions estimating the probability of a banking crisis as a function of the financial liberalization dummy variable and the set of control variables are presented in table 2, along with the usual diagnostic tests to assess the fit of the model.¹⁰

Effects of Control and Financial Liberalization Variables

The baseline specification fits the data well and correctly classifies 77 percent of the observations. The macroeconomic control variables are all significant at least at the 5 percent level and have the expected signs: banking crises tend to be associated with low GDP growth, adverse terms of trade changes, high real interest rates, and high inflation. Of the characteristics of the banking sector, vulnerability to a speculative attack against the currency is significant at the 1 percent level, while credit growth lagged by two periods is significant at the 10 percent level. The other variables are not significant. Per capita GDP is significantly and negatively correlated with the probability of a banking crisis, suggesting that, other things being equal, developing countries are more vulnerable to such events.

The financial liberalization dummy variable is strongly and positively correlated with the probability of a banking crisis, a result that holds regardless of the treatment of reversals (see regression 2 in table 2). These results suggest that financial liberalization is a significant factor leading to banking sector fragility and that this effect is at work even after variables capturing the state of the macroeconomy (including the level of the risk-free short-term real interest rate) are controlled for. This suggests that financial liberalization increases financial fragility even if it is carried out after macroeconomic stabilization is achieved, as McKinnon (1993) recommends.

An important question is whether the effect of liberalization on the probability of a crisis tends to be a transitional effect—that is, one that manifests itself only in the years immediately following the change in policy. To test this hypothesis, we estimated the baseline regression using a liberalization dummy that takes the value of 1 only in the first three, four, five, or six years after liberalization, as opposed to the entire period following the policy change (see regressions 3–6 in table 2). The redefined dummies are all less significant than in the baseline, and the overall fit of the model does not improve. In fact, the dummy corresponding to a transition of only three years is not significant, and that corresponding to a transition of four years is significant only at the 10 percent level. Thus the effect of financial liberalization on banking fragility does not appear in the immediate aftermath of the change in policy but rather emerges only over time. This result may also be due to the fact that we chose the beginning of deregulation as the date of the policy change even though in a number of countries interest rate deregulation was gradual.

Table 2. Financial Liberalization and Banking Crises

Variable	Regression						
	1	2	3	4	5	6	7
<i>Control variables</i>							
GDP growth rate	-0.168*** (0.040)	-0.164*** (0.039)	-0.163*** (0.039)	-0.162*** (0.039)	-0.167*** (0.039)	-0.168*** (0.039)	-0.191*** (0.044)
Change in terms of trade	-0.052** (0.023)	-0.050** (0.022)	-0.043** (0.020)	-0.043** (0.020)	-0.049** (0.022)	-0.049** (0.022)	-0.050** (0.025)
Real interest rate	0.047*** (0.015)	0.046*** (0.015)	0.048*** (0.015)	0.050*** (0.015)	0.051*** (0.015)	0.050*** (0.015)	0.044*** (0.015)
Inflation	0.027*** (0.009)	0.027*** (0.008)	0.027*** (0.009)	0.027*** (0.009)	0.027*** (0.009)	0.028*** (0.009)	0.022** (0.011)
M2/reserves	0.022*** (0.007)	0.021*** (0.007)	0.016*** (0.007)	0.017*** (0.007)	0.017*** (0.007)	0.017** (0.007)	0.024*** (0.007)
Private credit/GDP	0.007 (0.012)	0.007 (0.013)	0.006 (0.012)	0.006 (0.012)	0.006 (0.012)	0.006 (0.012)	0.013 (0.013)
Liquid reserves/ bank assets	-0.018 (0.014)	-0.019 (0.014)	-0.020 (0.014)	-0.020 (0.014)	-0.021 (0.014)	-0.020 (0.014)	-0.022 (0.016)
Credit growth _{t-2}	0.023* (0.013)	0.022* (0.013)	0.023* (0.013)	0.023* (0.013)	0.023* (0.013)	0.023* (0.013)	0.013 (0.014)
Per capita GDP	-0.108** (0.051)	-0.103** (0.051)	-0.078* (0.051)	-0.077* (0.051)	-0.079* (0.051)	-0.080* (0.051)	-0.101* (0.057)
<i>Financial liberalization variables</i>							
Financial liberalization		1.761*** (0.634)					1.449** (0.712)
Financial liberalization (R)		1.423*** (0.589)					
Financial liberalization (3)			0.488 (0.434)				
Financial liberalization (4)				0.639* (0.415)			
Financial liberalization (5)					0.892** (0.415)		
Financial liberalization (6)						0.811** (0.418)	
Financial liberalization x initial interest rate							-0.026 (0.020)

(Table continues on next page)

Another interesting question is whether the effects of financial liberalization on financial fragility differ between countries that were severely repressed at the time of liberalization and countries that were only financially restrained. To explore this issue, we interacted the financial liberalization dummy variable with the average real interest rate in the three years before liberalization and introduced this interaction term as an

Table 2. Financial Liberalization and Banking Crises (continued)

Variable	Regression						
	1	2	3	4	5	6	7
<i>Past crisis variables</i>							
Duration of last crisis	0.108** (0.051)	0.115** (0.051)	0.139*** (0.051)	0.147*** (0.050)	0.139*** (0.050)	0.140*** (0.051)	0.130** (0.062)
Number of past crises	32	32	31	32	32	32	26
Number of observations	639	639	602	639	632	632	525
Share correct (percent)	77	77	77	76	76	77	78
Share of crisis countries correct (percent)	63	63	68	59	59	56	62
Model χ^2	61.42***	58.79***	52.52***	54.49***	57.32***	56.48***	55.95***
AIC	217	219	218	224	219	221	177

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level.

Note The regressions correspond to different definitions of the financial liberalization dummy. In regression 1, the baseline specification, the dummy is 0 for periods when interest rates are subject to controls and 1 when liberalization begins. The dummy remains 1 even if liberalization is temporarily reversed, under the assumption that the effects of liberalization persist even through short reversals. In regression 2 the dummy variable is modified by treating periods of reversal as 0s. For regressions 3–6 the liberalization dummy takes a value of 1 only in the first three, four, five, or six years after liberalization. Regression 7 introduces, as an additional regressor, an interaction term capturing the interaction between financial liberalization and the average real interest rate in the three years before liberalization. Numbers in parentheses are t-statistics. See the appendix for a description of the variables and their sources.

Source See appendix

additional regressor. A negative and significant coefficient for the new variable would suggest that fragility is less severely affected by liberalization in countries that were more financially repressed at the beginning of liberalization. The estimated coefficient is negative, but it is not significantly different from zero (see regression 7 in table 2).

To illustrate the magnitude of the effect of financial liberalization on financial fragility, according to our empirical model, we estimated the probability of a crisis using the baseline model for the 26 crisis episodes that took place in liberalized regimes (table 3). We also recalculated the probability of a crisis had the country not liberalized by setting the liberalization dummy equal to 0 (see the fourth column in table 3). For all countries the predicted probability of a banking crisis falls substantially. Of the 20 episodes that were correctly classified as crises, 11 would not have been crises in the absence of financial liberalization. Thus the effect of financial liberalization on the probability of a banking crisis is both statistically significant and of a nontrivial magnitude.

The Role of the Institutional Environment

Theory suggests that the adverse effect of financial liberalization on banking sector fragility is stronger when the institutions needed for the correct functioning of financial markets are weak. To test whether this effect is supported by the data, we added to the baseline regression various alternative variables in the form of interaction terms between the liberalization dummy and proxies of the quality of the insti-

Table 3. Predicted Effects of Financial Liberalization on Financial Fragility

<i>Country</i>	<i>First year of bank crisis</i>	<i>Probability of crisis predicted by baseline on crisis date^a</i>	<i>Predicted probability of crisis had country not liberalized on or before bank crisis date</i>
Chile	1981	0.174	0.035
Colombia	1982	0.047	0.008
Finland	1991	0.119	0.023
Guyana	1993	0.028	0.005
India	1991	0.221	0.047
Indonesia	1992	0.306	0.071
Italy	1990	0.028	0.005
Japan	1992	0.071	0.012
Jordan	1989	0.786	0.387
Kenya	1993	0.412	0.108
Malaysia	1985	0.170	0.034
Mexico	1994	0.207	0.043
Nigeria	1991	0.044	0.008
Norway	1987	0.031	0.006
Papua New Guinea	1989	0.259	0.057
Paraguay	1995	0.114	0.022
Peru	1983	0.347	0.084
Philippines	1981	0.052	0.009
Portugal	1986	0.133	0.026
Sri Lanka	1989	0.104	0.019
Sweden	1990	0.033	0.006
Turkey	1991	0.221	0.047
	1994	0.443	0.121
United States	1980	0.459	0.126
Uruguay	1981	0.358	0.087
Venezuela	1993	0.424	0.113

Note Probabilities are not reported for El Salvador, Israel, Mali, Mexico in 1982, Tanzania, and Thailand because these countries had not liberalized before the banking crisis

a Countries in the baseline specification are classified as crisis cases if the predicted probability of a crisis is greater than 0.05, which is equal to the ratio of the number of crisis observations to the total number of observations

Source: See appendix

tutional environment (table 4). Negative and significant coefficients for the interaction variables mean that a better institutional environment tends to weaken the effect of financial liberalization on the probability of a banking crisis.

The first proxy for the institutional environment is per capita GDP, which was also used as a control variable in the baseline regression. The other proxies, as noted above, are indexes of the respect for rule of law ("law and order"), the extent of bureaucratic delay, the quality of contract enforcement, the quality of the bureaucracy, and the degree of lack of corruption. The indexes measuring law and order, the quality of the bureaucracy, and corruption range from 0 to 6, while the indexes of bureaucratic delay and contract enforcement range from 0 to 4. The proxies measure factors that affect the extent to which laws and regulations are enforced, not the quality of the laws and regulations in a particular country.

Table 4. Importance of the Institutional Environment

Variable	Regression					
	1	2	3	4	5	6
<i>Control variables</i>						
GDP growth rate	-0.171*** (0.040)	-0.214*** (0.054)	-0.233*** (0.072)	-0.238*** (0.070)	-0.219*** (0.054)	-0.223*** (0.054)
Change in terms of trade	-0.054** (0.023)	-0.040* (0.027)	-0.056* (0.034)	-0.060* (0.033)	-0.042* (0.026)	-0.040* (0.026)
Real interest rate	0.045*** (0.015)	0.052** (0.024)	0.053** (0.021)	0.050*** (0.021)	0.049** (0.024)	0.049** (0.023)
Inflation	0.026*** (0.009)	0.027* (0.015)	0.022* (0.013)	0.020* (0.013)	0.021 (0.015)	0.022 (0.015)
M2/reserves	0.022*** (0.007)	0.018* (0.010)	0.025** (0.012)	0.025** (0.012)	0.022** (0.010)	0.019** (0.010)
Private credit/GDP	0.002 (0.011)	-0.003 (0.011)	0.005 (0.012)	0.006 (0.012)	-0.003 (0.011)	-0.003 (0.011)
Liquid reserves/bank assets	-0.018 (0.014)	-0.030 (0.023)	0.020 (0.026)	0.015 (0.026)	-0.030 (0.022)	-0.027 (0.021)
Credit growth $t-2$	0.024* (0.013)	0.013 (0.018)	0.045*** (0.017)	0.043*** (0.016)	0.011 (0.018)	0.009 (0.018)
<i>Financial liberalization and institutional variables</i>						
Financial liberalization	1.956*** (0.657)	1.770* (0.986)	4.053*** (1.542)	4.732*** (1.557)	1.803* (1.082)	1.323* (1.030)
Financial liberalization x per capita GDP	-0.089*(6%) (0.048)					
Financial liberalization x law and order		-0.405** (0.205)				
Financial liberalization x bureaucratic delay			-0.727 (0.678)			
Financial liberalization x contract enforcement				-0.938* (0.574)		
Financial liberalization x quality of bureaucracy					-0.380* (0.223)	
Financial liberalization x corruption						-0.403*(6%) (0.215)
<i>Past crisis variables</i>						
Duration of last crisis	0.112** (0.051)	0.181** (0.081)	0.028 (0.067)	0.131 (0.067)	0.171** (0.079)	0.156** (0.078)
Number of past crises	32	22	21	21	22	22
Number of observations	639	425	406	406	418	418
Share correct (percent)	77	72	78	80	72	73
Share of crisis countries correct (percent)	63	55	67	71	59	59
Model χ^2	60.08***	35.69***	49.65***	51.34***	34.16***	34.77***
AIC	218	161	140	138	162	162

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level.

Note Numbers in parentheses are t-statistics. See the appendix for a description of the variables and their sources
Source. See appendix

All six interaction variables have the expected negative sign, and all except the index of bureaucratic delay are significant at least at the 10 percent level. Law and order, per capita GDP, and corruption have the highest significance levels, and the size of the effect is not trivial. For a country with a law and order index of 0 (the lowest score), for example, the net impact of financial liberalization on the probability of crisis is 1.770. This impact falls to 0.555 when the law and order index rises to 3, and it becomes negative when the index rises to 6, suggesting that financial liberalization tends to make banking crises less likely. Similarly, an increase in the contract enforcement index from 0 to 4 reduces the impact of liberalization on the probability of crisis from 4.732 to 0.980. These results suggest that improving the quality of the institutional environment, and especially reducing corruption and strengthening the rule of law, can curb the tendency of liberalized financial markets to harbor systemic banking crises.

Sensitivity Analysis

A number of robustness tests were performed on the baseline regression. These tests confirmed that the relationship between financial liberalization and banking sector fragility appears to be robust to various changes in the specification of the logit regression.

The first test concerns the treatment of years during which the crisis is under way. These years were omitted from the baseline specification because including them requires accurate information on the year in which a crisis ended. Because the end of a crisis may be difficult to determine, we also estimated the baseline regression using three alternative panels: one that omits all years following a crisis, one that treats all crisis years as 1s, and one that treats all crisis years (except the first year) as 0s. The results show that while there are some changes in the coefficients and standard errors of the control variables, the liberalization dummy remains strongly significant in all specifications (table 5).

A second set of sensitivity tests uses different definitions of a banking crisis relative to the baseline (table 6). Under the more stringent definition, nonperforming loans must represent at least 15 percent of total loans or the cost of the crisis must equal at least 3 percent of GDP. Under the less stringent definition, nonperforming loans must represent at least 5 percent of total loans or the cost of the crisis must equal at least 1 percent of GDP. Changes in the control variables are minor, and the liberalization dummy remains significant, although only at the 10 percent level.

A third methodological issue is whether to include country (time) fixed effects to allow for the possibility that the dependent variable may vary across countries (years) independently of the explanatory variables included in the regression. In logit estimation the inclusion of fixed effects requires that countries (years) in which there was no crisis during the period under consideration be excluded from the panel (Greene 1997). Doing so excludes a large amount of information. For this reason we omitted fixed effects from the baseline and estimated a model with fixed effects as part of the sensitivity analysis (second and third columns in table 7). For

Table 5. Sensitivity Analysis: Different Treatment of Crisis Years

Variable	Baseline	No years after first crisis	Years of crisis = 1	Years of crisis = 0
<i>Control variables</i>				
GDP growth rate	-0.168*** (0.040)	-0.136*** (0.041)	-0.067*** (0.023)	-0.137*** (0.036)
Change in terms of trade	-0.052** (0.023)	-0.043** (0.023)	-0.014 (0.014)	-0.047** (0.021)
Real interest rate	0.047*** (0.015)	0.046*** (0.017)	0.016*** (0.007)	0.013** (0.006)
Inflation	0.027*** (0.009)	0.025*** (0.010)	0.016*** (0.005)	0.004 (0.005)
M2/reserves	0.022*** (0.007)	0.017*** (0.007)	0.017*** (0.004)	0.008* (0.005)
Private credit/GDP	0.007 (0.012)	0.015 (0.012)	0.011** (0.005)	-0.003 (0.009)
Liquid reserves/bank assets	-0.018 (0.014)	-0.007 (0.014)	-0.016** (0.008)	-0.005 (0.012)
Credit growth _{t-2}	0.023* (0.013)	0.018 (0.014)	0.002 (0.008)	0.019* (0.012)
Per capita GDP	-0.108** (0.051)	-0.134*** (0.052)	-0.091*** (0.022)	-0.080** (0.041)
<i>Financial liberalization variable</i>				
Financial liberalization	1.761*** (0.634)	2.154*** (0.618)	2.187*** (0.343)	1.178** (0.557)
<i>Past crisis variables</i>				
Duration of last crisis	0.108** (0.051)		-0.133*** (0.030)	0.144*** (0.049)
Number of past crises	32	29	128	32
Number of observations	639	531	735	735
Share correct (percent)	77	77	72	73
Share of crisis countries correct (percent)	63	66	69	59
Model χ^2	61.42***	50.50***	141.82***	42.67***
AIC	217	197	562	245

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level

Note Numbers in parentheses are t-statistics See the appendix for a description of the variables and their sources
Source See appendix

both country and time fixed effects the hypothesis that the coefficients of the country and time dummies are jointly significantly different from 0 is rejected, suggesting that there are no fixed effects. In any case, the liberalization dummy is still positively and significantly correlated with the probability of a crisis.

Another sensitivity test involves using lagged values of the explanatory variables to reduce the risk that the regressors may not be exogenous determinants of a crisis (see the fourth column in table 7). The drawback of using lagged values on the right-hand side is that if the macroeconomic shocks that trigger the crisis work relatively quickly, their effect would not be evident a year before the crisis erupts. In this regression most

Table 6. Sensitivity Analysis: Different Crisis Definitions

<i>Variable</i>	<i>Baseline</i>	<i>More stringent definition</i>	<i>Less stringent definition</i>
<i>Control variables</i>			
GDP growth rate	-0.168*** (0.040)	-0.126*** (0.044)	-0.160*** (0.039)
Change in terms of trade	-0.052** (0.023)	-0.04** (0.023)	-0.045** (0.022)
Real interest rate	0.047*** (0.015)	0.067*** (0.023)	0.044*** (0.014)
Inflation	0.027*** (0.009)	0.032*** (0.012)	0.025*** (0.009)
M2/reserves	0.022*** (0.007)	0.009 (0.007)	0.020*** (0.007)
Private credit/GDP	0.007 (0.012)	-0.003 (0.017)	0.001 (0.011)
Liquid reserves/bank assets	-0.018 (0.014)	-0.017 (0.018)	-0.021 (0.015)
Credit growth _{t-2}	0.023* (0.013)	0.022 (0.015)	0.027** (0.013)
Per capita GDP	-0.108** (0.051)	-0.150** (0.071)	-0.069* (0.044)
<i>Financial liberalization variable</i>			
Financial liberalization	1.761*** (0.634)	1.098* (0.692)	1.732*** (0.607)
<i>Past crisis variables</i>			
Duration of last crisis	0.108** (0.051)	0.106* (0.059)	0.109** (0.047)
Number of past crises	32	24	6
Number of observations	639	639	623
Share correct (percent)	77	78	74
Share of crisis countries correct (percent)	63	58	61
Model χ^2	61.42***	52.88***	59.73***
AIC	217	176	239

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level

Note. Numbers in parentheses are t-statistics. See the appendix for a description of the variables and their sources
Source: See appendix

macroeconomic control variables (except the real interest rate) lose significance, while the other controls remain significant. More interestingly, the liberalization dummy continues to be positively and significantly correlated to the probability of a crisis.

Financial Liberalization and Bank Franchise Values

The results of the previous sections suggest that liberalization increases the fragility of the financial system. One reason why financial liberalization may

Table 7. Sensitivity Analysis: Country and Time Fixed Effects and Lagged Explanatory Variables

<i>Variable</i>	<i>Baseline</i>	<i>Country fixed effects</i>	<i>Time fixed effects</i>	<i>Lagged explanatory variables</i>
<i>Control variables</i>				
GDP growth rate	-0.168*** (0.040)	-0.246*** (0.060)	-0.177*** (0.047)	0.057 (0.044)
Change in terms of trade	-0.052** (0.023)	-0.054* (0.031)	-0.044* (0.026)	-0.004 (0.022)
Real interest rate	0.047*** (0.015)	0.122*** (0.042)	0.049*** (0.015)	0.007* (0.004)
Inflation	0.027*** (0.009)	0.064*** (0.027)	0.028*** (0.009)	0.004 (0.003)
M2/reserves	0.022*** (0.007)	0.026** (0.012)	0.024*** (0.007)	0.007** (0.003)
Private credit/GDP	0.007 (0.012)	-0.011 (0.039)	0.012 (0.014)	-0.001 (0.012)
Liquid reserves/bank assets	-0.018 (0.014)	0.002 (0.024)	-0.016 (0.015)	-0.002 (0.009)
Credit growth _{t-2}	0.023* (0.013)	0.032* (0.021)	0.024* (0.014)	0.019* (0.012)
Per capita GDP	-0.108** (0.051)	-0.402 (0.423)	-0.138*** (0.056)	-0.077* (0.046)
<i>Financial liberalization variable</i>				
Financial liberalization	1.761*** (0.634)	1.962* (1.196)	2.077*** (0.702)	1.113** (0.555)
<i>Past crisis variables</i>				
Duration of last crisis	0.108** (0.051)	0.501*** (0.132)	0.229** (0.113)	0.073 (0.049)
Number of past crises	32	32	32	31
Number of observations	639	33	565	605
Share correct (percent)	77	75	76	67
Share of crisis countries correct (percent)	63	44	53	58
Model χ^2	61.42***	81.85***	66.39***	22.44***
AIC	217	210	235	246

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level

Note Coefficients for the country and time dummy variables are not reported. Numbers in parentheses are t-statistics. See the appendix for a description of the variables and their sources.

Source See appendix.

increase fragility is that the removal of interest rate ceilings or the reduction of entry barriers reduces bank franchise values, exacerbating moral hazard. As Caprio and Summers (1993) and Hellmann, Murdock, and Stiglitz (1998) note, interest rate ceilings and entry restrictions create rents that increase the value of a banking license. The risk of losing this valuable license induces banks to become more stable by introducing better incentives to monitor the firms they finance and manage the risk of their loan portfolios. Financial liberalization increases bank

competition and lowers profits, eroding franchise values and distorting risk-taking incentives. Unless liberalization is accompanied by the adequate strengthening of prudential regulations and supervision, lower franchise values will likely increase fragility.¹¹

This section uses bank-level data from Fitch IBCA's BankScope database to investigate whether there is empirical evidence that bank franchise values fall with financial liberalization. The dataset includes bank-level accounting data for 80 countries covering 1988–95. In most countries the banks covered by the database account for at least 90 percent of the banking system. For each bank we constructed three profitability measures: net interest margin, after-tax return on assets, and after-tax return on equity.

Because none of these measures is a perfect indicator of future profitability, we also looked at additional balance sheet ratios that may be associated with a drop in franchise value: a measure of capital adequacy (the ratio of the book value of equity to total assets), a measure of liquidity (the ratio of liquid assets to total assets), and the ratio of deposits to total liabilities. These ratios are country averages of bank-level figures. High capitalization and high liquidity should have an adverse effect on franchise value, because they decrease the volume of loans that a bank can extend for a given volume of deposits.¹²

We also examined the behavior of an indicator of market concentration (the ratio of assets of the three largest banks to total banking assets) and an indicator of foreign bank penetration (the ratio of foreign bank assets to total bank assets). High market concentration and low foreign bank penetration should be associated with more monopolistic powers for domestic banks and, therefore, with higher franchise values.

The correlations of these banking variables with the financial liberalization dummy variable are reported in table 8. Although simple correlations do not imply causality, these results can at least indicate whether the hypothesis that financial liberalization leads to lower bank franchise values should be dismissed out of hand or taken seriously. The correlations in the first column of the table are calculated using a dummy variable that is equal to 1 in all periods in which the financial market is liberalized and equal to 0 otherwise. In the remaining columns the liberalization dummy is redefined to take a value of 1 during the transition to a liberalized system, with the transition taken to last three, four, five, or six years, and 0 otherwise. By comparing these correlations we can see the extent to which a drop in bank franchise value, if there is one, is a temporary or permanent effect of liberalization.

The results in the first column of table 8 indicate that liberalization leads to permanently lower bank profits as measured by the return on equity, while neither the net interest margin nor the return on assets are significantly correlated with the liberalization dummy. Financial liberalization appears to be correlated with higher capitalization (which should reduce bank profitability) and lower liquidity (which should have the opposite effect). The extent of long-term deposit mobilization does not appear to change significantly with liberalization. More interestingly, liberaliza-

Table 8. Correlation Coefficients between Financial Liberalization and Bank Franchise Value Indicators

Variable	Financial liberalization variables				
	Baseline	(3)	(4)	(5)	(6)
Net interest margin	0.024 <i>0.653</i>	0.175*** <i>0.001</i>	0.150*** <i>0.006</i>	0.157*** <i>0.004</i>	0.158*** <i>0.004</i>
Return on assets	0.088 <i>0.139</i>	0.202*** <i>0.001</i>	0.168*** <i>0.006</i>	0.167*** <i>0.006</i>	0.132** <i>0.030</i>
Return on equity	-0.118** <i>0.028</i>	0.120** <i>0.029</i>	0.097* <i>0.076</i>	0.077 <i>0.158</i>	0.068 <i>0.212</i>
Capital adequacy	0.207*** <i>0</i>	0.058 <i>0.289</i>	0.119** <i>0.028</i>	0.116** <i>0.032</i>	0.121** <i>0.026</i>
Liquidity	-0.155*** <i>0.004</i>	0.154*** <i>0.005</i>	0.184*** <i>0.001</i>	0.152*** <i>0.005</i>	0.168*** <i>0.002</i>
Deposits/liabilities	-0.033 <i>0.541</i>	0.069 <i>0.210</i>	0.161*** <i>0.003</i>	0.170*** <i>0.002</i>	0.121** <i>0.026</i>
Market concentration	-0.087 <i>0.137</i>	0.092 <i>0.121</i>	0.053 <i>0.377</i>	0.042 <i>0.476</i>	0.035 <i>0.552</i>
Foreign bank penetration	0.109** <i>0.062</i>	-0.012 <i>0.840</i>	0.015 <i>0.799</i>	0.020 <i>0.734</i>	0.031 <i>0.606</i>

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level

Note Net interest margin is interest income minus interest expenses divided by total assets. Return on assets is net profits divided by total assets. Return on equity is net profits divided by book value of equity. Capital adequacy is the book value of equity divided by total assets. Liquidity is the ratio of liquid assets to total assets. Deposits/liabilities is the ratio of deposits (customer and short-term funding) to total liabilities. Market concentration is the ratio of assets in the three largest banks to total bank assets. Foreign bank penetration is the ratio of foreign bank assets to total bank assets. All bank-level variables are average ratios for all banks in the BankScope database in a country in a given year. Numbers in italics are Pearson correlation coefficients.

Source: Fitch IBCA BankScope database

tion appears to be permanently associated with lower market concentration (albeit significant only at the 13 percent level) and higher foreign bank concentration. Both effects are consistent with lower bank franchise values because of the reduced monopolistic profits resulting from increased competition.

When we look at the correlations with the transition to a liberalized system, we see that bank margins, profits, capital, liquidity, and deposit mobilization are higher during the transition period. However, comparison with the correlations in the first column suggests that most of these effects do not persist in the long run. During the transition the coefficients for market concentration and foreign bank penetration are not significant, suggesting that the structure of the banking sector changes only slowly after liberalization begins.

These results are broadly consistent with theories that liberalization increases bank fragility through its negative impact on bank franchise values. The next logical step would be to test whether low bank franchise values are associated with increased bank fragility. Unfortunately, we are unable to pursue this question because the number of banking crises occurring during the period covered by the BankScope database is too small.

Financial Liberalization, Banking Crises, Financial Development, and Growth

We have established that financial liberalization increases financial fragility. Do these results imply that policymakers should abandon liberalization in favor of increased direct intervention in financial markets? The answer depends on whether the welfare costs of financial fragility exceed the welfare benefits of liberalization and whether governments can design and implement regulations that correct rather than reinforce market failures. An answer to these complex questions is beyond the scope of this article. Nonetheless, it is possible to use our dataset to explore one aspect of this issue—namely, whether financial liberalization and banking crises affect economic growth through their effect on financial development.

The focus on growth effects through financial development is suggested by the large body of literature documenting how financial development increases long-run growth rates (King and Levine 1993; Levine 1997). Presumably, one of the main benefits of financial liberalization is that it fosters financial development and, as a result, increases long-run growth. But the disruption caused by a systemic banking crisis is likely to have a direct adverse effect on financial development (at least in the short or medium term) and, accordingly, reduce growth. The question addressed in this section is whether these effects can be detected in our dataset and, if so, how the magnitude of the adverse effect of banking crises on financial development compares with that of the positive effect of financial liberalization.

To verify whether financial development tends to increase growth in our sample, we estimated growth regressions using a panel obtained by splitting the sample period (1980–94) into three subperiods of five years each. The regressors include a set of control variables and four alternative indicators of financial development proposed by King and Levine (1993): the ratio of liquid liabilities of the financial system to GDP (liquidity), the share of bank credit that goes to the private sector (private credit), the ratio of domestic bank assets to the sum of central bank domestic assets and domestic bank assets (bank assets), and the ratio of central bank domestic assets to GDP (central bank). (The control variables, which are also similar to those used by King and Levine 1993, are the logarithm of real per capita GDP, the logarithm of secondary school ratio at the beginning of the subperiod, the share of government consumption expenditure in GDP, the inflation rate, the ratio of the sum of imports and exports to GDP, the real interest rate, and a period dummy variable.) The first three indicators increase with financial development, while the fourth decreases.

The results of the growth regressions are reported in the top panel of table 9. Although the R^2 figures are low, two of four indicators (bank assets and central bank) have significant coefficients of the expected sign. Thus there is some evidence that financial development is positively correlated with growth in our panel.

To assess the impact of financial liberalization and banking crises on financial development, we then regressed each financial development indicator on a constant, the liberalization dummy, and the banking crisis dummy, using the same panel as in

the growth regressions.¹³ The estimated coefficients have a simple interpretation: the constant is the mean level of financial development for observations in which neither financial liberalization nor a banking crisis took place. The coefficient of the liberalization dummy indicates the difference between the level of financial development in countries or time periods in which financial liberalization but no banking crisis occurred, and the level of financial development in countries or time periods in which neither liberalization nor a banking crisis took place. Similarly, the coefficient of the banking crisis dummy, if significantly less than 0, would indicate that, on average, observations corresponding to banking crises are accompanied by lower financial development, provided no liberalization occurred. Finally, if the difference between the coefficients of the two dummies is significantly greater than 0, then a country or time period in which both financial liberalization and a banking crisis occurred has, on average, a higher level of financial development than a country or time period in which no crisis or liberalization took place.

Table 9. Growth, Financial Development, Financial Liberalization, and Banking Crises—Full Sample

	<i>Liquidity</i>	<i>Private credit</i>	<i>Bank assets</i>	<i>Central bank</i>
<i>Growth regressions^a</i>				
Financial development	-0.407 (0.765)	0.243 (1.007)	3.450** (1.633)	-2.010* (1.166)
Adjusted R ²	0.11	0.11	0.14	0.11
Number of observations	136	136	137	134
<i>Financial development regressions^b</i>				
Constant	0.466*** (0.044)	0.252*** (0.032)	0.682*** (0.028)	0.187*** (0.048)
Financial liberalization dummy	0.108** (0.050)	0.202*** (0.044)	0.152*** (0.034)	-0.103** (0.043)
Banking crisis dummy	-0.104* (0.055)	-0.085* (0.047)	-0.066* (0.037)	0.040 (0.039)
Adjusted R ²	0.03	0.09	0.10	0.03
Number of observations	156	156	159	153
Aggregate impact on financial development	0.004 F = 0	0.117** F = 4.62	0.086* F = 3.32	-0.063 F = 0.88
Coefficient in growth regression	-0.407	0.243	3.450**	-2.010*
Impact on growth	0.002	0.028	0.297	0.127

* Significant at the 10 percent level

** Significant at the 5 percent level.

*** Significant at the 1 percent level

a The dependent variable is real per capita GDP growth. Each growth regression includes an alternative financial development indicator, as specified in the column head. Liquidity is the ratio of liquid liabilities of the financial system to GDP. Private credit is the ratio of bank credit to the private sector to GDP. Bank assets are the ratio of domestic bank assets to the sum of central bank domestic assets and domestic bank assets. Central bank is the ratio of central bank domestic assets to GDP. In addition to the financial development indicators, the regressions include the logarithm of initial real per capita GDP, the logarithm of initial secondary school enrollment, the ratio of government consumption expenditure to GDP, the inflation rate, the ratio of exports plus imports to GDP, the real interest rate, and dummy variables for five-year periods. Numbers in parentheses are White heteroscedasticity-consistent standard errors.

b The dependent variable is the financial development indicator listed in the column header.

Source: See appendix.

The estimation results reported in table 9 show that the coefficient of the liberalization dummy is positive and significant in all the specifications, while the banking crisis dummy has a negative coefficient that is significant in all specifications except one. These results suggest that financial liberalization and banking crises significantly affect financial development. Countries and periods in which both banking crises and financial liberalization occurred appear to have greater financial development as measured by private credit or bank assets; when financial development is measured by the liquidity and central bank indicators, the difference in the coefficients is not significantly different from 0. Private credit, however, does not have a significant effect on growth in our panel, as shown in the first row of table 9. In only one regression, which uses bank assets as an indicator of financial development, are the net effect of the dummies on financial development and the effect of financial development on growth significant. At least in the medium term, then, these tests do not support the hypothesis that choosing financial liberalization at the cost of a banking crisis pays off in terms of higher growth through higher financial development.¹⁴

Additional insights on this issue can be obtained by splitting the sample between countries that were repressed at the time of financial liberalization and countries that were only restrained. Countries are classified as repressed if they had a negative interest rate (on average) during the three years preceding financial liberalization; they are classified as restrained if interest rates were positive before liberalization. Countries that maintained controlled financial markets during the entire sample period are omitted from this panel because they cannot be included in either group.¹⁵ The results for the restrained countries are quite similar to those for the sample as a whole (bank assets and central bank are significant), while for the repressed group private credit also has a significant effect on financial development (tables 10 and 11).¹⁶

More interestingly, when we regressed the financial development indicators on the liberalization dummy and the crisis dummy, banking crises did not seem to be associated with significantly lower financial development in repressed countries (where financial development is in any case lower than in the restrained group). By contrast, in restrained countries at least two of four regressions revealed a correlation between banking crises and reduced financial development (see tables 10 and 11). The positive impact of financial liberalization is evident in both groups of countries.

These estimated coefficients indicate that countries that liberalized from a position of financial restraint and experienced banking crises had levels of financial development similar to those of countries that did not liberalize and escaped banking problems. By contrast, countries that liberalized from a position of financial repression experienced higher financial development following liberalization, even if they experienced a banking crisis. Based on the coefficient estimated in the growth regression, the net positive effect on growth for this group of countries is of the order of 0.7–0.9 percentage points a year (see table 11).

To summarize, empirical evidence supports the hypothesis that financial liberalization is associated with higher financial development and, through it, higher output growth. Banking crises have the opposite effect. For countries that liberalize from a position of financial restraint, the gains from liberalization in terms of finan-

Table 10. Growth, Financial Development, Financial Liberalization, and Banking Crises—Financially Restrained Countries

	<i>Liquidity</i>	<i>Private credit</i>	<i>Bank assets</i>	<i>Central bank</i>
<i>Growth regressions^a</i>				
Financial development	-0.735 (0.841)	-0.775 (1.007)	12.418*** (4.757)	-13.417* (7.362)
Adjusted R ²	0.09	0.09	0.25	0.13
Number of observations	64	64	64	62
<i>Financial development regressions^b</i>				
Constant	0.518*** (0.075)	0.363*** (0.059)	0.788*** (0.030)	0.094*** (0.012)
Financial liberalization dummy	0.157* (0.084)	0.173** (0.074)	0.112*** (0.033)	-0.038*** (0.014)
Banking crisis dummy	-0.019 (0.111)	-0.082 (0.082)	-0.074* (0.040)	0.038** (0.019)
Adjusted R ²	0.01	0.04	0.14	0.10
Number of observations	72	72	72	69
Aggregate impact on financial development	0.138 F = 0.86	0.091 F = 0.75	0.038 F = 0.51	0 F = 0
Coefficient in growth regression	-0.735	-0.775	12.418***	-13.417*
Impact on growth	-0.101	-0.071	0.472	0

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level

a See table 9 for descriptions. Numbers in parentheses are White heteroscedasticity-consistent standard errors

b The dependent variable is the financial development indicator listed in the column header

Source: See appendix

cial development are comparable to the costs of a banking crisis. In financially repressed countries the gains from financial liberalization are greater.

Although these results are suggestive, it is important to stress that they are tentative and are based on imperfect methodology. Growth regressions are intended to study the determinants of long-run growth rates, based on many years of data. But to incorporate enough data points, we were forced to use five-year averages, which may not capture the long-run economic growth rate. In fact, the low R² in the growth regressions may indicate that cyclical and other factors not controlled for are important in explaining the dependent variable.

Another potential problem involves the bias in the estimate of the coefficient of the financial development indicator that is introduced if variables that are correlated with the development indicators are omitted. This problem is not serious, however, because it affects only the growth regressions, for which the linkage between financial development and growth is established both here and in other, more rigorous studies.

More problematic are our tests of the relationships among financial development, financial liberalization, and banking crises. These tests, which look only at short- and medium-term horizons, are basically differences of means and ignore the fact that factors other than liberalization and banking crises affect financial development. Moreover, the effect of financial liberalization on the probability of a bank-

Table 11. Growth, Financial Development, Financial Liberalization, and Banking Crises—Financially Repressed Countries

	<i>Liquidity</i>	<i>Private credit</i>	<i>Bank assets</i>	<i>Central bank</i>
<i>Growth regressions^a</i>				
Financial development	0.421 (2.217)	5.189** (2.266)	4.466** (2.018)	-2.865** (1.453)
Adjusted R ²	0.04	0.12	0.10	0.08
Number of observations	57	57	58	57
<i>Financial development regressions^b</i>				
Constant	0.411*** (0.065)	0.178*** (0.024)	0.607*** (0.048)	0.267*** (0.100)
Financial liberalization dummy	0.060 (0.073)	0.163*** (0.048)	0.183*** (0.058)	-0.162* (0.097)
Banking crisis dummy	-0.085 (0.058)	-0.022 (0.061)	-0.009 (0.060)	0.026 (0.079)
Adjusted R ²	0	0.08	0.11	0.02
Number of observations	64	64	66	64
Aggregate impact on financial development	-0.025 F = 0.09	0.141*** F = 6.17	0.174** F = 5.68	-0.136 F = .97
Coefficient in growth regression	0.421	5.189**	4.466***	-2.865**
Impact on growth	-0.011	0.732	0.777	0.390

* Significant at the 10 percent level

** Significant at the 5 percent level

*** Significant at the 1 percent level

a See table 9 for descriptions. Numbers in parentheses are White heteroscedasticity-consistent standard errors

b The dependent variable is the financial development indicator listed in the column header

Source See appendix

ing crisis is not explicitly incorporated in the analysis. We leave more sophisticated explorations of this important issue to future research.

Conclusion

The analysis in this article indicates that financial fragility is affected by many factors, including adverse macroeconomic developments, bad macroeconomic policies, and vulnerability to balance of payments crises. When these factors are controlled for, financial liberalization has an independent negative effect on banking sector stability, and the magnitude of this effect is not trivial. But strong institutions, characterized by effective law enforcement, an efficient bureaucracy, and limited corruption, can curb the adverse effects of liberalization on the financial system.

These findings suggest that institutional development needs to be emphasized early in the liberalization process. In countries where institutions are weak, achieving macroeconomic stabilization before or during liberalization would control an important independent source of financial instability. But even in an otherwise well-functioning economy, weaknesses in the institutions and regulations needed for financial markets to operate efficiently may fail to check perverse behavior on the part of financial intermediaries—laying the foundation for systemic financial sector problems. Because strong institutions cannot be created overnight, financial liberal-

ization should be pursued gradually, with the benefits of each step toward liberalization carefully weighed against the risks. That financial liberalization appears to increase the fragility of the banking system also suggests the need to design and implement prudential regulations and supervision as part of the liberalization process, especially in developing countries.

Support for a gradual approach to financial liberalization also comes from our findings on the effects of liberalization and fragility on financial development and, accordingly, on growth. In countries that were financially repressed the positive effect of liberalization on financial development appears stronger than the negative effect of a banking crisis. By contrast, in countries that liberalized from a situation of financial restraint the two effects roughly offset each other. One way to interpret these findings is that, once financial sector reforms establish positive interest rates, steps toward further liberalization may not necessarily yield gains that offset the negative impact of increased fragility.

Appendix. Definitions and Data Sources for Variables Included in the Logit Regressions

<i>Variable</i>	<i>Definition</i>	<i>Source</i>
GDP growth rate	Rate of growth of real GDP	IFS where available; otherwise, WEO
Change in terms of trade	Change in terms of trade	WEO
Real interest rate	Nominal interest rate minus the contemporaneous rate of inflation	IFS. Where available, nominal rate on short-term government securities. Otherwise, a rate charged by the central bank to domestic banks (such as the discount rate). Otherwise, the commercial bank deposit interest rate
Inflation	Rate of change of the GDP deflator	IFS
M2/reserves	Ratio of M2 to the foreign exchange reserves of the central bank	M2 is money plus quasi-money (lines 34 + 35 in IFS) converted into U.S. dollars. Reserves are line 1dd of IFS
Private/GDP	Ratio of domestic credit to the private sector to GDP	Domestic credit to the private sector is line 32 in IFS
Cash/bank	Ratio of bank liquid reserves to bank assets	Bank reserves are line 20 of IFS. Bank assets are lines 21 + lines 22a–22f of the IFS
Credit growth	Rate of growth in real domestic credit to the private sector	IFS line 32 divided by the GDP deflator
Per capita GDP	Real per capita GDP	GDP data are from the World Bank's National Accounts database. Population is line 99z of IFS
Law and order	Index ranging from 0 to 6	ICRG

Variable	Definition	Source
Bureaucratic delay	Index ranging from 0 to 4	BERI
Contract enforcement	Index ranging from 0 to 4	BERI
Quality of bureaucracy	Index ranging from 0 to 6	ICRG
Corruption	Index ranging from 0 to 6	ICRG

Note: IFS = *International Financial Statistics*, BERI = *Business Environmental Risk Intelligence*, ICRG = *International Country Risk Guide* WEO = *World Economic Outlook*.

Notes

1. Empirical research on the relationship between interest rates and savings in countries that have liberalized their financial markets has generally failed to find clear evidence of a significant and sizable positive correlation. This failure is generally attributed to the strong positive wealth effect of interest rate increases (see Fry 1997 for a survey). Yet empirical studies tend to support the view that moderately positive real interest rates have a positive effect on growth (see, among others, Roubini and Sala-i-Martin 1992 and Bandiera and others 1997).

2. Financial markets allow agents to diversify and hedge risk, thereby making high-risk, high-return investments attractive to investors. Financial markets also allow the pooling of liquidity risk (Diamond and Dybvig 1983). Stock markets disseminate information on corporate values (although dissemination of too much information may reduce the incentives for collecting information, as Stiglitz 1985 argues) and allow a market for corporate control to emerge. Financial intermediaries, such as banks, make savings available to entrepreneurs who may lack resources of their own to finance investment and technology acquisition, and they screen and monitor loan applicants, thereby improving the allocation of resources. By exploiting economies of scale, intermediaries can also make savings mobilization more efficient (Levine 1997).

3. The 1981 Chilean experience, which shares many features with the current East Asian crisis, is analyzed in Diaz-Alejandro (1985). Other case studies of banking crises are presented in Sundararajan and Baliño (1991), Drees and Pazarbasioğlu (1995), and Sheng (1996).

4. In some countries the authorities may forbid commercial banks from entering certain segments of the credit market that are deemed excessively risky, such as extending credit to security dealers. Such restrictions are sometimes relaxed as part of liberalization.

5. This problem is exacerbated if financial liberalization takes place before macroeconomic stabilization (McKinnon 1993).

6. For lack of data, the observations for some countries included in the panel do not cover the entire 1980–95 period.

7. Leaving the outliers in the panel reduces the estimated coefficient for inflation and the real interest rate but does not change the results very much. Peru also experienced hyperinflation during the sample period, but the hyperinflation years are excluded from the panel because of missing data.

8. For more details on the relationship between the theory of banking crises and the choice of control variables, see Demirgüç-Kunt and Detragiache (1997).

9. To minimize potential endogeneity problems, in measuring the real interest rate we used the rate on short-term government paper or a central bank rate (such as the discount rate), and not a bank interest rate. In six countries, however, neither measure was available, and we used the bank deposit rate.

10. The model χ^2 tests the joint significance of the regressors by comparing the likelihood of the model with that of a model with the intercept only. The AIC (Akaike's Information

Criterion) is calculated as minus the log likelihood of the model plus the number of parameters being estimated, and is therefore smaller for better models. This criterion is useful in comparing models with different degrees of freedom. The percentage of crises and the total percentage of observations that are correctly classified are reported to assess the predictive accuracy of the model. A crisis is deemed to be accurately predicted when the estimated probability exceeds the frequency of crisis observations in the sample (around 5 percent). This criterion tends to downplay the performance of the model, because in a number of episodes the estimated probability of a crisis increases significantly a few years before the episode begins, and those observations are considered as incorrectly classified by the criterion (see Demirgüç-Kunt and Detragiache 1997 for examples).

11. Keeley (1990) presents empirical evidence that supports this view. First, he shows that in the 1970s U.S. thrift institutions began to lose charter value as a result of technological changes and the relaxation of regulatory entry restrictions. Second, he shows that banks with higher charter values were less risky, as measured by the risk premium on uninsured certificates of deposit.

12. Of course, for a given franchise value, high capitalization and liquidity should create fewer incentives to take on risk.

13. The financial liberalization dummy variable takes a value of 1 if interest rate liberalization began in any of the years of the subperiod or if markets were liberalized in the preceding subperiod; the banking crisis dummy variable takes a value of 1 if a crisis was ongoing in any of the years of the subperiod. The results are robust to redefining the dummy variables by treating a subperiod as a 1 only if the change in policy (crisis) occurs in the first three years of the subperiod. If the change in policy (crisis) takes place in the last or second-to-last period, the dummy for the following period is set to 1.

14. The coefficients of a growth regression including the banking crisis dummy and the financial liberalization dummy are not significant, suggesting that the dummies have a negligible direct impact on growth.

15. The panel includes countries that liberalized well before the beginning of the sample period. It may be argued that whether those countries were financially repressed or restrained at the time of liberalization should not affect their economic performance in 1980–94. As a robustness test, we repeated the tests described below, dropping those countries from the panel. The basic results remain unchanged.

16. Roubini and Sala-i-Martin (1992) find the negative growth effects of financial repression to be stronger in financially repressed countries than in financially restrained countries.

References

- Bandiera, O., Gerard Caprio, Patrick Honohan, and F. Schiantarelli. 1997. "Does Financial Reform Raise or Reduce Saving?" World Bank, Washington, D.C.
- Calvo, Guillermo A. 1996. "Capital Flows and Macroeconomic Management: Tequila Lessons." *International Journal of Finance and Economics* 1 (July): 207–24.
- Caprio, Gerard Jr., and Daniela Klingebiel. 1996. "Bank Insolvencies. Cross-Country Experience." Policy Research Working Paper 1620. World Bank, Washington, D.C.
- Caprio, Gerard Jr., and Lawrence Summers. 1993. "Finance and Its Reform: Beyond Laissez-Faire." Policy Research Working Paper 1171. World Bank, Washington, D.C.
- Chari, V. V., and C. R. Jagannathan. 1988. "Banking Panics, Information, and Rational Expectations Equilibrium." *Journal of Finance* 43: 749–61.
- Demirgüç-Kunt, Asli, and Enrica Detragiache. 1997. "The Determinants of Banking Crises: Evidence from Industrial and Developing Countries." Policy Research Working Paper 1828. World Bank, Washington, D.C.

- Diamond, D., and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (June): 401–19.
- Diaz-Alejandro, Carlos. 1985. "Good-Bye Financial Repression, Hello Financial Crash." *Journal of Development Economics* 19 (September–October): 401–19.
- Drees, Burkhard, and Ceyla Pazarbasioglu. 1995. "The Nordic Banking Crises. Pitfalls in Financial Liberalization?" IMF Working Paper 61. International Monetary Fund, Washington, D.C.
- Fry, M. J. 1997. "In Favor of Financial Liberalization." *Economic Journal* 107: 754–70.
- Greene, William H. 1997. *Econometric Analysis*. Upper Saddle River, N.J.: Prentice-Hall.
- Hellmann, Thomas, K. Murdock, and Joseph E. Stiglitz. 1998. "Liberalization, Moral Hazard in Banking and Prudential Regulation: Are Capital Requirements Enough?" Research Paper 1466. Stanford University, Graduate School of Business, Stanford, Calif.
- IMF (International Monetary Fund). Various years. *International Financial Statistics*. Washington, D.C.
- . Various years. *World Economic Outlook*. Washington, D.C.
- Kaminsky, George, and C. M. Reinhart. 1996. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." Federal Reserve Board, Washington, D.C.
- Keeley, Michael C. 1990. "Deposit Insurance, Risk, and Market Power." *American Economic Review* 80 (5): 1183–200.
- King, Robert G., and Ross Levine. 1993. "Finance and Growth: Schumpeter may be Right." *Quarterly Journal of Economics* 108: 717–37.
- Levine, Ross. 1997. "Financial Development and Growth: Views and Agenda." *Journal of Economic Literature* 35: 688–726.
- Lindgren, Carl-Johan, Gillian Garcia, and Matthew Saal. 1996. *Bank Soundness and Macroeconomic Policy*. Washington, D.C.: International Monetary Fund.
- McKinnon, Ronald I. 1973. *Money and Capital in Economic Development*. Washington, D.C.: Brookings Institution.
- . 1993. *The Order of Economic Liberalization: Financial Control in the Transition to a Market Economy*. 2nd ed. Baltimore, Md.: Johns Hopkins University Press.
- Roubini, N., and Xavier Sala-i-Martin. 1992. "Financial Repression and Economic Growth." *Journal of Development Economics* 39: 5–30.
- Shaw, E. 1973. *Financial Deepening in Economic Development*. New York: Oxford University Press.
- Sheng, Andrew. 1996. *Bank Restructuring: Lessons from the 1980s*. Washington, D.C.: World Bank.
- Stiglitz, Joseph E. 1985. "Credit Markets and the Control of Capital." *Journal of Money, Credit, and Banking* 17: 133–52.
- . 1994. "The Role of the State in Financial Markets." In Michael Bruno and Boris Pleskovic, eds., *Proceedings of the World Bank Annual Conference on Development Economics 1993*. Washington, D.C.: World Bank.
- Sundararajan, Vasudevan, and Tomas Baliño, eds. 1991. *Banking Crises: Cases and Issues*. Washington, D.C.: International Monetary Fund.

Comment on “Financial Liberalization and Financial Fragility,” by Asli Demirgüç-Kunt and Enrica Detragiache

Charles A.E. Goodhart

I will start with a riddle. What do Cambodia, the Democratic People’s Republic of Korea, Denmark, Germany, Iraq, Libya, and the United Kingdom have in common that distinguishes them from Egypt, Finland, Kuwait, the Republic of Korea, Spain, Thailand, and the United States? If your answer is that the first group of countries prefers their beer warmer, good try. But that is not the right answer. The right answer is that the countries in the first group did not suffer a banking crisis over the data period covered in recent work by the International Monetary Fund and the World Bank (Lindgren, Garcia, and Saal 1996; Caprio and Klingebiel 1996). The composition of this group suggests that maintaining a repressed, nonmarket economy in which the banking sector has virtually no freedom may be worse than suffering a banking crisis.

To be sure, there are numerous reasons why liberalization can create solvency problems, not all of which are mentioned in Asli Demirgüç-Kunt and Enrica Detragiache’s article. For example, repression—or even restraint by direct control—is usually associated with a large share in bank portfolios of government debt and of loans to state enterprises and to other large companies that the government explicitly or implicitly guarantees. This setup may not offer optimal resource allocation for growth or social welfare, but it does mean that banks in repressed socialist countries were considerably safer than they are in liberalized financial systems.

The period following financial liberalization is subject to asset price volatility—boom followed by bust—excessive risk taking, and financial fragility (see Lindgren, Garcia, and Saal 1996). In the United Kingdom the main liberalization reform, titled *Competition and Credit Control* and enacted in 1971, was followed by an asset price upsurge in 1972–73 that collapsed in 1973–74. The accompanying contagious financial crisis, known as the fringe bank crisis, was one of the few contagious crises to hit the United Kingdom in the 20th century.

Demirgüç-Kunt and Detragiache deserve praise for trying to submit the interaction between liberalization and financial fragility to econometric testing. My main concern is whether they have done as good a job as possible of taking the analysis beyond descriptive historical account into statistical hypothesis testing.

Charles A.E. Goodhart is professor of economics at the London School of Economics.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

Some General Themes

Liberalization tends to create fragility for two rather different reasons. First, a liberalized financial system is more likely to be competitive. Competition, including competition from foreign banks, erodes oligopolistic supranormal profits and thus reduces franchise values, profitability, and capital. The authors investigate this issue but—as they are the first to admit—their results are based on too short a data period and so are preliminary rather than conclusive. The effect of competition on franchise values is presumably quasi-permanent; as a result liberalized systems may be more fragile than other systems, regardless of how long ago liberalization occurred.

The second reason is more transitional. Bankers operating under a repressed system are not trained to assess risk or monitor loan performance. With liberalization, these bankers fall prey to the pitfalls of excessive portfolio expansion. Assessment and monitoring skills usually develop only after painful lessons of experience.

My reading of banking history indicates that the transitional danger is greater and more severe than the permanent levels effect. The authors find otherwise. The levels dummy in table 2 of their article—which takes a value of 1 for countries with liberalized systems, regardless of when the system was liberalized—remains highly significant. The results on transitional effects, measured by the variables financial liberalization 3–6, are much less so.

What might account for these results? One possible explanation is that the model has been misspecified. The authors treat financial liberalization as an exogenous variable. But surely liberalization is not exogenously determined. Liberalization is a policy choice, and the dangers associated with that choice are fully appreciated. Indeed, liberalization is likely to be undertaken only when economic conditions seem propitious—usually when the economy and private sector demand for loans are depressed. If such is the case, it is no wonder that it takes time for the potentially dangerous effects of liberalization to feed through.

In fact, as the authors' table 8 shows, banking margins and profitability rise in the years immediately after liberalization, because liberalization is usually adopted when the economy is near the bottom of a trough. Liberalization ushers in a period of upturn that all too often overshoots into an unsustainable boom. An interactive term, whereby liberalization interacts with an asset price boom to cause fragility, might be tried as the appropriate transitional variable. A companion exercise should be run examining the conditions under which liberalization is undertaken. In brief, because liberalization is an endogenous, and not a purely exogenous, variable, the study should be a simultaneous equation exercise.

Questions and Cavils

A few other issues in the article also raised my concern. For example, why are certain control variables, such as the rate of change in equity prices, the exchange rate regime (pegged, floating, or fixed), and the use (or not) of exchange controls not included? Why were some countries included and others left out? The United States,

for example, is included, but the United Kingdom is not. Portugal is in, but Spain is out. Zambia is in, but South Africa is out. Most of Latin America is in, but Brazil is out. India is in, but Pakistan is out. Jordan is in, but Kuwait is out. The authors also include the city states of Hong Kong (China) and Singapore, omit most Middle Eastern countries, and describe Italy as having experienced a banking crisis in 1990–94—a fact that will come as a surprise to my friends at the Banca d'Italia.

More generally, I found it hard to follow what was being done about the problem of simultaneity or about coping with different initial conditions. The financial liberalization variables should reflect the fact that some countries may have liberalized and experienced a crisis before 1980, when the dataset begins. Insofar as I understand what the authors have done, I am not sure that I regard their methodology as best practice.

I do not want to end on a critical note, however. The authors have run worthwhile exercises. For example, their ingenuity in looking at the role of institutional conditions—such as the rule of law and the prevalence of corruption—in inducing crisis deserves praise. And the connection they make between the effect of liberalization and the extent of repression before reform is insightful.

References

- Caprio, Gerard Jr., and Daniela Klingebiel. 1996. "Bank Insolvencies: Cross-Country Experience." Policy Research Working Paper 1620. World Bank, Washington, D.C.
- Lindgren, Carl-Johan., Gillian Garcia, and Matthew Saal. 1996. *Bank Soundness and Macroeconomic Policy*. Washington, D.C.: International Monetary Fund.

Comment on “Financial Liberalization and Financial Fragility,” by Aslı Demirgüç-Kunt and Enrica Detragiache

Thomas Hellmann

Aslı Demirgüç-Kunt and Enrica Detragiache have made an important contribution to ongoing World Bank research on financial development. The authors are part of a small group of researchers who have engaged in extremely valuable data collection on the experiences with financial development of a large number of countries from a wide variety of sources. Armed with this unique dataset, the group has set out to establish the empirical facts on financial development and to address a difficult set of policy implications.

Demirgüç-Kunt and Detragiache continue this tradition by addressing a number of questions that are central to the debate on the effectiveness of financial sector liberalization. While financial liberalization has often been a core element—if not the flagship—of economic reform, subsequent banking crises in a number of liberalizing countries have tarnished the outcome that economic reformers hoped to achieve. In their article the authors ambitiously set out to answer three questions. First, they ask whether banking crises have become more common in the aftermath of financial liberalization. Second, they explore some of the ways in which financial liberalization increases instability, focusing on the institutional environment and on banks’ incentives. Finally, they provide preliminary evidence on whether financial liberalization is worth it, given the risk of financial crisis.

Liberalization and Banking Crises

On the first question, the article provides some convincing evidence on the link between financial liberalization and the likelihood of a banking crisis. The authors provide a set of robustness checks that increases confidence in their results. One problematic thought lingers in the background, however. Caprio and Klingebiel (1996) establish that while banking crises are typically triggered by macroeconomic shocks, they often uncover weaknesses in the underlying microeconomic structure of the financial system. Thus a financial crisis is a discovery of information about an underlying problem.

Thomas Hellmann is assistant professor in the Graduate School of Business at Stanford University

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

One may therefore be concerned that fundamental problems that would trigger a crisis in a liberalized system might never surface in a highly regulated system. Put differently, unfavorable information may be discovered more often in a liberalized system either because there are more unfavorable events or because for the same number of such events the probability of discovery is higher. In their defense, the authors identify financial crises in highly regulated financial systems, indicating that there can be discovery in those systems. Moreover, we may care most about the events of crisis, where the discovery of the problem often affects other economic activity—including a recession.

Subject to this concern, I consider the finding of a link between liberalization and the likelihood of a financial crisis very important, and it is noteworthy that the result is not only statistically robust but also economically large. This finding provides an extremely useful starting point both for this article and for this line of research more broadly. It establishes that we cannot ignore the issue of financial crises when we discuss the merits of financial liberalization. This result alone should improve many policy debates that have had friends and foes of financial liberalization quibbling about the existence of a problem, rather than thinking about solutions.

Liberalization and Instability

Obviously, one would like to know more about the link between financial liberalization and instability. For example, country experiences with liberalization differ in terms of their speed, their extent, and the emphasis they place on the different aspects of liberalization. Moreover, financial crises differ in their magnitude and duration.¹ The question of timing seems particularly important. The article presents evidence that suggests that financial crises take time to develop, at least as measured relative to the initial liberalization of interest rates. What explains this delay? And if the likelihood of a crisis is initially increasing, is there an indicator of when it might be coming down again?²

One possible explanation relates to the previous conceptualization of a financial crisis as a discovery of information: to fully understand the delay and the timing of a crisis, I suspect that we need to look at the internal workings of a bank. In a highly regulated environment bank officials have limited responsibilities because they are only implementing government directives. After deregulation banks and their loan officers become somewhat more accountable. In the early days of liberalization a lot of learning is taking place, but the mistakes of this learning are unlikely to come to the surface quickly. Loan officers may hide their mistakes, and in the process they will make more mistakes—leading to the well-known problem of banks throwing good money after bad. This phenomenon may then account for the observed lags between liberalization and financial crises.

This conjecture on the timing of crises naturally leads to the question of why there is a link between financial liberalization and financial fragility. I have already noted that banking crises are typically triggered by macroeconomic shocks, but that these only uncover weaknesses in the underlying microeconomic structure of the financial

system. Demirgüç-Kunt and Detragiache examine two sets of such microeconomic arguments that may affect the underlying strength of the financial sector. First, they look at a number of quality measures in the institutional environment. Second, they examine the strength of a bank's incentives as measured by its franchise value.

The analysis of the institutional environment reveals some intriguing patterns. Institutional strength is measured by high respect for the rule of law, low corruption, good contract enforcement, and high-quality, fast bureaucracies. Somewhat surprisingly, the authors omit a discussion of the direct effect of institutional strength on the likelihood of a banking crisis (and on the likelihood of financial liberalization in the first place). What the authors do report, however, are some interesting interaction effects: the better is the institutional environment, the smaller is the impact of financial liberalization on banking crises. This finding suggests that the workings of a liberalized market require well-functioning institutions and norms. And the choice of policy instruments and the effectiveness of policy depend on the underlying strength of the microeconomic structure, reminiscent of the theoretical distinction between first-best and second-best solutions.

To make further progress on this set of questions, it would be interesting to link the concept of institutional strength more closely to the problem of financial stability. Do imperfections in private contracting, and especially norms about corruption and insider lending, undermine the soundness of banking systems? Or does the institutional environment mainly affect the quality of prudential supervision? Put differently, does the problem of a weak institutional structure stem mainly from a market or a government failure? To answer some of these questions, the authors might want to include variables that directly capture the problems in banking—particularly variables about the formal and informal structure of bank supervision.

The second set of issues that the authors consider relates to the question of franchise value. Hellmann, Murdock, and Stiglitz (1998) derive a general condition that states that a bank is willing to take a gamble on its portfolio if the short-term profits from winning the gamble exceed its franchise value—that is, the net present value of its future profits when invested prudently.

Theory tells us that the expectation of future profits drives banks' incentives but the empirical measurement of the franchise value is not trivial. Probably the best measure is the stock price of a bank's equity. But because many banks do not have reliable stock prices—mostly because of the absence of efficient stock markets—the authors have to use indirect instruments. I question the use of contemporaneous return measures (return on equity or assets) because they omit precisely those aspects that are emphasized by theory—namely, that it is not so much current profits that matter, but the expectation of future profits. In fact, contemporaneous profits may be negatively correlated with franchise value. If banks expect lower profits in the future from a prudent investment strategy, they may be tempted to engage in an imprudent gambling strategy that yields higher returns in the short run, with the risk of insolvency down the line. I return to this point below.

Measuring franchise value through the extent of competition strikes me as reasonable, as long as it is understood that this is an indirect measurement that rests on

the reasonable assumption that competition will put downward pressure on bank profits.³ For the other measures I agree that higher capital adequacy and greater liquidity are likely to increase the cost of capital. The likely effect of the share of deposits on franchise value seems unclear, however, given that the composition of liabilities is an endogenous choice for banks.

The empirical results are clearly preliminary and deserve further research. The franchise value hypothesis is that if financial liberalization reduces banks' franchise value, then this is one of the reasons why financial liberalization may lead to banking crises. The authors acknowledge that they are only examining whether financial liberalization reduces franchise value, not whether such a reduction can account for the higher incidence of banking crises.

To examine this first step of the hypothesis, the authors distinguish between a temporary effect and a long-term correlation between financial liberalization and their measures of franchise value. While I question the methodology—countries that have recently liberalized are compared with a default category that mixes countries with no liberalization and countries with more distant liberalization—the result that financial liberalization increases the return on assets in the short run but not in the long run is actually consistent with the franchise value hypothesis. As noted, the return on assets is not a measure of franchise value, but a measure of short-term profits. But if financial liberalization reduces the franchise value of banks, they will want to increase their short-term returns.⁴

It is also worth noting that capital gradually increases with liberalization, presumably as a result of regulatory pressures. This lowers the franchise value (the cost of capital increases for banks) and leads to the question of when during the process of liberalization banks are most likely to pursue risky investment strategies. It would also be interesting to link these considerations with the apparent fact established earlier in the article that banking crises occur not in the immediate aftermath of liberalization, but take some time to develop.

Is Liberalization Worth It?

The final section of the article discusses the relationships between financial liberalization, banking crises, and growth. After having presented convincing evidence establishing the link between liberalization and banking crises, the authors seem to be trying to offer an olive branch to protagonists of financial liberalization. They argue that while there are problems with financial liberalization, at the end of the day all is still well, because the benefits of liberalization outweigh its cost—at least for countries that practice financial repression rather than financial restraint.

Regardless of whether we believe that all is well at the end of the day, the current evidence is clearly preliminary. The authors find that even with a financial crisis, liberalization increases private credit. This is hardly surprising, because liberalization inherently transfers responsibilities from the public to the private sector. In fact, the authors then have difficulty replicating some of the results from other studies on the link between financial development variables, such as private credit, and economic growth.

In summary, the paper makes a valuable contribution by establishing a link between financial liberalization and financial crises, and it provides intriguing preliminary evidence on some of the microeconomic drivers of this link, such as the quality of the institutional environment and the incentive environment of banks, as measured by the franchise value.

Notes

1. Future research might return to the data that allowed for the construction of the two crucial dummy variables and develop a more detailed picture of the link.
2. The authors might want to think about extending their multinomial logit approach with duration models that can explain the hazard rate of a financial crisis as a function of a number of explanatory variables, such as the duration of financial liberalization. Such an analysis might also clarify the effect of a reversal of financial liberalization and the effect of a previous financial crisis.
3. I question, however, the use of three firm concentration measures that may not accurately reflect the incentives of small banks. Such banks may be particularly prone to engage in unsound investment strategies with the advent of competition in the newly liberalized environment.
4. Moreover, the fact that long-term returns are the same (as opposed to being lower) might be the result of a survivor bias in the measurement of returns.

References

- Caprio, Gerard Jr., and Daniela Klingebiel. 1996. "Bank Insolvencies. Cross-Country Experience." Policy Research Working Paper 1620. World Bank, Washington, D.C.
- Hellmann, Thomas, H. Murdock, and Joseph E. Stiglitz. 1998. "Liberalization, Moral Hazards in Banking, and Prudential Regulation: Are Capital Requirements Enough?" Stanford University Graduate School of Business Working Paper 1466R. Stanford, Calif.

Floor Discussion of “Financial Liberalization and Financial Fragility,”

by Aslı Demirgüç-Kunt and Enrica Detragiache

A participant from the University of Maryland asked about the effects of financial liberalization on banking crises. In addition to its direct and negative effect, liberalization increases GDP, and an increase in GDP lowers the probability of a banking crisis. The models developed by Aslı Demirgüç-Kunt and Enrica Detragiache (presenters) included GDP growth on the right-hand side of the equation, so they essentially showed only the direct (negative) effects of liberalization on banking crises. But the authors’ conclusion that liberalization increases the probability of a banking crisis draws on both the direct and indirect effects of liberalization. Thus, the participant concluded, it would be useful to have regressions that did not include GDP growth on the right-hand side.

Demirgüç-Kunt said that their growth regressions used the standard right-hand side variables, and that liberalization and crisis had no direct effect on growth. Thus they examined the indirect effects of liberalization and crisis by assessing how financial development affected growth, and found that both liberalization and crisis had a significant effect on growth.

A participant from the World Bank believed that a banking crisis occurs when financial liberalization is followed by a number of bank failures—as in Italy in 1994. There is nothing wrong with such a crisis, said the participant, because the purpose of financial liberalization is to increase competition in the market. Thus banks that are not competitive go out of business. But could Demirgüç-Kunt and Detragiache explain why some countries have not experienced financial crisis even after financial liberalization? Do institutional factors, such as deposit insurance, protect banks from falling into crisis?

Their models, replied Detragiache, did not distinguish between crises that had high costs to the economy and those that did not. In previous work, however, Demirgüç-Kunt and Detragiache had found that explicit deposit insurance seems to increase the probability of a crisis.

This session was chaired by Gary L. Perlin, vice president and treasurer at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

A banking crisis may not eliminate inefficient banks, asserted Thomas Hellman (discussant). The data indicate that a crisis makes banks, through a common macroeconomic shock, face bad loan portfolios that are highly correlated—that is, they are all making the same mistakes. In the new, liberalized environment, bankers may be willing to take large risks simply because everyone else is doing the same thing. This pattern of behavior, said Hellman, is different from entrepreneurial risks involving venture capital, which are much better for the economy.

A participant from the Carnegie Endowment for International Peace objected to the idea that because strong institutions cannot be created overnight, financial liberalization should proceed gradually. A financial crisis is typically the bust that follows a boom. The boom is caused by excessive monetary expansion, which is usually the result of financial liberalization. Thus there are early benefits from liberalization, but eventually government must impose restraints because of the negative effects of the monetary expansion. As a result liberalization ends up being gradual—as in Finland, Norway, and Sweden in the 1980s. Would it not be more accurate, wondered the participant, to define a financial crisis as a problem of a gradualist approach to developing regulation?

In their models, answered Demirgüç-Kunt, liberalization was a dummy variable measured as 0 or 1. She and Detragiache had not considered gradual liberalization. Still, they strongly believed that institutions cannot be created overnight, especially in developing countries, so it followed that financial liberalization should be gradual. Countries that have achieved good results with rapid liberalization, such as the United Kingdom, have highly developed institutions and so did not face the same issues.

A participant from the World Bank noted that the East Asia crisis had made many observers wonder why the Bank and the International Monetary Fund (IMF) had not paid more attention to banks in developing countries. Bruce Greenwald (presenter in another session) had suggested that rather than building factories, international institutions should be teaching local banks to build factories. The participant wondered whether this is an appropriate role for the Bank and the IMF. And if it is, why have they avoided it in the past?

What is needed, said Charles A.E. Goodhart (discussant), is international oversight of national supervisory bodies in both industrial and developing countries. A World Bank-IMF team could perform this function, visiting countries to assess supervisory institutions and publishing the results. Goodhart hoped that the two institutions would work together to take on that responsibility.

Ethnic Conflicts

Structure and Strategy in Ethnic Conflict: A Few Steps toward Synthesis

Donald L. Horowitz

There is a longstanding difference in approach between those who see ethnic groups as firmly bounded, durable communities, inclined to ethnocentrism, hostility to outsiders, and passionate conflict, and those who see them as social constructs within which solidarity is based on material rewards and conflict behavior is based on calculation. This difference of approach ought to yield to a new synthesis based on an understanding of ethnicity as a powerful Gemeinschaft affiliation that can induce both calculative and passionate action. The article outlines the alternative perspectives, describing 10 often mutually incompatible theories of ethnic conflict. It then advances a view of ethnicity that is grounded in a deep sense of sociality, buttressed by the birth nature of the affiliation, the sense of similarity among group members, and their difference from others. The article argues that this view of the ethnic group is compatible with changing ethnic boundaries and other claims of social constructionists but that the latitude of elites to steer the process is limited. Because a powerful affiliation, such as ethnicity, attracts the interest of those who wish to use it instrumentally, ethnic group behavior is likely to be both passionate and calculative. The recurrent tendency of groups to cleave from other groups and to claim a disproportionate share of rewards fits remarkably well with the incentives to bifurcation that typically obtain in democratic politics. This tendency and these incentives make it easy for political leaders to mobilize along ethnic lines and difficult to break the centrifugal tendencies that prevail in divided societies. In ethnic relations and in democratic politics, maximal inclusiveness is a strongly disfavored outcome.

For decades, ethnicity has been recognized as a powerful force in the politics of many countries, with profound effects on prospects for democracy (Horowitz 1993; Young 1993, pp. 18–19) and for economic development and the distribution of public goods (Easterly and Levine 1997; Klitgaard 1991; Meerman 1979). There is less agreement than ever, however, on the causes of ethnic conflict. The dis-

Donald L. Horowitz is James B. Duke Professor of Law and Political Science at Duke University. The author is greatly indebted to Layna Mosley for research assistance.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

agreements relate less to the facts surrounding ethnic conflict than to varying ways of interpreting those facts. Anthropology, remarks Clifford Geertz (1973, p. 29), "is a science whose progress is marked less by a perfection of consensus than by a refinement of debate. What gets better is the precision with which we vex each other." Ethnic conflict studies have been characterized by decreasing consensus and increasing mutual vexation.

Not all of the dissensus is necessary. On many issues, there has been a fruitless stalemate of opposed, polar positions that ought to give rise to a synthesis. The aim of this article, therefore, is not to break new ground but to rake over old ground (particularly for an audience that has not seen the terrain before) and to seek some common ground. Its purpose, in short, is to take a few short steps toward synthesis.

Conflict Theory: Hard and Soft Perspectives

Several schools of thought advance conflicting theories of ethnic conflict, in various permutations and combinations. The issues on which they differ are several, but the differences can be reduced to hard views of ethnic conflict versus soft views, where *hard* and *soft* refer to the nature of group affiliations and the ends of conflict behavior. For a hypothetical theorist who adheres to all the hard positions (no one actually does), ethnic groups are ascriptive, firmly bounded entities based on a strong sense of commonality, producing considerable loyalty, persisting over time, providing large affective rewards to group members, inclined to ethnocentrism and to hostility to and a desire to dominate outsiders, liable to conflict behavior based on passion (even to the exclusion of calculation), and engendering a great willingness on the part of group members to sacrifice for collective welfare. For a hypothetical theorist who adheres to all the soft positions (again, no one does), ethnic groups are entities whose boundaries are problematic and malleable, whose solidarity is based on the material rewards they provide for their members rather than on diffuse affection, whose behavior, based on the interests of their members, is vulnerable to strategic manipulation, whose apparent affect can often be reduced to calculation, and whose severe conflicts with others often result less from irreconcilable objectives than from strategic dilemmas. These are hard and soft positions in the sense that the first sees ethnic affiliations as made of stone, while the second sees them as made of putty.

The range of issues on which disagreement is possible along these continua is very great. At least three issue-clusters can be distinguished.

One has to do with the very concept of an ethnic group and the features it may or may not share with other affiliations. Is an ethnic group inevitably an exogenous variable, a given, or is it merely a vehicle created, say, for the extraction of resources from an environment? Where do ethnic groups come from, anyway? Once we discover that group boundaries change, that some ethnic groups die while others are born, what is left of the idea that groups are merely given? On the other hand, once we discover that ethnic affiliations seem to possess competitive advantages over other forms of affiliation in attracting the loyalty of their members, can they simply

be regarded as vehicles for social purposes that are more or less fungible with other affiliations? What is at stake is the very idea of ethnicity as a distinctive affiliation.

A second cluster relates to the opposition of passion and interest as wellsprings of human behavior. Here sides have been chosen in debate, and they are powerfully antagonistic. Where some analysts see love and hatred, others see straightforward calculation. Where some see expressiveness, others see instrumental action. Where some see perceptual distortion driven by affect, others see appropriate response resulting from the demands of the situation. Each side tries to reduce one to the other.

A third cluster concerns the extent to which the collectivity should be considered the central actor, the unit of analysis. A deeply embedded groupness, characterized by diffuse sentiments, altruism, and the willingness of individuals to make sacrifices is characteristic of the hard position. In contrast, the soft view posits that ethnic groups can be decomposed into the motives of their members, that they are instruments for the pursuit of those motives, and that their actions need to be explained in terms of individual calculations of utility in specific contexts rather than some transcendent collective purpose. On the whole, the importance accorded leaders is greater among adherents of the soft persuasion; the role of mass sentiments and behavior is greater among hards. To the antinomies that make progress in this field difficult is thus added the methodological issue of the proper locus of investigation, macrosocietal or microstrategic.

The incompatibility of these perspectives depends on whether they are seeking to explain the same aspects of the same phenomenon. If the question is how to understand ethnic conflict in the large, the views remain completely incompatible. If the question is how to explain some large chunk of conflict behavior, such as ethnic violence, the range of incompatibility remains great. But suppose ethnic violence is disaggregated into several forms: violent protest demonstrations, deadly riots, and secessionist warfare, for example. Is it not likely that these phenomena will be explicable in terms of varying configurations of independent variables? Affect may be more important in deadly riots, calculation more important in secessionist warfare—especially in situations where preexisting units, such as the Yugoslav republics, made decisions about warfare in structured institutional settings. Even then, a variety of fears and perceptual distortions on the part of decisionmakers can hardly be ruled out (Jervis 1976). But this would still leave open the contested foundations of affect (*Rationality and Society* 1993; de Figueiredo and Weingast 1997).

The opportunities for disaggregated analysis are enormous. Everywhere the issue of the birth and death of ethnic groups presents itself for explanation. In some states, ethnic conflict preempts conflict along other cleavage lines; in others, it coexists with them; elsewhere, ethnic conflict is restrained. In some states, interethnic attitudes remain unremittingly hostile; in others, they have undergone changes, whether benign or malign. In some states, seemingly manageable conflicts have become unmanageable; in others, the reverse is true. In some, political parties are ethnically based; in others, multiethnic coalitions have formed. There is abundant variation over time and space, and there are bite-sized pieces into which ethnic conflict can be sliced. There is no a priori reason to swallow it whole.

Ten Explanations for Ethnic Conflict

So far I have characterized the competing hard and soft perspectives rather abstractly. When we get down to the level of explanations actually advanced, the disarray is greater than a simple polarity would suggest. Before going any further, I provide capsule—and therefore somewhat caricatured—descriptions of 10 views of ethnic conflict that have had some currency. For each, I point out some shortcomings or some evidence that is not consistent with the explanation.

1. Ethnicity is a primordial affiliation, in the sense that it is connected to the things people cannot live without, among them traditionality, the persistence of the past into the present, and a sense of collective self-consciousness (Shils 1957, 1995; Geertz 1963; Connor 1993, 1994; Isaacs 1975). A sense of community of this sort—*Gemeinschaft* rather than *Gesellschaft*—necessarily generates awareness of other communities, which spills over (by unspecified mechanisms) into conflict and violence. Ethnic affiliations are highly charged and, on some accounts, nonrational.

It seems futile to gainsay the emotive power of ethnic affiliations. A good explanation will have to come to grips with the thick, compelling character of group membership. But primordialists have not provided a satisfying explanation for the emotive quality of these affiliations. They have identified the phenomenon, but most have not gone behind it. If ethnic behavior is primordial in the fundamental sense in which that term is used, why is ethnic conflict so variable, and why is even the definition of group boundaries so sensitive to shifting contexts?

2. Ancient hatreds between groups produce conflict. Few scholars currently embrace this view, which remains popular with some foreign policymakers, frustrated with the intractability of ethnic conflict in distant lands. Nevertheless, some students of ethnic warfare have pointed out that groups concerned about possible threats to their security from other groups will ask themselves how those other groups behaved in comparable circumstances on previous occasions. If the previous behavior was hostile, the current response deemed appropriate may also be hostile (Posen 1993; Fearon 1994; de Figueiredo and Weingast 1997). Although these students are skeptical of ancient hatreds arguments, this way of seeing the matter opens the door to antecedent hostility as an explanatory variable.

Some antipathies are very old, but many are not. Romanian-Hungarian conflict and Bulgarian-Turkish conflict certainly have something to do with turning tables on former imperial rulers. So does hostility between Cambodians and Vietnamese in Cambodia. But one needs to be very careful on this point, because the histories of some groups, such as Sinhalese and Tamils in Sri Lanka, have been rewritten to emphasize implacable enmity even in the face of considerable cooperation and even assimilation. Furthermore, a good many groups in conflict had never really encountered each other until relatively recently. Although it may play a role in some con-

flicts, traditional antipathy is far from a sufficient explanation for all current conflicts.

3. Ethnic conflict entails a clash of cultures. It pits against each other people whose values are in conflict, who want different things, and who do not really understand each other.

This view is challenged by the fact that, more often than not, people who seek different rewards do not impinge on each other: longstanding ethnic divisions of labor are usually shields, rather than swords, in conflict. Groups in conflict usually know each other well and dislike what they have experienced of each other, or they fear the traits and ambitions of other groups that threaten them. Certainly, the well-known contact hypothesis that proposes that cooperative intergroup contact will reduce antipathy has not been verifiable (Forbes 1997).

4. Ethnic conflict is brought on by modernization. Modernization makes people want the same things, not different things, and this sets up a great scramble for resources (Melson and Wolpe 1970; Tellis, Szayna, and Winnefeld 1998; Olzak and Nagel 1986).

If modernization produces conflict, why do some of the least modernized societies in the world (highland Burma, Chad, Mauritania, southern Sudan) have such intense ethnic conflict? And if modernization is the source of conflict, why does one so often hear that if economic development proceeds, conflict will be reduced? The relationship may be curvilinear, but as things now stand these are equal and opposite ideas.

5. Ethnic conflict is the result of economic competition between ethnically differentiated segments of the working class or between ethnically differentiated traders and customers (Bonacich 1972, 1973). In the one case, it is the underbidding of labor that causes resentment; in the other, it is the monopoly position minority entrepreneurs are able to exploit that is the cause.

Such conflicts are more difficult to find than might be imagined. As I suggested, an ethnic division of labor inhibits much conflict of the first sort, and the fact that middleman minorities are thought by their customers to be providing a valuable service, rather than exploiting them, inhibits much conflict of the second sort. (That is not to say that such groups are not objects of conflict on other grounds; some are.) In any case, if conflict is economically motivated, why does it take place along cross-class ethnic lines rather than along cross-ethnic class lines?

6. An answer of sorts to this last question is provided by those who see the ethnic group as "a service producing club" (Congleton 1996, p. 74), an affiliation designed to provide a reliable basis for exchange in the absence of enforcement mechanisms to prevent cheating (Wintrobe 1996). By engendering trust, ethnic networks reduce cheating, and they impose sanctions on members when cheating does occur. The high barriers to entering and leaving the group make ethnicity a particularly apt affiliation to perform these functions (Wintrobe 1996). From this approach there are two roads to conflict. Groups

use their influence to obtain state services, resulting in ethnically targeted policies and then redoubled group efforts in politics, an increase in conflict, and a desire of groups that lose in the political process to opt out of the state (Congleton 1996, pp. 87–90). Alternatively, some groups are so effective at providing a foundation for exchange that they receive very high returns, while others receive very low returns. Fear, envy, and hatred arise among members of the low-yield groups, who are in turn stigmatized by the high yielders (Wintrobe 1996, pp. 53–54). (This is one of many arguments to the effect that conflict proceeds from ethnic inequality.) Conflict is thus the dysfunctional outcome of the functions groups are called on to perform.

In these utilitarian accounts, there is no independent role for affect and no explanation for why group boundaries follow one set of lines rather than another. To show that groups already in being provide services to their members is not to show that services are the *raison d'être* of those groups. Services can be demanded of and performed by groups that come into being for quite different ends. That available institutions are used for some purpose is not evidence that they no longer fulfill any other. Moreover, the causal links are doubtful. Even without group-provided services, interethnic political competition might occur, and fear, hatred, and envy might arise.

7. The reduction of transactions costs also informs a quite different interpretation of group behavior (Hardin 1995). On this view, the ethnification of politics entails a coordination game, in which group members emphasize their ethnic identity in the expectation that other group members will do the same. Various rewards follow from acting tacitly in concert. Group boundaries might run arbitrarily along one or another dimension, ethnic or nonethnic, and ethnic identity might not take hold at all (Hardin 1995, p. 51). But once it does coordination begins, and coordination slides easily into conflict and violence. All that is involved is tipping—that is, passing a certain point of escalation, often “from more or less random shocks” (Hardin 1995, p. 155). Interethnic conflict and violence thus derive from accidents of social relations (compare Fearon and Laitin 1996), rather than from feelings of antipathy, for ordinary people obtain their demonic ideas about others from manipulative elites (Hardin 1995).

In many severely divided societies, however, the evidence is the other way. In some, the attitudes of elites are at least as hostile as those of their followers (Horowitz 1997, pp. 439, 457 n. 31), but in many other societies hostile ideas arise out of interethnic juxtapositions that make it difficult for political leaders to transcend those ideas. Some political elites, cross-pressured to be moderate on ethnic issues because of an externally derived universalism, have found, to their enduring disadvantage, that they are soon outbid on the extremist flanks by leaders more in tune with widely distributed hostile attitudes (Rabushka and Shepsle 1972; Milne 1980).¹

8. Elite competition and the actions of “ethnic entrepreneurs” drive ethnic conflict (Brass 1985, 1996, 1997; compare Vail 1989, p. 11; and Kasfir 1979).

Elites manipulate ethnic identities in their quest for power, and they “construct” ethnic conflict (Brass 1997).

Why does interelite conflict proceed along ethnic lines, and why do the followers of elites follow them if the benefits flow solely to the elites whose interest motivates the struggle? Perhaps the followers do not know that they are being manipulated, but this requires proof. It is especially unlikely that people will go to war or participate in a deadly riot in behalf of the interests of others. Not surprisingly, the elite competition view has been criticized for creating an inaccurate image “of evil politicians and innocent masses” (Kakar 1996, pp. 150–51), for leaving too little room for individual agency on the part of ordinary people who engage in conflict behavior (Pandey 1992, p. 41).

9. Rational-choice theorists, motivated by the wars in the former Yugoslavia, have focused on ethnic warfare, but they have not always felt obliged to offer a theory of ethnic conflict in general. For them, ethnic war is produced by the insecurity that emerges when an actor is unsure of the intentions of another actor and the two are already mutually hostile (de Figueiredo and Weingast 1997; Fearon 1994; Hechter 1995; Lake and Rothchild 1996; compare Posen 1993).

A virtue of this conception is that its emphasis on the felt necessity of violence requires an understanding of how ordinary people come to be convinced that violence is unavoidable. Their fears move to the foreground of analysis.

In one variant (Fearon 1994), ethnic war results from the fear of a minority that it cannot trust the majority’s guarantees that it will not abuse power, to the acute disadvantage of the minority, under a new regime. Separatist warfare is thus preventive war, in which a minority decides to secede at the outset rather than chance participation in a regime that would weaken its position and make secession difficult later.

This predicament is produced by a failure of credible commitment. Buried in the formulation of the commitment problem is an assumption of group intentions: groups would prefer to live together on mutually advantageous terms, if only they could exchange commitments about those terms that were not subject to renunciation. But perhaps the problem is not that the majority is unable to commit itself credibly to respect minority interests. Perhaps it does not wish to do so. Perhaps the majority prefers to live in a state in which it dominates the minority, and it supposes—usually correctly, to judge by the net results of separatist warfare around the world—that it can handle minority resistance and so need make no concessions at the outset. Embedded in the commitment view of ethnic warfare, in other words, is an understanding of ethnic conflict that is more benign than the one that may well animate the participants.

In strategic-dilemma formulations, the role of interethnic antipathy in causing conflict is rejected. Take the issue of atrocities and what they might

signify about hostility. Fearon (1994) speculates that atrocities are not produced by affect but are intended to make later cohabitation impossible by deliberately deepening hatred or by generating such fear of the target group's revenge among moderate members of the attacking group that they will no longer be able to live in close proximity to members of the target group. But if future hatred can be engendered by inflicting atrocities, why is it not plausible to consider that current atrocities are the result of a hatred engendered earlier, rather than the result of a deliberate strategy? Do the emotions produced by a strategic use of violence dissipate, so that they do not become independent variables in later episodes of conflict? And if they do, of what use is the strategy of engendering hatred? The antipathy to antipathy leads to an argument that proves too much.

10. The strength of ethnic emotions, including the willingness to risk death for the sake of group interests, has led some writers to develop theories of ethnic conflict based on evolutionary conceptions of kin selection and genetic imprint (van den Berghe 1981; Shaw and Wong 1989). Stern (1997) has proposed a genetically encoded "primordial sociality" that he asserts promotes inclusive fitness. Groups, he posits, gain an adaptive advantage when they can develop rules that proscribe selfish individual behavior. The deep primacy of group interests leaves them open to emotional appeals about threats to those interests. Although ethnic groups may not have an evolutionary advantage over other groups, political leaders make ethnic groups objects of primordial attachment by tapping sentiments initially developed for such attachments.

Others have gone farther, suggesting that groups inclined toward conflict and warfare have tended to be successful in competition with others (for a summary, see LeVine and Campbell 1972, pp. 72–80). Brewer (1997a) has also argued that group boundaries have evolutionary utility, because individuals historically were unable to survive alone. The benefits of cooperation, however, decline as groups expand; and so there is an optimal level of group distinctiveness beyond which groups lose the loyalty of their members (Brewer and Miller 1996, p. 46). Outside the boundaries lie ethnocentrism and hostility.

If confirmed, these theories might provide an understanding of the passionate attachment of individuals to ethnic groups. By stressing the functionality of intergroup aggression, some versions might provide clues to sources of conflict. But it is always dangerous to reason backward from a contemporary, recurrent phenomenon to provide it with an evolutionary rationale that pins current behavior to hard-wired functionalism. Deep sociality and tendencies to identify with ascriptive groups certainly exist, but it is not clear that their manifestations in ethnic conflict behavior can be explained by phylogenetic evolution. Even adaptive behavior can be manifested in maladaptive ways, and maladaptive behavior can persist over very long historical periods. If ethnic conflict could be shown to fulfill adaptive functions, it would still be wrong to

think that these functions could not be fulfilled in alternative ways, given the problem-solving capacity possessed by human beings.

The various approaches reviewed here do not build on each other so much as they compete with and purport to displace each other. They run the gamut from passionless to passionate, from structural to strategic, from rational to irrational, from culturally primordial to merely constructed and now to hard-wired primordial. They do not provide evidence of a field that has produced a great deal of cumulative knowledge.

Theoretical overclaiming is in evidence here.² Portions of the phenomenon of ethnic conflict have been discovered or rediscovered and then asserted to be characteristic of the whole phenomenon. If there are powerful emotions, or if ethnic group loyalty supersedes loyalty to other groups, all ethnic conflict must be primordial. If politicians benefit from calculative behavior, or if groups struggle over resources, ethnic conflict must be instrumental. Clearly, what is needed is a theory that can embrace the disparate manifestations of ethnic conflict behavior.

It would be idle to pretend to be able to accomplish in one clean exercise what has not been done in many iterations. To indulge in such a pretense would be to recapitulate the very overclaiming that has produced the current disarray. But it is not idle to try to deal with some of the bite-sized issues into which the larger issues can be decomposed.

Central to any such analysis must be a consideration of the foundations of ethnic loyalty, of the relation between passion and interest, and of the significance of institutional context. A number of other issues, such as the roles of individual and collective gratifications, can be touched on along the way, but these basic issues get to the heart of the current dissensus.

Sociality and Ethnicity

If we have come full circle from cultural primordialism to evolutionary primordialism, it is because the power of ethnic affiliations stubbornly presses in on us and demands explanation. Such an explanation resides in the need of individuals to belong to groups. Individuals require the cooperation that groups provide. They possess a deep sociality.

That sociality, however, is not maximally inclusive. Whenever groups form, their members sense the existence of boundaries that divide them from other groups. Both the impulse to form groups and the impulse to differentiate them from others are so strong that they are easily activated. No sense of birth connection, no sense of common history, no sense of prior intracategory similarity is necessary at the outset. In a laboratory setting, random assignment to categories will activate a sense of groupness (Tajfel 1981). On small, homogeneous islands, similar cleavages develop among people living under identical conditions (Firth 1957). Whatever the exact sources, the tendency to cleavage is so well established as to be undeniable.

With cleavage comes comparison. Ingroups are said to have certain qualities that distinguish them from outgroups. Generally, insiders evaluate their own collective

qualities more highly than those of outsiders (see, for example, Edelstein 1972 and Dubb 1974). They also evaluate their own products as superior, even when detached third-party judgments suggest otherwise (Ferguson and Kelley 1964), and they apportion rewards so as to favor themselves and to disfavor outsiders, even when it costs them some portion of the reward to create the intergroup difference in apportionment (Tajfel 1974). They behave this way even in the absence of any hint by experimenters of intergroup competition and even in the face of the possibility of positive-sum outcomes if they wish to choose them. They use their perceptual apparatus to categorize people into classes and to exaggerate the similarities among themselves and their differences from others, a phenomenon well known in social judgment theory as assimilation and contrast effects.³ More than this, groups are given to in-group bias, and they appear to be motivated by a desire for favorable collective evaluation.

Although, as we have seen, some theories suggest that groups derive value from their ability to satisfy the goals of individuals within them, the opposite appears to be the case: individuals derive value from the groups to which they belong (Brewer 1997b, p. 205; 1991, p. 476). Indeed, they derive satisfaction from the success of the group, even when their own contribution to that success is palpably absent (Brewer 1979, p. 322). By the same token, the willingness of individuals to sacrifice for group interests and participate in collective action is predicted more by a sense of collective deprivation than by individual deprivation (Brewer 1991, pp. 478–79). The improvement of the group's condition, in other words, may be a more powerful motivation for participation in collective effort than is improvement of the participating individual's condition. This finding casts doubt on the aptness of methodological individualism as a starting assumption—or at least the sole starting assumption—in understanding group dynamics.

Given the compelling power of group affiliations, it is not surprising that individuals should find them useful vehicles for pursuing their own interests as well; it would be surprising if they did not. It is not possible, however, to reduce groups to their fulfillment of individual goals or to reduce affect to instrumental behavior. Both are present, as I argue in more detail later.

So far, the discussion has pertained to groups but not specifically to ethnic groups. Members of ethnic groups seem to partake of all of these tendencies to cleave, to compare, to specify inventories of putative collective qualities, to seek a favorable evaluation, to manifest in-group bias, to exaggerate contrasts with outgroups, and to sacrifice for collective interests. Ethnic groups also seem to carry matters further. They appear frequently to engender more loyalty from their members than competing group types do and to engage in severe conflict with other ethnic groups.

To appreciate the special power of ethnic loyalty, one must view ethnic groups both as groups like all others and as groups possessing qualities that not all other groups possess in the same measure. The most important of these qualities is a strong sense of similarity, with roots in perceived genetic affinity, or early socialization, or both. Belonging to a group implies thinking of oneself as possessing char-

acteristics that are somehow representative of the social category that the group embraces (Brewer and Miller 1996, p. 22). This feature, it can be hypothesized, is what leads individuals to submerge their own identities in the collective identity, to favor in-group members and make sacrifices for them. One sees oneself, so to speak, in other group members. Similarity engenders empathy and in extreme cases even obliterates the boundary between one individual group member and another.

There is ample evidence that people like others whom they believe to be similar to themselves in tastes, attitudes, and values (Byrne, Clore, and Smeaton 1986; Stephan 1973; Byrne 1969). They are also attracted to those they believe like them in turn, but this seems to be because "interpersonal attraction is a very sound basis for assuming commonality; it results in exaggerated perceptions of similarity" (Hogg and Turner 1985, p. 61). People assume that in-group members are similar to each other—that, after all, is the result of assimilation effects—and that assumption strengthens their attraction to them, even when they have no actual knowledge of their qualities (Brewer and Miller 1996, pp. 29–31). The irony here is that egocentrism leads people to favor those seen to be like themselves, and this leads beyond egocentrism to ethnocentrism and sacrifice—in other words, to a broader concept of self and self-interest.

Ethnicity is a powerful affiliation, both because similarity is valued and because genetic (or putatively genetic) origins and early socialization are potent sources of similarity or, in any case, of cues that signal similarity: appearance, customs, gestures, language, clothing, tastes, and habits. The assumption of similarity follows quickly once groups are formed, even in laboratories, and reinforces group identity. How much more powerful is the assumption of similarity when it derives from birth or common experiences of childhood.

Underlying genetic similarity and early socialization is, of course, the family. The family is the unit that provides constant replenishment of ethnic group members. It is the first group to which individuals belong; because of the long period of human maturation, its influence is extraordinarily durable. At an early age, children express fear toward strangers, and they learn the difference between family members and others. There is little doubt that they internalize the significance of the birth principle in cooperative organization and the significance of distinctions based on birth. In most conditions, they appreciate that, however attenuated, genetic connections are a surer source of similarity than are any others.

Ethnic groups, whatever their actual composition, purport to be founded on descent, and they, too, offer a greater sense of similarity than do groups founded on other premises. Given the general preference for people with similar attributes, this fact alone accounts for ethnic affinity. Experimental studies show that the greater the felt similarity within a group, the greater the degree of in-group bias (Brewer 1979, p. 318). Within ethnic groups, the high level of felt similarity should lead to expectations that in-group bias and differentiation from outgroups will also be great. The ascriptive character of ethnic affiliations accounts for their potency.

Among family choices bearing on ethnicity, marriage is perhaps the most important. Endogamy gives concreteness to conceptions of ethnic affinity and makes the

group a descent affiliation. Two subgroups that previously did not intermarry may begin to do so if they begin to see the line between them as insignificant compared with the line separating them from others. Although marriage patterns can and do change, there are spiral effects that flow from them. Groups in accelerating conflict tend to practice less exogamy. This leads to less porous boundaries between them, and the resulting attenuation in ties facilitates (or at least does not prevent) further conflict. Ethnic groups that are in severe conflict rarely have exogamy rates much higher than 10 percent. Groups with low levels of conflict (such as many in Latin America) have porous boundaries, both created by and facilitating exogamy, thus bolstering ties that can be a barrier to conflict.

In addition to being the fount of descent affiliations, the family is the source of behavioral lessons likely to be transferred to ethnic relations. Diffuse, unflinching mutual support and affection are the widespread model for family relations (however discrepant reality may be). It is likely that this model is invoked for ethnic affiliations that are grounded in the same birth principle and that grow out of family affiliations as well.

The extent of actual similarities of traits within ethnic groups and of actual differences between them is undoubtedly highly variable, but they need not be substantial. There may be fairly similar inventories of actual cultural attributes across group lines. Nevertheless, it is common for a few traits that distinguish one group from another to be built on and exaggerated as interaction proceeds (see, for example, Weiner 1978, pp. 240–41). Cleavage drives culture more than culture drives cleavage.

Notice that this account, although referable to birth, gives ample play to the social construction of ethnicity. Just as intergroup boundaries are constructed in the laboratory, so are they constructed in social life. The scope of group boundaries is not foreordained, and boundary change is common (Edel 1965; Young 1965; Bruner 1974; Vail 1989). In the process of expanding or contracting boundaries, cultural and political elites play their part, just as social constructivists claim, emphasizing those features in the situation of their followers, including those affinities and disparities that conduce to one or another definition of the group and its boundaries (Horowitz 1977).

Some social constructivists, however, go further, claiming that elites do not merely steer the process in the light of existing ethnic juxtapositions but that they more or less shape them, that the determination is made in the material interest of elites, and that they also have wide latitude to foment conflict and violence. These broader claims are highly contestable (see Kasfir 1979, pp. 375–76). The constraints of the field in which group interactions occur limit what elites can do and what interests they can pursue (Horowitz 1985, pp. 64–75). The strong perceptual basis of ethnic affinities and disparities is underappreciated by many constructivists. By the same token, the freedom of elites to foment conflict and violence is limited by their followers' definition of the situation and by what they would be willing to fight over. Hindu nationalists in India often attempt to incite attacks on Muslims, but they rarely succeed in the southern states of Kerala and Tamil Nadu, where caste

affiliations have more resonance than does the Hindu-Muslim polarity (Wilkinson 1997). Constructivism can survive without seeing ethnicity as an altogether opportunistic and infinitely malleable affiliation.

The account I have given also takes seriously the claims of primordialists that ethnicity is a thick, *Gemeinschaft* affiliation without acceding to the extreme claims of some primordialists to its mysterious, ineffable, invariably conflict-producing quality. To the contrary, a hard emphasis on the responsiveness of ethnic groups to the deep needs of group members is not at odds with a keen sense of the variability of ethnic phenomena (compare Eller and Coughlin 1993). It follows that group members may entertain sentiments so intense that theorists identify them as primordial, even though group identities are socially constructed, recently constructed, founded on relatively little in the way of palpable differences, and mutable as environmental conditions change. Intergroup sentiments can and do change. Georgians once viewed the Abkhaz as a people with a rich traditional culture; in recent years, these stereotypes have given way to a view of them as wild and uncivilized (Anchabadze 1997, p. 6). Some antipathies and some affinities are durable, and some are not. The determinants of such variations remain open to explanation.

Stereotypes of group characteristics, aptitudes, and disabilities amplify whatever intergroup differences have been identified. Many of the traits imputed to out-group members are threatening. The putative differences that accompany intergroup differentiation may motivate groups to mobilize. Not only do ethnic groups seek favorable evaluations, but the favorable outcomes they seek in competition are endangered by the presence of traits, such as diligence, clannishness, or intelligence, that they sense in their adversaries, even as they prefer their own mix of attributes. Where stereotypes are invidious, which is to say threatening to positive group evaluation, conflict is likely to be severe (Horowitz 1985, pp. 166–81).

If we return to the emergence of cleavage and ask why the tendency is so persistent, we might speculate that differentiation results from a decline in the benefits of cooperation as group size increases. But another explanation can be inferred from the behavior that immediately accompanies cleavage: group bias that seeks favorable evaluation through discriminatory action. Perhaps this behavior also manifests a desire to reap a disproportionate share of rewards in a given environment. Such a desire would not be at odds with the quest for a favorable evaluation, since the relative superiority of evaluation would provide a justification for unequal distribution. This desire would also comport with observed political behavior in ethnically divided societies, in which a part of the society claims to be the whole, places the status of ethnic strangers at sufferance, demands a variety of privileges, dominates the environment, and, if possible, establishes a status hierarchy.

Nothing I have said here, it should be emphasized, obviates the possibility that individuals may sense more affinity with groups other than the ethnic group into which they were born. Some people may find that, for them, occupational cleavages displace ethnic cleavages. On the similarity principle, professors may prefer the company of professors, physicians that of physicians, to the company of members of their own ethnic group or family. Some individuals may find that they have more

than one strong birth affiliation. A Belgian may be born Flemish and Catholic or Walloon and Catholic or Flemish and secularist in the Freemason tradition that, passed down from generation to generation, produces a birthlike affiliation in Belgium. Beneath these overarching levels of affiliation, there are likely to be subethnic ties as well. A Fleming from the southeast of Flanders may also see himself or herself as a Limburger and may view Antwerpers as possessing a certain unwonted arrogance.

These variations depend on particular social developments. In the first case, the growth of a complex set of educational structures may strongly differentiate academics and doctors from other occupational groups. In the second case, a history of religious warfare and then Enlightenment ideas produced *familles spirituelles* based on religious orientation, which were overlaid on, but did not correspond to, those based on language. In the third case, since many ethnic groups are amalgams (that is, are built from the ground up), the residue of sentiment attaching to the earlier, lower-level materials may be more or less salient, depending in large measure on how significant overarching conflicts are.

In many countries, the first and second variations—occupational identities (or class identities) and cross-cutting birth-based identities—do not exist. Occupational identity is a function of an industrial or post-industrial, highly complex division of labor. Cross-cutting birth-based identity is a function of specific histories. Multiple, sweeping continental movements of the sort that, because of their inconclusive results, produced multiple affiliations in Belgium or Switzerland are not universal. Still, they are present where, for example, Islam and Christianity or Islam and Hinduism cross-cut ethnic affiliations, as they do for the Yoruba in Nigeria and for many groups in India. But in many countries in which such developments did not break ethnic lines, ethnicity, while it always encounters competing affiliations, does not encounter affiliations quite so powerful in their pull. In such cases, a single cleavage can rend a polity. Usually, however, especially after relatively recent amalgamation, subethnic affiliations are available for invocation when the context shifts to a lower level. This fact, as I mention briefly below, provides a policy handle for the mitigation of ethnic conflict.

In short, a view of ethnicity as a strong affiliation is not incompatible with variation in its political saliency. History is not reducible, after all, to how members of one group see members of another, and it is a mistake of maximal primordialists to think otherwise. But the *Gemeinschaft* position, a minimalist primordialism, is not indefensible—although the term *primordial* has misleading overtones that ought to be avoided altogether.

Affective and Instrumental Behavior

It should be clear by now that, at least in this field, efforts to subsume passion in calculation or calculation in passion are futile (see *Rationality and Society* 1993). Neither the view that ethnicity is impervious to reason nor the view that ethnic conflict results solely from the suboptimal consequences of the rational behavior of

actors can be sustained. Bell's observation that ethnicity is salient "because it can combine an interest with an affective tie" puts the matter well, and it is easy to see that, once the affective side of ethnic affiliations is recognized, those affiliations will also become "a strategic site" (Bell 1975, p. 169). When it is clear that people give preference to members of their own group, members of that group will appeal to other members for such preference. Just as the family is simultaneously an emotional and an economic unit, so the ethnic group takes on instrumental tasks. It cannot, however, be described solely in terms of, or reduced to, the performance of those tasks, even though performance of such tasks should make the affiliation more rewarding and therefore more valuable than it was before (Patchen 1995).

The point is important for many reasons. One of these is that reduction of passions to interests can lead to a cribbed view of the emotions. In one account (Wintrobe 1996, pp. 53–54), fear, envy, and hatred arise essentially as a result of distributive shortfalls. But a good case can be made that the repertoire of ethnic emotions, conditions evoking them, and actions based on them are considerably wider. Petersen (forthcoming), for example, differentiates fear, hatred, and resentment, suggesting that they derive from varying conditions and are activated under varying circumstances of ethnic relations. Furthermore, to deny an independent role to affect might be to expect a proportionate (that is, reasonable) response to stimuli such as grievances. But ethnic conflict is frequently characterized by disproportionate responses: excessive fears of harm and excessive reactions to harm.

The reduction of instrumental to affective behavior has adverse consequences as well. In anger there can be reason, even if the response goes beyond reason. Experimental studies of aggression reveal that attacks believed to be unjustified arouse greater aggression than do justified or explicable attacks (Geen 1990, p. 44). People gauge the appropriateness of their level of anger by reference to justification, and they reason about justification.

Despite links between passion and interest, each can occupy a somewhat separate domain, or at least be present in variable proportions. Violence provides an excellent example, for there is abundant evidence that some interpersonal violence is impulsive and some is instrumental (Berkowitz 1974). Some psychologists emphasize instrumental rewards in generating aggression (Lange and van de Nes 1973), others the anger that follows from frustration (Fitz 1976). Whatever the exact balance, the two motives can be distinguished. In a study of Scottish prison inmates convicted of serious assaults, 76 percent were found to have engaged in impulsive aggression; only 18 percent had engaged in "controlled and planned aggression" (Berkowitz 1986, p. 97).

Personal dispositions play a role. Emotionally reactive people are likely to attribute hostile intentions to others and, on that account, to respond with aggression; they are sensitive to insults and easily provoked. In contrast, instrumentally aggressive people have personality profiles resembling those of nonaggressive people (Russell and Arms 1995). In a study of the *Seductions of Crime*, Katz (1988) notes the low rate of injuries in robbery, suggesting that robbery is an example of instrumental violence. But then he examines the violence that takes place in the

minority of robberies, concluding that it is not typically the result of panic or a transaction gone awry but is gratuitous, even “recreational violence.” It is, he says, a manifestation of a “commitment to be a hard man—a person whose will, once manifested, must prevail, regardless of calculations of practical self-interest” (Katz 1988, p. 187). The person who behaves in this way has chosen to cultivate a “distinctive morally insensitive will,” to be mean, to generate “dread” (Katz 1988, pp. 194, 135–38). Undoubtedly, people who cultivate reputations for excessive, unpredictable violence gain rewards that might not be available to others (see Schelling 1966, pp. 37–39, 118); for that reason, willfulness can become a strategy. But it is not always a strategy; it can and does proceed from impulsiveness or from deliberate immorality and “delight in violence” (Katz 1988, p. 138). The rewards that accrue may reinforce the behavior, but the personality inclined to needless violence is not likely to be impelled toward it in the first instance by the rewards.

There are some analogies here to ethnic violence. Deadly ethnic riots are large, passionate events, characterized by brutality and atrocities. There has been much speculation about the extent to which they are spontaneous or organized episodes (see, for example, Tambiah 1996; Brass 1996, 1997; and Wilkinson 1997). In the end, however, how could they be anything but both? A large-scale riot could scarcely be organized in the absence of a great deal of anger toward the targets. A riot that was the spontaneous expression of deep hostility to the targets would soon attract the attention of those who stood to gain from planning violence, and they would then surely attempt to foment later riots, drawing on (and undoubtedly trying to exacerbate) the hostile sentiments of those whose participation they wished to engage in violence. Passion might come first; organization could not succeed without it. But passion would attract organization. Interest can mobilize people along the lines of their passion but only if there is passion to mobilize around. Quite a lot of instrumental behavior is grafted onto attachments that have such value for people that they become magnets for mobilization.

Ethnic Behavior in Democratic Politics

Strategic behavior really comes into its own in the structured settings of democratic politics. Even here, however, there is a circular relationship between the situations that actors confront and those that they create. Changes in ethnic group boundaries respond strongly to the political environment, particularly the territorial frame, in which groups find themselves. Very often, in Asia and Africa, the approach of independence produced a tendency to create larger ethnic agglomerations, better suited to political competition in the whole territory. This tendency was called “supertribalism” or “artificial ethnicity,” as if to say that lower-level affiliations were more genuine.

In many cases, new identities at high levels of generality—such as North and South, Christian and Muslim, or Malay and non-Malay—were embraced, even though the component groups did not abandon lower-level identities for all purposes. Mobilization of one group to such new levels often created a need for others

to increase subethnic cohesion in response, but the responses were not perfectly symmetrical. The category of "northerner," or at least northern Muslim, came to have great resonance in Nigeria, but southerners remained divided among Ibo, Yoruba, and others; and the Yoruba and Ibo categories were subject to further fission. The strategic imperatives that produced the reconstruction of group boundaries did not work their will evenly among groups or do so free of a variety of social constraints.

In some states, the configuration of antecedent cleavages and the structure of the polity inhibited the growth of wider identities. In those states, such as Tanzania, group identities remained at much more local levels, and ethnic conflict, although not absent, did not divide the state down the middle (Young 1976, pp. 216–73). Much the same was long true of India, although the newly nationalized Hindu-Muslim cleavage now threatens to rend the society (Hardgrave 1993).

It is useful to contrast the range of variation outside of Asia and Africa. At one end are largely fluid societies, often with immigrant origins, that contain groups that historically had initial difficulty securing acceptance but came to be accepted as the immigration of still newer groups facilitated their incorporation. This is true in large measure of groups of European origin in Australia, France, New Zealand, and the United States. In the middle of the continuum are societies in which ethnic groups compete as groups but in a complex setting that mitigates the singleminded character of such conflicts. In these countries—Belgium, Canada, and Switzerland, all of them close to powerful neighbors and all of them federations—external forces historically fostered internal integration, but in the context of multiple lower-level loyalties. To return to a point made earlier, religious warfare, the Enlightenment, and the Industrial Revolution gave rise to powerful cleavages that compete with the ethnic cleavage: Protestant-Catholic, religious-ant clerical, social class, and in some countries region or canton. Ethnic issues broke into the party system late in relation to these established cleavages and never quite displaced them or the political parties to which they had given rise. Party politics is thus not a perfect reflection of ethnic conflict. States in which such complexity has produced what are essentially multiple-birth or near-birth affiliations are surely advantaged over those, such as Sri Lanka or Kenya, in which one ethnic group confronts another, external forces are neutral or disintegrative, and parties reflect just one cleavage line. This is the other end of the continuum.⁴

In Asia, Africa, Eastern Europe, and the former Soviet Union, many more states are characterized by a predominant ethnic cleavage than by multiple, alternating ethnic axes of conflict. That many of these states also have two or three large ethnic agglomerations confronting each other is due to a combination of social perception and strategic imperatives in a bounded environment.

That social perception is important is suggested by the fact that, despite a tendency to expand group identity to a level of up to half the population of a state (see Chai 1996), many smaller groups survive although they do not approach the 50 percent mark, and in some cases subgroups actually split off from larger groups of which they formerly were a part. Not all boundary change is upward.

Still, there is no denying the utility of larger affiliations or a general trend toward them. That strategy, as well as social perception, is involved is indicated not merely by the trend toward larger groups but also by a trend toward bifurcation of polities and the frequent proclivity (albeit far from ineluctable) for tripolar conflicts to become bipolar, as groups that are smallest or most divided internally are pressured to align with one of the other two groups.

The dynamics of bifurcation can be understood by adverting to a phenomenon first elaborated, without reference to ethnicity, by Riker (1962, pp. 32–101). Riker exposed the distributive advantages underlying the creation of minimum winning coalitions, those enjoying the support of just over half the electorate. Supporters of the minimum winning coalition could expect significantly greater rewards than they would receive if the winning party had to satisfy 80 or 90 percent of the electorate. Majoritarian rules for the formation of governments create no incentive to seek majorities larger than the required 50 percent plus one.

In the case of ethnic bifurcation, majoritarian political institutions cause a special problem. Segmented electoral markets, each exploited by ethnically based political parties, mean that prospects for alternation in office are thwarted by ascriptive affiliations.

Ironically, of course, majority rule was intended to mean something different, an exercise of choice. Instead, in divided societies, majority rule usually means group head-counts, and it encourages the formation of groups whose heads can be counted to reach the majority threshold. Brubaker (1996, p. 24) perhaps puts the matter too strongly when he says that political context does not merely constrain but constitutes; ethnic groups and nations, he contends, are cultural forms “governed by the properties of political fields, not by the properties of collectivities” (p. 17). We know too much by now of the properties of collectivities to accede to the hyperbole, but it is certainly true that groups are strongly responsive to the power of political fields, and Brubaker is right to stress the point.

What I have said so far puts general prescriptions of democracy as a solvent for ethnic conflict (compare Collier in this volume) in a problematic light, for serious ethnic conflict implies that majorities will want to dominate and that alternation in office is improbable once parties crystallize along ethnic lines. Democracy can work to reduce ethnic conflict, but only if measures are taken to encourage interethnic compromise.

Ethnically based parties, with orientations directed exclusively toward the welfare of the group that comprises the bulk of their members, arise in divided societies for many reasons. Among them are the fact that parties build on antecedent social relations and the tendency to see ethnic relations in competitive, even zero-sum terms. But there is another, recurrent reason embedded in the formal institutions to which parties respond. As Bates (1974, pp. 470–71) has pointed out, legislative constituencies in most countries are ethnically homogeneous. Once they begin that way, they are likely to stay that way in subsequent electoral boundary delimitations. Even in some countries in which groups tended to be intermixed, the number of heterogeneous constituencies declined soon after independence and kept declining later

(Lim 1997). It is safe and rewarding for politicians to make ethnic appeals to homogeneous electorates; it is more difficult to craft appeals to heterogeneous electorates. Once legislators are elected on the basis of such appeals, they will resist changes in constituency composition, but winners in heterogeneous constituencies will not resist changes in them, unless—improbably—they have found a formula that enables them to secure votes across group lines in such a constituency.

Politicians who have made ethnic appeals to secure their election from homogeneous constituencies will, of course, find ethnically based parties congenial, and they will not be attracted by the idea of securing election or reelection in such constituencies if they belong to a multiethnic party that, to survive as such, has to make compromises inimical to the untrammelled interests of their constituents. There are, in short, bottom-up reasons for ethnically based parties once constituencies are homogeneous.

Policymakers interested in promoting interethnic compromise have occasionally sensed the need for heterogeneous constituencies. Some countries have experimented with functional equivalents, requirements that candidates secure secondary support from differently composed constituencies adjacent to theirs. In 1978, Nigeria required that a successful candidate for president receive an overall plurality plus minimal geographic distribution of support among the Nigerian states. The latter requirement was designed to provide the functional equivalent of heterogeneous support. The winning candidate in the 1979 election managed to meet the dual requirements and, accordingly, behaved moderately in office. The same could not be said of the vast majority of legislators, whom the policymakers had not thought to make dependent on multiethnic support. They came overwhelmingly from homogeneous constituencies and pursued ethnically exclusive policies in office.

In country after country, political parties that began with multiethnic support have become ethnically based, irrespective of the initial wishes of party leaders and often in explicit contradiction of their ideas about the desirability of multiethnic organization. Panethnic nationalists and socialists have been reduced, against their wishes, to the status of ethnic group leaders. Again and again, the competitive configuration is what governs. If group members want a differential apportionment of rewards—including rewards of status, prestige, and symbolic recognition—and if the modern state is able to provide these tangible and intangible rewards, interethnic compromise will not provide what group members want.

Nevertheless, leaders who engage in interethnic compromise could do so with impunity if they could forestall the growth of ethnically based opposition on their flanks. Short of outlawing opposition—which indeed occurs in response to this problem—they are unable to do so. If they wait until ethnically based opposition appears, as Sinhalese party leaders did in Sri Lanka in the 1950s, they will then need to compete with that opposition with the handicap of appearing tardy in taking up the ethnic cause and insincere in its espousal, a handicap that those Sinhalese leaders did in fact encounter. Hence it is preferable to move early, occupy the ethnic flank, preempt the competition, and try to inhibit its formation altogether, thus pre-

serving monopoly control over the votes of group members. Monopoly control has the additional advantage of quite possibly limiting the degree of ethnically virulent claims that will come with sharp interparty but monoethnic competition on the ethnic flank. With homogeneous constituencies, early action is strongly advisable.

This logic applies equally to the leaders of multiethnic parties who, if they opt instead to leave and lead ethnically based parties, will end up in a permanent minority position even if they capture all of their group's support, simply because their group constitutes a minority. The alternative in both cases, once competition on the flanks gets going, is to be left with no support at all if these leaders remain with the multiethnic party. Better to capture the support of one's own group, become a leader of a minority party, and hope—generally futilely—for the day when a new multiethnic party or a new coalition can be formed.

Typically, therefore, multiethnic parties dissolve before polarization is complete. Their dissolution, however, usually represents a large step on the road to polarization.

All of this highlights a point made earlier in connection with claims that members of ethnic groups learn their antipathy toward other groups from manipulative leaders. Here we see that even the preferences of leaders for multiethnic parties yield to the danger that their supporters will abandon them for monoethnic parties. It is most certainly not the case that all leaders are less bigoted than their followers, but it is equally true that their followers do not need to be taught whom to be hostile to or how to make claims in politics that will disadvantage those they see as ethnic enemies.

This brings us to the question of why politics takes this sort of centrifugal turn. Most of the answer is already in place. As Rabushka and Shepsle (1972, pp. 62–86) explain, intense preferences, distributive politics, and an oversized (more than majority) coalition lead to strongly centrifugal tendencies. All that is missing is the political entrepreneur who, excluded from leadership in the multiethnic party, proceeds to make extreme ethnic claims. The presence of such politicians is the reason that compels leaders of multiethnic parties to move—in anticipation if possible, in reaction if necessary—to cultivate support on the flank where the political entrepreneur will be located. The alternative course, to deliver preferred goods by interethnic compromise, is foreclosed by the intensity of preferences along a single axis of ethnicity, which makes the exchange of one value for another a very difficult undertaking (Rabushka and Shepsle 1972, p. 84). In-group bias, majoritarian institutions, and homogeneous constituencies thus generally produce ethnically based parties, and the attempts of party leaders to compromise produce reactions on the flanks.

None of this is inevitable. Elsewhere I have tried to explain that institutions can alter conflict outcomes (Horowitz 1989), that some electoral institutions are more hospitable to maintaining interethnic moderation than are others because they counter the centrifugal incentive structure of politicians (Horowitz 1991, pp. 163–203), that devolution to substate units can mitigate bifurcation by activating lower-level identities (Horowitz 1985, pp. 601–13), and that, on the other hand, some decision rules actually foster ethnic exclusion and can even promote minority

rule (Horowitz 1993, pp. 28–31). These variations create opportunities for the amelioration of conflict. However, under most conditions, leaders are likely to be punished for opting for ameliorative institutions, just as they are punished for embarking on a compromise course in general. There is a large body of literature on ethnic conflict management (see, for example, McGarry and O’Leary 1993), but little systematic work has been done on the timing and conditions conducive to embarking on a course of conflict management.

Speculations on Synthesis

I claimed earlier that various perspectives on ethnic conflict were not as mutually incompatible as they might seem. I suggested, too, that while passion and interest sometimes operate separately, there are reasons why they are often found together (those with interests seek to harness passions). One can now also connect patterns of group formation and group loyalty to ethnic political behavior in the institutional framework of the democratic state.

We have seen that ethnic groups are formed and re-formed in bounded fields, based on a sense of similarity of character and of common fate. The tendency of people to form groups, to cleave from others, to compare themselves with others, and to discriminate against others implies, to say the least, a less than maximal sociality. Here we have seen that the democratic state provides a frame within which ethnic conflict is played out, within which groups seek the practical benefits of ingroup bias. In democratic politics, too, maximal inclusiveness is a strongly disfavored outcome, and there are strong incentives to maximize differences in the issue positions and the rewards obtained by groups.

To be sure, some intergroup differentiation has taken place in anticipation of competition within the political system of the state (Chai 1996). If this were all there were to it, if groups were only organized for politics in institutional settings, the isomorphism of the two processes would be illusory, for the two would amount to the same thing. Indeed, both might be reducible to Riker’s minimum winning coalition theorem. But what is striking is that intergroup differentiation has the same characteristics even in noninstitutional settings, including experimental settings. This might lead, upon further exploration, to a conclusion that the minimum winning coalition theorem is actually an instance of the more general human tendency to cleave, compare, and apportion rewards in a biased fashion.⁵

I do not mean to make such a claim with any special vigor. I advance it merely as a speculation to undermine further the prevailing wisdom that processes characterized as either affective or calculative do not have some prominent underlying similarities. One could make an even more radical claim—that the attraction of analysts to seemingly irreconcilable hard and soft positions on all these issues is itself a manifestation of the same underlying propensities to cleavage, comparison, and self-definition by opposition. If I am right that there is less mutual exclusivity than is supposed by those who have taken the opposing positions, then this speculation would be of interest far beyond the field of ethnic conflict.

Be that as it may, the current state of affairs demands a new synthesis. I am sure I have not produced any such synthesis here, and I probably have not done justice to any of the prevailing positions in attempting to articulate a medium-hard, medium-soft, at once structural and strategic perspective, but I hope I have shown that there is promise in the combination.

Notes

1. For helpful observations on elites and masses, see chapter 2 of Petersen (forthcoming).
2. Some of the writers I have cited, such as Fearon (1994), are careful to confine themselves to a particular slice of conflict behavior, such as separatist warfare. Their analysis, however, is founded on broader assumptions and has broader implications for conflict theory, which is why I placed them where I did.
3. Where values are arrayed along a continuum, subjects simplify the array by compressing small differences to something approaching their median value (assimilation effects) and by exaggerating larger differences (contrast effects) (see Eiser 1990). Put another way, differences judged to belong to the same class (for example, heavy) are perceptually reduced, while those judged to belong to a different class from the first (for example, light) are judged to be more different from the first class than they actually are. These phenomena constitute intra-class assimilation and interclass accentuation (see Zebrowitz 1990).
4. This paragraph draws on Horowitz (1993).
5. After all, one of the advantages of the ingroup-outgroup models I have described is that they are sensitive to context and so fit well with the state's framing of interactions.

References

- Anchabadze, Yurii. 1997. "Georgia-Abkhazia: The Hard Road to Agreement." Paper presented at the conference Georgians and Abkhazians: the Search for Agreement and the Role of the International Community, June 12–14, Brussels, Belgium.
- Bates, Robert H. 1974. "Ethnic Competition and Modernization in Contemporary Africa." *Comparative Political Studies* 6 (4): 457–84.
- Bell, Daniel. 1975. "Ethnicity and Social Change." In N. Glazer and D. P. Moynihan, eds., *Ethnicity: Theory and Experience*. Cambridge, Mass.: Harvard University Press.
- Berkowitz, Leonard. 1974. "Some Determinants of Impulsive Aggression." *Psychological Review* 81 (2): 165–76.
- . 1986. "Some Varieties of Human Aggression: Criminal Violence as Coercion, Rule-Following, Impression Management, and Impulsive Behavior." In A. Campbell and J. J. Gibbs, eds., *Violent Transactions: The Limits of Personality*. Oxford: Basil Blackwell.
- Bonacich, Edna. 1972. "A Theory of Ethnic Antagonism: The Split Labor Market." *American Sociological Review* 37(5): 547–59.
- . 1973. "A Theory of Middleman Minorities." *American Sociological Review* 38 (5): 583–94.
- Brass, Paul R. 1985. *Ethnicity and Nationalism: Theory and Comparison*. Newbury Park, Calif.: Sage Publications.
- . 1996. *Riots and Pogroms*. London: Macmillan.
- . 1997. *Theft of an Idol: Text and Context in the Representation of Collective Violence*. Princeton, N.J.: Princeton University Press.

- Brewer, Marilynn B. 1979. "In-Group Bias in the Minimal Intergroup Situation. A Cognitive-Motivational Analysis." *Psychological Bulletin* 86 (2): 307-24.
- . 1991. "The Social Self. On Being the Same and Different at the Same Time." *Personality and Social Psychology Bulletin* 17 (5): 475-82.
- . 1997a. "Social Identity, Group Loyalty, and Intergroup Conflict." Paper presented at the meeting of the Midwest Consortium for International Security Studies, December 5, Oak Brook, Ill.
- . 1997b. "The Social Psychology of Intergroup Relations: Can Research Inform Practice?" *Journal of Social Issues* 53 (1): 197-211.
- Brewer, Marilynn B., and Norman Miller. 1996. *Intergroup Relations*. Buckingham, England: Open University Press.
- Brubaker, Rogers. 1996. *Nationalism Reframed: Nationhood and the National Question in the New Europe*. Cambridge, U.K.: Cambridge University Press.
- Bruner, Edward M. 1974. "The Expression of Ethnicity in Indonesia." In C. Abner, ed., *Urban Ethnicity*. London: Tavistock.
- Byrne, Donn. 1969. "Attitudes and Attraction." In L. Berkowitz, ed., *Advances in Experimental Social Psychology*. New York: Academic Press.
- Byrne, D., T. L. Clore, and G. Smeaton. 1986. "The Attraction Hypothesis: Do Similar Attitudes Affect Anything?" *Journal of Personality and Social Psychology* 51(6): 1167-70.
- Chai, Sun-Ki. 1996. "A Theory of Ethnic Group Boundaries." *Nations and Nationalism* 2(2): 281-307.
- Congleton, Roger D. 1996. "Ethnic Clubs, Ethnic Conflict, and the Rise of Ethnic Nationalism." In Albert Breton, Gianluigi Galeotti, Pierre Salmon, and Ronald Wintrobe, eds., *Nationalism and Rationality*. Cambridge, U.K.: Cambridge University Press.
- Connor, Walker. 1993. "Beyond Reason: The Nature of the Ethnonational Bond." *Ethnic and Racial Studies* 16 (3): 373-89.
- . 1994. *Ethnonationalism: The Quest for Understanding*. Princeton, N.J.: Princeton University Press.
- de Figueiredo, Rui, and Barry R. Weingast. 1997. "The Rationality of Fear: Political Opportunism and Ethnic Conflict." Stanford University, Department of Political Science, Palo Alto, Calif.
- Dubb, A. A. 1974. "The Impact of the City." In W.D. Hammond-Tooke, ed., *The Bantu-Speaking Peoples of South Africa*. London: Routledge & Kegan Paul.
- Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112 (4): 1203-50.
- Edel, May. 1965. "African Tribalism: Some Reflections on Uganda." *Political Science Quarterly* 80 (3): 857-72.
- Edelstein, Melville Leonard. 1972. *What Do Young Africans Think?* Johannesburg: South African Institute of Race Relations.
- Eiser, J. Richard. 1990. *Social Judgment*. Pacific Grove, Calif.: Brooks/Cole Publishing.
- Eller, Jack David, and Reed M. Coughlin. 1993. "The Poverty of Primordialism. The Demystification of Ethnic Attachments." *Ethnic and Racial Studies* 16 (2): 183-202.
- Fearon, James D. 1994. "Ethnic War as a Commitment Problem." Paper presented at the annual meeting of the American Political Science Association, September 3-7, New York.
- Fearon, James D., and David D. Laitin. 1996. "Explaining Interethnic Cooperation." *American Political Science Review* 90 (4): 715-35.
- Ferguson, Charles K., and Harold H. Kelley. 1964. "Significant Factors in Overevaluation of Own Group's Product." *Journal of Abnormal and Social Psychology* 69 (2): 223-28.

- Firth, Raymond. 1957. *We, The Tikopia*. London: Allen & Unwin.
- Fitz, Don. 1976. "A Renewed Look at Miller's Conflict Theory of Aggression Displacement." *Journal of Personality and Social Psychology* 33 (6): 725-32.
- Forbes, H. D. 1997. *Ethnic Conflict: Commerce, Culture, and the Contact Hypothesis*. New Haven, Conn.: Yale University Press.
- Geen, Russell G. 1990. *Human Aggression*. Milton Keynes, U.K.: Open University Press.
- Geertz, Clifford. 1963. "The Integrative Revolution: Primordial Sentiments and Civil Politics in the New States." In C. Geertz, ed., *Old Societies and New States*. New York: Free Press.
- . 1973. *The Interpretation of Cultures*. New York: Basic Books.
- Hardgrave, Robert L., Jr. 1993. "India: The Dilemmas of Diversity." *Journal of Democracy* 4 (4): 54-68.
- Hardin, Russell. 1995. *One for All: The Logic of Group Conflict*. Princeton, N.J.: Princeton University Press.
- Hechter, Michael. 1995. "Explaining Nationalist Violence." *Nations and Nationalism* 1 (1): 53-68.
- Hogg, Michael A., and John C. Turner. 1985. "Interpersonal Attraction, Social Identification and Psychological Group Formation." *European Journal of Social Psychology* 15 (1): 51-66.
- Horowitz, Donald L. 1977. "Cultural Movements and Ethnic Change." *The Annals* 433: 6-18.
- . 1985. *Ethnic Groups in Conflict*. Berkeley, Calif.: University of California Press.
- . 1989. "Incentives and Behaviour in the Ethnic Politics of Sri Lanka and Malaysia." *Third World Quarterly* 11 (4): 18-35.
- . 1991. *A Democratic South Africa? Constitutional Engineering in a Divided Society*. Berkeley, Calif.: University of California Press.
- . 1993. "Democracy in Divided Societies." *Journal of Democracy* 4 (4): 18-38.
- . 1997. "Self-Determination: Politics, Philosophy, and Law." *NOMOS* 39: 421-63.
- Isaacs, Harold R. 1975. *Idols of the Tribe*. New York: Harper & Row.
- Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press.
- Kakar, Sudhir. 1996. *The Colors of Violence*. Chicago: University of Chicago Press.
- Kasfir, Nelson. 1979. "Explaining Ethnic Political Participation." *World Politics* 31 (3): 365-88.
- Katz, Jack. 1988. *Seductions of Crime*. New York: Basic Books.
- Klitgaard, Robert. 1991. *Adjusting to Reality: Beyond "State versus Market" in Economic Development*. San Francisco, Calif.: Institute for Contemporary Studies Press.
- Lake, David A., and Donald Rothchild. 1996. "Containing Fear: The Origins and Management of Ethnic Conflict." *International Security* 21 (2): 41-75.
- Lange, Alfred, and Ad van de Nes. 1973. "Frustration and Instrumentality of Aggression." *European Journal of Social Psychology* 3 (2): 159-77.
- LeVine, Robert A., and Donald T. Campbell. 1972. *Ethnocentrism: Theories of Conflict, Ethnic Attitudes, and Group Behavior*. New York: John Wiley & Sons.
- Lim, Hong Hai. 1997. "The Malayan Electoral System: Its Formulation and Change." Ph.D. dissertation. University of Malaya, Faculty of Economics and Administration, Kuala Lumpur.
- McGarry, John, and Brendan O'Leary, eds. 1993. *The Politics of Ethnic Conflict Regulation*. London: Routledge.
- Meerman, Jacob. 1979. *Public Expenditure in Malaysia: Who Benefits and Why*. New York: Oxford University Press.

- Melson, Robert, and Howard Wolpe. 1970. "Modernization and the Politics of Communalism: A Theoretical Perspective." *American Political Science Review* 64 (4): 1112-30.
- Milne, R.S. 1980. *Politics in Ethnically Bipolar Societies*. Vancouver: University of British Columbia Press.
- Olzak, Susan, and Joane Nagel. 1986. "Introduction." In S. Olzak and J. Nagel, eds., *Competitive Ethnic Relations*. New York: Academic Press.
- Pandey, Gyanendra. 1992. "In Defense of the Fragment: Writing About Hindu-Muslim Riots in India Today." *Representations* 37 (winter): 27-55.
- Patchen, Martin. 1995. "Ethnic Group Loyalties and Societal Strife." *International Journal of Group Tensions* 25 (3): 227-45.
- Petersen, Roger. Forthcoming. *Fear, Hatred, Resentment: Delineating Paths to Ethnic Violence in Eastern Europe*.
- Posen, Barry. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (1): 27-47.
- Rabushka, Alvin, and Kenneth R. Shepsle. 1972. *Politics in Plural Societies*. Columbus, Ohio: Charles E. Merrill Publishing.
- Rationality and Society*. 1993. Special Issue on "Emotions and Rational Choice." 5 (2): 155-293.
- Riker, William R. 1962. *The Theory of Political Coalitions*. New Haven, Conn.: Yale University Press.
- Russell, Gordon W., and Robert L. Arms. 1995. "False Consensus Effect, Physical Aggression, Anger, and a Willingness to Escalate a Disturbance." *Aggressive Behavior* 21 (5): 381-86.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, Conn.: Yale University Press.
- Shaw, R. Paul, and Yuwa Wong. 1989. *Genetic Seeds of Warfare: Evolution, Nationalism, and Patriotism*. Boston, Mass.: Unwin Hyman.
- Shils, Edward. 1957. "Primordial, Personal, Sacred and Civil Ties." *British Journal of Sociology* 8 (1): 130-45.
- . 1995. "Nation, Nationality, Nationalism, and Civil Society." *Nations and Nationalism* 1 (1): 93-118.
- Stephan, Cookie. 1973. "Attribution of Intention and Perception of Attitudes as a Function of Liking and Similarity." *Sociometry* 36 (4): 463-75.
- Stern, Paul C. 1997. "Why Do People Sacrifice for Their Nations?" In J. Comaroff and P. Stern, eds., *Perspectives on Nationalism and War*. Amsterdam: Gordon & Breach.
- Tajfel, Henri. 1974. "Intergroup Behavior, Social Comparison and Social Change." Katz-Newcomb Lectures. University of Michigan, Ann Arbor.
- . 1981. *Human Groups and Social Categories*. Cambridge, U.K.: Cambridge University Press.
- Tambiah, Stanley J. 1996. *Leveling Crowds: Ethnonationalist Conflicts and Collective Violence in South Asia*. Berkeley: University of California Press.
- Tellis, Ashley J., Thomas S. Szayna, and James A. Winnefeld. 1998. *Anticipating Ethnic Conflict*. Santa Monica, Calif.: RAND Corporation.
- Vail, Leroy, ed. 1989. *The Creation of Tribalism in Southern Africa*. Berkeley: University of California Press.
- van den Berghe, Pierre L. 1981. *The Ethnic Phenomenon*. New York: Elsevier.
- Weiner, Myron. 1978. *Sons of the Soil: Migration and Ethnic Conflict in India*. Princeton, N.J.: Princeton University Press.

- Wilkinson, Steven I. 1997. "The Electoral Origins of Ethnic Violence: Hindu-Muslim Riots in India." Ph.D. dissertation. Massachusetts Institute of Technology, Department of Political Science, Cambridge, Mass.
- Wintrobe, Ronald. 1996. "Some Economics of Ethnic Capital Formation and Conflict." In Albert Breton, Gianluigi Galeotti, Pierre Salmon, and Ronald Wintrobe, eds., *Nationalism and Rationality*. Cambridge, U.K.: Cambridge University Press.
- Young, Crawford. 1965. *Politics in the Congo*. Princeton, N.J.: Princeton University Press.
- . 1976. *The Politics of Cultural Pluralism*. Madison. University of Wisconsin Press.
- . 1993. "The Dialectics of Cultural Pluralism: Concept and Reality." In Crawford Young, ed., *The Rising Tide of Cultural Pluralism*. Madison: University of Wisconsin Press.
- Zebrowitz, Leslie A. 1990. *Social Perception*. Pacific Grove, Calif.: Brooks Cole Publishing.

Comment on "Structure and Strategy in Ethnic Conflict: A Few Steps toward Synthesis," by Donald L. Horowitz

Robert H. Bates

Donald Horowitz presents us with his assessment of the state of the art in the study of ethnic conflict. He focuses on the roots of ethnic conflict, by which he means the psychology that impels people to identify with ethnic groups and to define their private interests in accord with the collective interests of such groups. I find myself in agreement with his main argument, which I take to be that we need no longer choose sides in the debates over the foundations of ethnic politics—that is, as to whether they are based on rational calculation or emotional commitment. Indeed, I regard the debate as tired and, based as it is on misleading notions of what rational actor models require, unproductive.

Let me take inspiration from Horowitz's efforts to move beyond debates about individual motivations and to address arguments about collective behavior and the implications of ethnicity for development. As Horowitz argues, ethnic groups are plastic and adaptive. Building on this insight, I would argue that they are economic organizations that promote and safeguard the process of investment. In the contemporary jargon, they constitute forms of social capital. This renders them economically productive, but it can also render them politically dangerous.

The Formation and Protection of Human Capital

One of the features of ethnic groups that has long confounded scholars is that such groups combine rural and urban populations. Given their rural roots, it has been tempting to view ethnic groups as traditional or atavistic. Given the urban focus of these groups, however, it would clearly be a mistake to do so. Rather, it makes more sense to regard ethnic groups as instruments for the formation of human capital. Ethnic groups forge links between generations. Within them elders invest in the young by financing their education and launching their careers. Often the elders live in rural areas; following their education, the young pursue opportunities in town. In a later period the young pay back the old by remitting a portion of their urban wages

Robert H. Bates is professor of government at Harvard University.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

to the rural areas, financing community improvements (such as schools or infrastructure), or providing collective services (such as political leadership and representation). The earlier sacrifices of the older generation thus bear fruit in later periods, as the young provide resources and services in return.

A notable feature of ethnic groups is that they employ family sentiment and cultural values to secure the return on such investments. The elders often gain control over the rites of passage: the ceremonies that accompany birth, death, and marriage. To receive the sacraments—figuratively and sometimes literally—those who have succeeded in their careers must preserve ties with their community by tithing and by observing ritual practices. The possibility of exclusion constitutes a threat that can be used to exact elite cooperation in collective enterprises. And the possibility of failing to receive the blessings of the community for oneself and one's family gives the community influence over those who, once successful, may be tempted opportunistically to renege on the implicit terms of the contract and fail to repay the debt they owe the community.

Not only do ethnic groups thus link flows of resources between generations; in many settings they also provide defenses for those investments. Particularly in areas where the public sector is weak, groups such as ethnic groups provide private means of securing property rights. As the accumulation of wealth increases, so too do the incentives to steal. Thus increases in wealth require increased investment in the means of coercion (Bates, Greif, and Singh 1998). Ethnic groups not only promote investment; they also promote the organization of protection.

Members of groups that are powerful are safe; so too are their possessions. The creation of such powerful groups provides security by promoting deterrence. Groups, including ethnic groups, secure the power to defend property and possessions by establishing reputations for violence and by instilling the belief that retaliation for wrongful acts is sufficiently certain that the expected benefits from malfeasance are small. But this solution suffers from several important defects. It is hair-triggered—to preserve reputations, infractions must be punished. It is costly—when there is an infraction, cycles of retaliation may result. And it generates perverse incentives—it encourages machismo styles of provocative behavior by those who are certain of the backing of their kin.

The Politicization of Ethnic Groups

Normally, ethnic groups are not politically dangerous. In every major city in the world, local politicians find ethnic organizations to be a useful means of detecting the interests of their constituents and organizing their followers. It makes little sense to see ethnic groups as inherently dangerous politically when so much normal, everyday politics takes place peacefully within the settings that they organize.

And yet we also know that ethnic groups can be highly destructive. What transforms ethnic politics from the benign form that political scientists love and economists despair of into the kind that we all find so horrible?

One cause of the transition into the "red zone" is that in majority-rule settings ethnic domination can be highly inefficient and yet immune from easy correction.

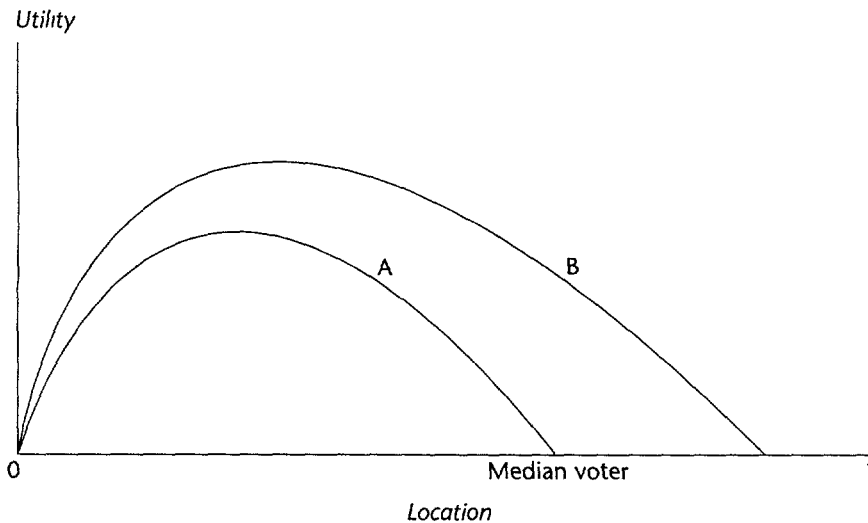
The result is that minorities are compelled to consume inferior public services and see no chance of gaining the power to upgrade them. To counter the impact of “permanent exclusion,” they therefore turn to alternative forms of political action, including violence. To see this, consider the following model:

1. Constituents are spread along an interval $[0,1]$. One group is concentrated at one end point, another at the other.
2. Politicians seek followers whom they can tax; they keep a portion of unspent public funds. One kind of politician enters from the left (0), the other from the right (1).
3. Followers value public goods and political identities; they prefer projects that are located “close” to their ideal points (or physical location) and politicians from their own group.
4. The farther a good is located from an end point, the more expensive it is to provide.

Under these circumstances a politician from a majority group can continue to win power—or at least the support of the median voter—even though he or she provides inferior services and pockets public revenues.

To see why, consider figure 1. One of the two utility curves—curve A—results purely from the provision of a public good. The good generates the maximum utility for the follower located at the point of its provision; benefits decline as the distance from that point increases. The second utility curve—curve B—includes the increased utility gained from having the provider of that good come from the same ethnic group as the follower who consumes it. Curve B can be shifted back and to the left and still generate the same level of utility as curve A to the median voter. A politician who offers inferior services to the median voter will nonetheless retain his or her support, given the advantage the politician draws from belonging to the same ethnic group.

Figure 1. Utility As a Function of Location



The provider of public goods would benefit from the provision of lower-value services: goods located closer to the end point are less costly to provide, leaving the politician with greater amounts of unspent—and therefore privately appropriable—public revenues. All followers residing to the right of previous locations of public goods would suffer, however, and those who do not place a positive value on the politician's ethnic label would gain no compensation for their loss. Despite the redistribution of benefits to the self-seeking elite and the loss of welfare to many citizens, the outcome nonetheless remains stable, for the median voter experiences no decline in welfare.

Under such a system, then, those not from the politician's ethnic group suffer losses. Because they cannot undermine the government's political support by censuring its public performance, they face permanent exclusion from the benefits of the political regime. Thus they may seek to secede from the political community—something the government will surely resist, if only because of the impact on its revenues. The result may therefore be violence.

From this perspective, then, the red zone becomes larger the more intense the ethnic preference and the stronger the preference for "proximate" public goods.¹ Increases in each variable generate increases in losses for citizens from minority groups and may heighten their conviction that they face permanent exclusion from the benefits of the political process. This dynamic appears to characterize political communities in the United States, where suburbs seek to secede from central cities (Miller 1973). The turn to violence appears more characteristic of developing countries, where secession sometimes leads to civil war (Gurr 1993).

The red zone is also marked by the size distribution of ethnic groups. Say, for example, that plurality voting prevails and that one group accounts for 48–49 percent of the electorate. Its leaders then know that if they can render their bloc solid and add another small group to their coalition, they will gain control over the state. Other groups know that if they can police their ranks, they can deny the dominant group adherents and therefore victory. Under such circumstances it pays to organize ethnically and, in particular, to use force to ensure group solidarity. Ethnic tensions rise as the larger group seeks to secure the allies that will allow it, the winner, to take all and as the smaller groups seek to prevent privately profitable defections. Such tensions would not arise if the largest ethnic group were able to supply a majority on its own or if the distribution of ethnic groups were so splintered that no group could reasonably aspire to capture power and render the government an agency of its will.

Post-independence Nigeria exhibited the relevant danger marks: the north's search for confederates with whom to ally in the seizure of power heightened political tensions and helped unleash the forces of Biafran secession (see Ademoyega 1981). In Uganda the prospect of Bugandan hegemony aroused ethnic political tensions and contributed to the rise of political violence in the post-independence period (Omara-Otunnu 1995). Tanzania, by contrast, possessed so many groups and Lesotho and Swazi so few that ethnic tensions and communal violence remained rare in those countries (see Bates 1973; Jackman 1978; and Collier in this volume).

Ethnic groups organize the pursuit of wealth by building economic organizations. They also organize the pursuit of power. While promoting investment, they can also seek domination. Under certain circumstances the struggle for power can become economically destructive, as groups turn from peaceful political competition to violence.

Ethnic Demobilization

Historically, the demilitarization of regional groups, such as ethnic groups, has been secured through the political distribution of economic rewards. Central political elites tamed regional factions by attaching their economic fortunes to the state and by conferring on their leaders the capacity to allocate resources—jobs, patronage, or other forms of political spoils—among their followers. In modern times, political elites have used their command and control over centrally planned economies to form multiethnic states in Eastern Europe and the Balkans, thereby bringing peace to a region once characterized by violence. In the developing world, political elites used state-led development programs to ensure the political loyalty of different groups. National politics became machine politics in both socialist and developing countries.

When competitive market forces caused the collapse of central planning and the abandonment of state-led development, politicians lost the ability to extract from markets the resources with which to organize political followings. That is, they lost the means of targeting economic resources politically. It is notable that countries, such as Colombia, that had long maintained conservative fiscal regimes experienced higher levels of regional violence and that nations, such as Yugoslavia, entered the red zone when moving away from central planning. Changes in forms of economic management appear to have left the central state with fewer means of co-opting powerful regional elites. Ethnic groups that generate and promote investment and wealth by providing economic organizations now also destroy it by promoting political violence.

Note

1. In the model increased corruption and graft accompany the strengthening of such sentiments.

References

- Ademoyega, Ademola. 1981. *Why We Struck: The Story of the First Nigerian Coup*. Ibadan, Nigeria. Evans Brothers.
- Bates, Robert H. 1973. *Ethnicity in Contemporary Africa*. Syracuse, N.Y.: Program of Eastern African Studies, Syracuse University.
- Bates, Robert H., Avner Greif, and Smita Singh. 1998. "Organizing Violence."
- Cohen, David. 1985. *Law, Violence, and Community in Classical Athens*. Cambridge, U.K.: Cambridge University Press.

- Gurr, Ted Robert. 1993. *Minorities at Risk*. Washington, D.C.: United States Institute of Peace Press.
- Jackman, Robert. 1978. "The Predictability of Coups d'Etat Using African Data." *American Political Science Review* 72: 1262-75.
- Miller, Gary. 1973. *Cities by Contract*. Cambridge, Mass.: MIT Press.
- Omara-Otunnu, Ammi. 1995. "The Dynamics of Conflict in Uganda." In Oliver Furley, ed., *Conflict in Africa*. New York: St. Martin's Press.

Comment on "Structure and Strategy in Ethnic Conflict:
A Few Steps toward Synthesis,"
by Donald L. Horowitz

Michael Chege

Donald Horowitz has made a laudable effort to demolish the conceptual divide separating theories of ethnic identity based on "hard" primordial attributes from those grounded on "soft" cultural engineering. Against this backdrop, I would like to reiterate the most fundamental finding of five decades of anthropological inquiry into the origins of contemporary African ethnicity: most African ethnic consciousness, as we know it today, is of colonial-era vintage, the product of cultural entrepreneurship in the crucible of social tensions in the modern sector and especially of urban economic and political competition (Helm 1968; Vail 1989). Horowitz's hard primordial factors—ancestry, customs, skin color, language, geographic home—serve as a foil in inventing new ethnic solidarity, while newly formed ethnic unity in turn lurches onto new formulations of ancestral origin, culture, and ethnic homeland.

Far from representing cut and dried, mutually hostile ethnolinguistic units, African ethnic categories (like all others worldwide) are by definition multilayered, mutating constantly with respect to time and place. Depending on leadership and its policy orientation, ethnic groups in Africa might be rallied to a development-friendly, a neutral, or an economically deleterious cause. With the potentially benign side of African ethnicity thus admitted into the development equation, and without brushing aside the malignant aspect, I would like to bring attention to the unexplored potential of merging the rich and varied findings on African ethnic behavior with interest group theory, with the insights obtained from decentralization and fiscal federalism, and with institutional reform theories that accommodate cultural diversity and intragroup competition in their own right. This is a more constructive policy alternative, but one that is habitually thrown out by some colonial-sounding contemporary interpretations that construe African ethnic diversity as a unidirectional menace to economic development—which need not always be the case.

Michael Chege is director of the Center for African Studies at the University of Florida, Gainesville.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

A Century of "Baleful" African Ethnicity and Its Malcontents

As Thomas Packenham recounts in his epic narrative of the European scramble for African colonies in the late 19th century, the rhetorical justification for African colonization was underpinned by a consensus that the relative backwardness of the "dark continent" was a result of endemic interethnic warfare and the slave trade—the antidote to which was to be found in philanthropist and explorer David Livingstone's three Cs: commerce, civilization, and Christianity (Packenham 1991). For years the containment of interethnic conflicts in Africa in the interest of peace, trade, and economic progress dominated the colonial-era literature on what we now call development.

In the wake of African nationalism and the run-up to decolonization in the 1960s, that governance formula was altered to provide incoming African governments with the intellectual justification for integrating the diverse ethnic groups within colonial-established territorial boundaries into a cohesive, functional nationhood consistent with the comprehensive development plans that were characteristic of that era. Ethnic identity, derisively referred to as "tribalism," acquired a pejorative connotation in contrast to the glamorized ideal of the new nation.

In his influential *Stages of Economic Growth*, Walt Rostow (1963, p. 7) listed as one of the "noneconomic preconditions for take-off to sustained growth" the need for "a new nationalism, in opposition to the traditional landed regional interests" whose divisive provincialism was inimical to economic growth and to the emergence of the all-important "strong modern national government." By that time resolution of the "crisis of national integration" had acquired a hallowed place among the six main problems of political development identified by the high-powered Comparative Politics Committee of the U.S. Social Science Research Council in its decade-long inquiry into the preconditions of modernizing political systems in the developing world (Binder and others 1971).

In the meantime a powerful intellectual backlash had already taken the offensive on two fronts: the colonial portrayal of African conflicts in precolonial societies and the notion of a mandatory repertoire of preconditions for economic take-off that each developing country would have to meet. A generation of Western and African historians produced evidence that, as explanatory variables of African underdevelopment, colonial-vintage "tribal" rivalries, geography, and sociocultural constraints were exaggerated, misunderstood, or fabricated (UNESCO 1985, 1989). From the Marxist-leaning dependency perspective of that era, Walter Rodney (1972) contended that far from being economically deleterious, many of the indigenous African social institutions destroyed by colonial rule had constituted the basis of much previous social and economic progress.

Defending a free enterprise approach to West African agriculture long before it became fashionable and espousing the polar opposite of Rodney's ideological view, Bauer and Yamey (1957) had earlier asserted that while some local customs were sometimes inimical to economic development, many ethnic-based institutions in the region had spurred the spectacular growth in the output of cocoa, kola nuts, coffee,

peanuts, cotton, and palm oil over the previous five decades. According to Bauer and Yamey, the main obstacles to agricultural development in West Africa were the monopsonistic networks of colonial marketing boards and state regulation of rural trade. At the nadir of the counteroffensive against the school of natural and socio-cultural obstacles to economic development, Hirschman (1965) demonstrated that where they could not be undermined by action on a different front, some of the perceived hindrances to progress—civic conflicts included—could turn out, like Adam Smith's "invisible hand," to have unintended benevolent consequences that should never be ignored by economic reformers. Through detailed case studies of social conflict and drought management in northeastern Brazil, land reform in Colombia, and inflation in Chile, Hirschman (1963, p. 6) sought to demonstrate that "a society can begin to move forward as it is, in spite of what it is, and because of what it is."

The burgeoning econometric literature on African ethnic conflicts and their relationship to negative per capita growth entirely ignores these findings. In a recent article, Easterly and Levine (1997) claim to have uncovered empirical evidence that African "ethnolinguistic fragmentation" is significantly negatively correlated with school attainment, financial depth, and the number of telephones per worker and positively associated with the black market currency premium. These factors in turn have negative consequences for overall economic progress, accounting for 25–40 percent of the growth differential between a regressing Sub-Saharan Africa and the then fast-growing economies of Southeast Asia. The way this works, we are told, is that over time African ethnic diversity, on its own momentum, generated a perverse one-dimensional cutthroat intergroup competition that made it exceedingly difficult to agree on the distribution of public goods and on sustainable, development-friendly macroeconomic policies.

No systematic case study evidence is produced to prove a precise cause-effect linkage between high ethnic diversity indexes and poor growth policy. Rather, the article is peppered with anecdotes and supportive (if controversial) opinions, including those of some African scholars, to drive home the connection between "Africa's growth tragedy" and cancerous ethnic rivalry that holds full-blown macroeconomic reform perpetually hostage to sectarian interests. According to Easterly and Levine, over the past 30 years ethnic fragmentation has wreaked havoc on African development in a way that appears alternately to be dependent on or independent of public policy. Throughout the period Africans, and their development collaborators, remained incapable of figuring a way out of this self-made hell of political instability, rent seeking, and inadequate provision of public goods.

If this model predicts anything, it must be as a result of multicollinearity and the spurious correlation that is asserted between African ethnolinguistic groups' perennially deleterious (as opposed to constructive) intergroup differences over macroeconomic policy, and their spillover into exchange rate mismanagement, lack of infrastructure, high public sector deficits, instability, and poor education. Ignoring the well-documented secular (as opposed to ethnic) policy origins of state intervention in Kwame Nkrumah's Ghana (Killick 1978) and Julius Nyerere's Tanzania (Lofchie 1989)—often with the support of mainstream economists and donors—

Easterly and Levine attribute economic debacle in both countries to the immanent logic of internal ethnic contestation, notwithstanding much country-specific evidence to the contrary (Young 1977, pp. 216–73; Schildkrout 1979).

It bears pointing out, following the Morrison, Mitchell, and Paden (1989) classification, that the mean distribution of ethnic groups (48) in the six African states credited by the World Bank (1994) with large improvements in macroeconomic reform between 1981–86 and 1987–91 is more than twice that in countries suffering from economic deterioration (21)—the exact opposite of what the model would suggest. Indeed, if, as Easterly and Levine assert, Botswana's supposed ethnic homogeneity is behind its record growth rates, then Somalia, which has an even more homogeneous ethnolinguistic and religious tradition, must pass the development test with flying colors—as must Burundi, Haiti, Rwanda, and Yemen.

Development studies swing like a pendulum from one extreme to the other and back. One of the most fortuitous consequences of Horowitz's article is that, much like the dissenting work against the mandatory "preconditions" for growth in the 1950s and 1960s, it could provide us with an authoritative excuse for questioning the ascendant paradigm attributing African economic regress to the continent's unusually high ethnic diversity, a theoretical approach with distinct echoes of the colonial era. Ethnic generalizations call for great circumspection, not only because ethnicity is such a slippery concept, but also because it is easy to stereotype and cause offense, even with the best of intentions. Aimed primarily at synthesizing the wide spectrum of academic opinions on the socio-structural origins of ethnic identity and ethnic-based political mobilization, Horowitz's article cautions against hard and fast generalizations much like those of the supposedly concrete ethnolinguistic classifications delineated above.

A considerable portion of the article is, in fact, taken up by a valiant attempt to reconcile the 10 contending schools—those inclined toward "primordial" origins of ethnic rivalry and those that maintain that ethnicity is a constantly renewable social construct, an "imagined community" (Anderson 1983) dependent on social context and the whims of cultural and political entrepreneurs, as is evident in the contemporary Balkans, California, the Democratic Republic of Congo, Kenya, South Africa, Sudan, and elsewhere. To wit, ethnic identity is a multilayered phenomenon that changes with time and place and is frequently mobilized in response to strategic positions assumed in situations of social conflict or its opposite—solidarity in expanding the domain of public and private goods. The empirical question in that context then becomes why at a given time and place a constellation of groups is likely to strike in a politically negative direction as opposed to a positive one.

Yet even with the reconciliation of the divergent positions on the structural and ideological bases of ethnicity, there is an acute need to juxtapose the primordial-constructivist spectrum against the broad span of concretely observed ethnic interaction (ranging from the negative to the positive) and to tease out a preliminary repertoire of policy and institutional norms that maximize development-inducing group behavior and penalize behaviors that are inimical to development. This perspective is now at the heart of the burgeoning literature on the optimal mix of indi-

vidual and group incentives and penalties within firms, institutions, and states. With the benefit of its insights, we can proceed to look at plausible policy responses to African ethnicity along the micro-macro behavior axis.

Synchronizing Ethnic Diversity with Institutional Reform in Africa

In *The Logic of Collective Action*, Mancur Olson (1965) observes that in the absence of “selective incentives”—that is, rewards for contributory membership and penalties for lack thereof—individuals will make only minimal personal sacrifice to achieve public goods because they can expect a free ride from their more civic-minded compatriots. But if free rider problems multiply as the domain of public concerns broadens (thus raising the probability of myriad other citizens playing an ameliorative part), personal contribution to civic groups or other collective goods increases the closer to home is the public problem. The smaller is the community, the lower is the probability of taking a free ride based on the sacrifice of a distant neighbor or a civic-minded far-off minority. To Olson this was the key reason why special-interest organizations, close-knit community groups, professional and trade associations, cartels, and oligopolies had a better track record of successfully lobbying for market-distorting and welfare-reducing policies than numerous but less organized citizens at large.

There are, of course, African communal groupings that are mobilized for less honorable goals than these—including the 1994 genocide in Rwanda, apartheid in South Africa, and corruption in Kenya. But those familiar with African urban ethnic associations after World War II can attest that they originated primarily as advocacy groups and community development organizations geared to uplifting the rural homeland (Hodgkin 1957). Much of the much-maligned ethnic advocacy that has surfaced under democracy in contemporary Africa is in fact no more than interest group mobilization that appeals to region, linguistic category, or culture in the context of highly centralized constitutional structures hostile to ethnic diversity and decentralization.

It might then be advisable to take seriously Olson’s remedy for welfare-minimizing special interests: first, developing a tight regime of “selective incentives” or, roughly speaking, the rule of law, and second, confining the provision of collective goods whenever possible to smaller and more transparent social and political units within which widespread individual concern for public issues is likely to be higher. That approach may well turn African ethnic solidarity into an asset, at least in part, rather than the intractable liability it is supposed to be. And it brings us straight into the issues of decentralization and fiscal federalism, much in vogue among economists, which have been almost totally divorced from the daily headlines on the push and pull of ethnic conflicts and regional loyalties from Albania to Zimbabwe.

In Federalist Paper 10—part of a series issued in 1787–88 to explain the U.S. Constitution—James Madison discussed factions at length, which he defined as “a number of citizens whether amounting to a majority or minority of the whole, who are united or actuated by some common impulse of passion or of interest, adverse to the rights of other citizens” (Rossiter 1961, p. 78). Far from being conventional in his approach to political factionalism, however, Madison argued that because factions

were inevitable in any genuinely democratic society, the best constitutional policy for handling them was to provide them with maximum legal room for political play. This approach stands in contrast to the perennial ethnic consensus endlessly urged on Africans by their development advisers—a solution Madison (rightly) considered potentially tyrannical. Competing factions might cancel out each other's ambition and leave the public good unharmed, if not enhanced, functioning in the same fashion as self-interest in Adam Smith's marketplace. Federalist Paper 10 provides two prescriptions: a territorially expansive state that would be too large to be preyed on by a few powerful factions and a federal state with multiple layers of government at the national, state, county, and neighborhood levels. The family resemblance between the proposed federalist constitution and disaggregated power on the one hand (bringing public responsibility closest to the community) and Olson's economic analysis on interest groups and public goods on the other ought to be immediately apparent.

That ethnic problems are a constant affliction of African national governance and often the scourge of economic reform should never be denied, even as we admit the shifting and multipurpose character of cultural identity described by Horowitz. What has been less obvious so far, however, is the extent to which centralizing programs by national governments and their external donors (many of which have decentralized governments at home) have helped to aggravate the problem, by ignoring the development potential of ethnic and other factions when given the right institutional home and policy treatment. Over the past decade Ethiopia, Nigeria, and South Africa have introduced varieties of federalism and decentralization to cope with ethnic demands. In South Africa constitutional devolution and power sharing were introduced as part of the historic 1994 "miracle" transition, a development about which the protagonists of perverse African ethnic fragmentation are silent. By mid-1998 the clamor for federalism and decentralization could be heard in Kenya, Mozambique, Senegal, Sudan, Tanzania, Uganda, not to mention Britain, South Asia, and Eastern Europe.

At the same time, issues of fiscal decentralization in the context of macroeconomic stabilization and economic liberalization have already edged their way to the forefront of economic policy research (Tanzi 1996; Hommes 1996). It is time to link the emerging initiatives of ethnic-driven constitutional decentralization to the growing concern of fiscal federalism in the context of market economics. This, and not the supposed high incidence of "ethnic fragmentation," is the challenge of the future. Using the accumulated sociological knowledge on ethnicity, it is also time to put to rest once and for all the colonial-era myths of African ethnic perversity, which add more rancor than value to an extremely complex issue and delay the solution of an urgent development policy problem.

References

- Anderson, Benedict. 1983. *Imagined Communities*. London: Verso.
- Bauer, Peter T., and Basil S. Yamey. 1957. *The Economies of Underdeveloped Countries*. Chicago, Ill.: University of Chicago Press.

- Binder, Leonard, James S. Coleman, Joseph LaPalombaba, Lucian W. Pye, Siden Verba, and Myron Weiner. 1971. *Crises and Sequences in Political Development*. Princeton, N.J.: Princeton University Press.
- Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy Policies and Ethnic Division." *Quarterly Journal of Economics* 112 (November): 1,203–50.
- Helm, June, ed. 1968. *Essays on the Problem of Tribe*. Seattle. University of Washington Press.
- Hirschman, Albert O. 1963. *Journeys toward Progress*. New York. Twentieth Century Fund.
- . 1965. "Obstacles to Development: A Classification and a Quasi-Vanishing Act." *Economic Development and Cultural Change* 13 (4): 385–93.
- Hodgkin, Thomas. 1957. *Nationalism in Colonial Africa*. New York: New York University Press.
- Hommel, Rudolf. 1996. "Conflicts and Dilemmas of Decentralization." In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1995*. Washington D.C. World Bank.
- Killick, Tony. 1978. *Development Economics in Action: A Study of Economic Policy in Ghana*. London: Heinemann Educational Books.
- Lofchie, Michael. 1989. *The Policy Factor: Agricultural Policy in Kenya and Tanzania*. Denver, Colo.: Lynne Rienner Publishers.
- Morrison, George, Robert C. Mitchell, and John C. Paden. 1989. *Black Africa: A Comparative Handbook*. New York: Paragon House.
- Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, Mass.: Harvard University Press.
- Packenham, Thomas. 1991. *The Scramble for Africa*. New York: Random House.
- Rodney, Walter. 1972. *How Europe Underdeveloped Africa*. Washington, D.C.: Howard University Press.
- Rossiter, Clinton, ed. 1961. *Federalist Papers*. New York: Penguin Books.
- Rostow, Walt W. 1960. *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge, U.K.: Cambridge University Press.
- Schildkrout, Emd. 1979. "The Ideology of Regionalism in Ghana." In William Schack and Elliot Skinner, eds., *Strangers in African Societies*. Berkeley: University of California Press.
- Tanzi, Vito. 1996. "Fiscal Federalism and Decentralization: A Review of Some Efficiency and Macroeconomic Aspects." In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1995*. Washington D.C.: World Bank.
- UNESCO (United Nations Educational, Social, and Cultural Organization). 1985. *General History of Africa*. vol. 7. Berkeley. University of California Press.
- . 1989. *General History of Africa*. vol. 6. Berkeley: University of California Press.
- Vail, Leroy, ed. 1989. *The Creation of Tribalism in Southern Africa*. Berkeley: University of California Press.
- World Bank. 1994. *Adjustment in Africa: Reforms, Results and the Road Ahead*. A Policy Research Report. New York: Oxford University Press.
- Young, Crawford. 1977. *The Politics of Cultural Pluralism*. Madison. University of Wisconsin Press.

Floor Discussion of "Structure and Strategy in Ethnic Conflict: A Few Steps toward Synthesis," by Donald L. Horowitz

A participant from the World Bank wondered what role nonpolitical institutions and mechanisms can play in reducing ethnic and other conflicts. Could conflict be mitigated with dynamic consumer markets, secure property rights, and easily enforced contracts?

Deadly ethnic violence has almost disappeared in Western countries, noted Donald L. Horowitz (presenter). In the postwar era there have been few cases where members of one ethnic group have killed members of other groups in large numbers. But while markets, property rights, and related institutions might limit conflict, ideological and other factors are also at work. And in developing countries there is no discernible correlation between conflict and such variables. Still, countered Michael Chege (discussant), leadership and the rule of law are important elements in reducing ethnic conflict.

Robert H. Bates (discussant) added that various nonpolitical institutions—pension funds, alternative forms of savings, property rights—play a major role in shaping the incentives to form ethnic groups. For example, some countries have perpetuated ethnic groups by maintaining communal property in rural areas. That way, urban workers can return to rural areas to retire, and their membership in an ethnic group will give them a claim on land. This approach also saves countries from having to build up pension systems in urban areas.

A participant from ICF Kaiser International believed that most ethnic conflicts result from ignorance—ignorance of how other groups live, or of how to appreciate and respect differences. If ignorance is exploited to create conflict, how can this ignorance be removed?

Horowitz disagreed with the participant's view, citing studies in the United States on the so-called contact hypothesis. Researchers had believed that if members of different groups met and got to know one another, intergroup relations would improve. But years of experience with many types of experiments had discredited that notion. The reason is that while relationships may improve at the individual level, those

This session was chaired by Joanne Salop, director, Operations Policy and Strategy at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

improvements will not necessarily extend to the group level. In fact, groups in conflict are often in close contact and know one another quite well—and conflict arises precisely because of the things they know about each other and the characteristics they impute to each other.

Bates added that some studies had seen conflict increase when groups learned more about each other. Chege agreed that ethnic conflict does not arise from ignorance—and that it sometimes comes from knowledge. Ethnic animosity and ethnic solidarity are powerful forces, and they help determine the overall balance of economic resources and distribution of power.

Joanne Salop (chair) closed by emphasizing the importance of ethnic issues and their effects on economic policies for the work of the World Bank. Ethnic concerns have implications for projects and for country assistance strategies, so it is vital to understand their roots. The Bank's operations and discussions with member countries would benefit greatly from more research on these topics—research that bridges all the social sciences, not just economic and legal studies.

The Political Economy of Ethnicity

Paul Collier

Ethnic diversity is associated with lower growth only in societies where political rights are limited. A similar relationship holds for World Bank projects, which are less successful in ethnically diverse societies with poorly developed democratic institutions. Yet highly fractionalized societies are less likely than homogeneous societies to experience violent conflict; only moderate fractionalization is associated with an increased risk of civil war. Moreover, higher incomes and democratic institutions reduce the risk of violent conflict. Once conflict escalates to full-scale civil war, the persistence of conflict and the sustainability of a settlement depend more on ethnic composition and less on income and political rights. Thus the maintenance or creation of extreme ethnic fractionalization may be more effective than economic development in ensuring that peace is maintained following violent conflict. Highly fractionalized societies must be democratic, however, if they are not to suffer high economic costs.

For many years social scientists ignored the ethnic diversity in developing countries. More recently, they have concluded that diversity is detrimental to economic performance. Easterly and Levine (1997) show that at the aggregate level ethnic diversity significantly reduces economic growth and suggest that diversity is the main explanation for slow growth in Africa, the most ethnically diverse continent.

A related perception is that ethnic diversity increases the risk of civil strife (World Bank 1997). This perception stems from vivid instances of interethnic violence around the world and from the aggregate correlation between diversity and civil war in Africa. If there is a relationship between diversity and conflict, it might account for the detrimental economic effects of diversity.

But even if it can be proven that ethnic diversity is economically dysfunctional, is this a helpful line of research? After all, there is nothing a country can legitimately do

Paul Collier is director of the Development Research Group at the World Bank. The author is grateful to Anke Hoeffler for considerable research assistance, Pablo Zoido-Lobaton for the regressions in table 3, and Mans Soderbom for the hazard functions in table 5.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

about its ethnic composition—and illegitimate acts, such as ethnic cleansing, are hardly to be encouraged. This article shows, however, that such research offers a hopeful and practical message. The economic effects of diversity are by no means as detrimental as is commonly thought, and diversity actually reduces the risk of violent conflict. Moreover, the detrimental effects of diversity have feasible political solutions.

The article examines the relationships among fractionalization, growth, and civil war from a highly aggregated, cross-country perspective. While no substitute for country-level studies, which bring to light factors undetectable in cross-country analysis, the approach is nevertheless one in which economics has a comparative advantage, and it can serve as a complement to country-level work.

Several important results emerge from the analysis. First, ethnic fractionalization can hold down income by reducing trust among citizens, thereby increasing transactions costs. Second, political institutions matter much more in ethnically fractionalized societies than in homogeneous societies. In fact, democracy can almost completely offset the economic damage associated with high fractionalization. Third, the relationship between ethnic fractionalization and the risk of violent conflict is more subtle than has been thought. Only moderate fractionalization is associated with an increased risk of civil violence; highly fractionalized societies are less likely than homogeneous societies to experience civil war. Indeed, the high diversity in Africa reduces rather than exacerbates the risk of civil conflict there. Moreover, democratic institutions can substantially reduce the risk of violence. Because income is also an important determinant of the risk of conflict, democracy reduces risks both directly, by helping to defuse conflict, and indirectly, by increasing the opportunity cost of rebellion. That democracy effectively eliminates the potentially negative effects of ethnic diversity on growth while high ethnic diversity reduces the risk of violent conflict is encouraging for highly diverse countries.

Finally, examination of the determinants of the escalation, maintenance, and resumption of violence in societies that have experienced civil war shows that ethnic fractionalization is very important and income levels are unimportant in all three processes. This result suggests that the creation or maintenance of extreme ethnic fractionalization may be more effective than economic development at ensuring that peace is maintained following violent conflict. Highly fractionalized societies must be democratic, however, if they are to avoid high economic costs.

Ethnic Diversity, Political Institutions, and Economic Performance

Easterly and Levine (1997) find that ethnic diversity is correlated with slow economic growth, a correlation they attribute partly to the fact that diversity is associated with poor economic policies. From this the authors infer that diversity may make it harder for policymakers to reach cooperative solutions to problems instead of engaging in zero-sum games. Testing this hypothesis, however, they find a residual negative effect of ethnic diversity over and above its effect through policy.

Alesina, Baqir, and Easterly (1997) analyze the effects of ethno-linguistic diversity at the local level using data on U.S. cities. They find that ethnic diversity reduces

the effectiveness of city government in delivering a range of public services. Diversity, they argue, lowers spending on productive public services while increasing rent seeking. Collier and Garg (forthcoming) present similar evidence using disaggregated data on public sector labor markets in Ghana. They find that unless restrained, kin groups function as patronage systems. In the public sector the locally dominant ethnic group exacts a 25 percent wage premium for its members. No such premium prevails in the private sector, presumably because the private sector has a stronger incentive to restrain kin group patronage. These studies suggest that in addition to leading to inferior macroeconomic policies, ethnic diversity reduces the efficiency of public service delivery.

A third route by which diversity undermines economic performance is by inhibiting the development of social capital and trust, thereby raising transactions costs. Ethnic fractionalization helps build trust within ethnic groups (Putnam 1993), but its effect on society as a whole has not, to my knowledge, been investigated. To address that issue, I use internationally comparable data on the level of trust in 23 countries, as measured through surveys asking respondents whether other people in the society can be trusted. Putnam makes trust endogenous to certain forms of social interaction. But because social interaction must be endogenous to its costs, I take the more conventional economic route of explaining trust in terms of the costs agents face in deciding whether to interact socially. Ethnic divisions are a potential additional barrier to social interaction and hence detract from trust. As proxies for the costs of social interaction, I use the density of the telephone system and the spatial density of the population (table 1).

Trust, measured on a scale of 0–100, turns out to be explicable in terms of the costs of social interaction, with telephone density being highly important.¹ Population density enters as a quadratic function: people trust their neighbors both when they have few of them and when population density is very high. The effect of ethnic fractionalization is hard to interpret. Although it is not statistically significant at an acceptable level, it is not so insignificant that it can be dismissed as hav-

Table 1. Ethnic Diversity and Trust

<i>Variable</i>	<i>Regression result</i>
Ethnic fractionalization	-0.14 (-1.3)
Telephone density (log)	16.35 (6.2)
Population density	-0.17 (-3.0)
Population density ²	0.00043 (2.9)
Adjusted R ²	0.62
F	10.11
Number of observations	23

Note. Numbers in parenthesis are t-statistics

Source: Author's calculations, see text

ing no effect on trust. The coefficient is negative and fairly large: the difference between the least and most ethnically diverse societies maps into a one standard deviation decrease in trust. Thus this evidence weakly suggests that ethnic diversity is a barrier to trust. Ethnic fractionalization, then, appears to be negatively correlated with economic growth, good macroeconomic policy, city government performance, efficient public sector wage determination, and the level of trust.

The Effect of Political Rights

How do political institutions affect these outcomes? Democracy has the potential both to discipline governments into delivering reasonable economic policies and to provide a framework in which groups can negotiate mutually beneficial outcomes. The baseline regression explains average per capita GDP growth between 1960 and 1990 using a standard set of nonpolicy structural characteristics—initial per capita GDP, the population growth rate, and whether the country is landlocked (table 2). (All policy variables are excluded because these must be presumed to be endogenous to ethnic diversity.) To this I add ethno-linguistic fractionalization, scored on a scale of 0 to 100, and the extent of democratic political rights, scored on a scale of 1 (fully democratic) to 7 (complete absence of political rights).²

All the variables are significant: democracy leads to higher growth, fractionalization to lower growth. Ethnic diversity appears to be very detrimental: the most diverse society grows about 1.6 percentage points less than a homogeneous society and ends up with a steady-state income about one-fifth that of the homogeneous society.

Table 2. Two Regressions of Growth on Ethnic Fractionalization and Democracy

<i>Variable</i>	<i>Baseline regression</i>	<i>Interaction effect</i>
Constant	10.73 (3.42)	9.20 (3.55)
GDP (log)	-0.90 (-2.55)	-0.81 (-2.59)
Population (log)	-0.73 (-2.05)	-0.86 (-2.54)
Landlocked	-1.01 (-1.85)	-0.99 (-1.84)
Ethno-linguistic fractionalization	-0.0156 (-2.22)	
Political rights	-0.26 (-1.73)	
Ethno-linguistic fractionalization-political rights interaction		-0.0043 (-3.22)
Adjusted R ²	0.16	0.18
F	4.63	6.20
Number of observations	94	94

Note: Numbers in parenthesis are t-statistics
Source: Author's calculations, see text

Interaction of Ethnic Diversity and Political Rights

How does ethnic fractionalization interact with democracy? At the risk of oversimplification I distinguish between performance politics and identity politics. In performance politics the political system helps improve economic policies. Cooperation and the development of social capital are facilitated, and governments that deliver poor economic performance are punished (and those that deliver good performance, rewarded) by voters. Identity politics, by contrast, emphasizes differences among citizens, limiting cooperation and the development of social capital. Citizens in such a system remain loyal to their party regardless of economic performance, and governments deliver patronage to loyalists rather than services to the median voter.

An implication of the result of Alesina, Baqir, and Easterly (1997) is that the interaction between ethnic fractionalization and democracy is likely to be negative. In U.S. cities, for example, ethnic diversity appears to be associated with a shift from performance to identity politics. The interaction could well be positive, however. Democracy might provide institutions in which potentially costly disputes between ethnic groups can be mediated. Cooperation might be sufficiently easy in homogeneous societies that it does not require democratic institutions, whereas in ethnically diverse societies these institutions may mean the difference between zero-sum and cooperative solutions. Thus an ethnically diverse society might gain more from democracy than a homogeneous society because the homogeneous society has less need for dispute resolution.

This result need not conflict with that of Alesina, Baqir, and Easterly (1997). The ethnic diversity of political decisionmaking is endogenous to political boundaries. Decentralization, however, offers the possibility of shifting certain economic decisions into a different ethnic space. The United States has performance politics at the federal level ("It's the economy, stupid") but identity politics at the local level. But this situation is unusual and probably reflects the immigrant nature of the society. In most countries that have identity politics it dominates at the national level because there is little ethnic diversity at the local level. But because national boundaries are much larger than those of ethnic groups, local politics is more likely to be performance politics. These conditions help explain the regional decentralization in Ethiopia: regional democracy is fairly homogeneous, despite diversity at the national level.

To test how ethnic diversity and democracy interact, I included an ethnolinguistic fractionalization-political rights interaction term, in the regression (see table 2). Once this term is included, the effects of ethnic diversity and political rights become completely insignificant (not reported), indicating that their effect works entirely through their interaction.³ Once these insignificant direct effects are eliminated from the regression, the interaction is the most significant variable in the regression, with a large and negative coefficient. Thus democracy significantly reduces the problem posed by ethnic diversity. Specifically, in a completely undemocratic political system a homogeneous society will grow by 3 percentage points

more than a maximally fractionalized society. By contrast, in a fully democratic political system the growth differential narrows to a modest 0.4 percentage points. This finding suggests that a lack of political rights is economically ruinous in highly fractionalized societies.

Ethnic Diversity, Political Institutions, and Performance of World Bank Projects

To investigate the same relationships at the project level, I examined the evaluations of the Operations Evaluation Department (OED), an independent unit within the World Bank, of Bank projects implemented between 1958 and 1996 in 87 countries. (For a more detailed discussion of the OED dataset, see Kaufmann and Wang 1995.) The dependent variable in the dataset is the proportion of projects in a country classified by the OED as unsatisfactory. The explanatory variables were initially the same as for the growth regressions. A landlocked location and the population growth rate were not even remotely significant, however, and were dropped. Thus the baseline regression includes average per capita GDP and the direct effects of ethnic diversity and political rights (table 3).

Ethnic diversity was most significant when it was entered as the square of the ethno-linguistic fractionalization index. In this baseline the direct effects of ethnic diversity and political rights are qualitatively similar to those in the baseline growth regression: diversity and a lack of rights increase the risk of project failure, although the lack of rights is not statistically significant. The second regression in table 3 explores the interaction effect between diversity and political rights. The square of the ethno-linguistic fractionalization index is used as a proxy for diversity. When the interaction term is introduced as an additional variable alongside the direct effects,

Table 3. Two Regressions of Project Performance on Ethnic Fractionalization and Democracy

<i>Variable</i>	<i>Baseline regression</i>	<i>Interaction effect</i>
Constant	0.43 (1.66)	0.30 (1.47)
GDP (log)	0.052 (1.72)	0.063 (2.39)
Ethno-linguistic fractionalization ²	-0.000015 (-1.94)	
Political rights	-0.013 (-1.04)	
Ethno-linguistic fractionalization ² -political rights interaction		-0.000003 (-2.22)
Adjusted R ²	0.19	0.20
F	7.95	11.82
Number of observations	87	87

Note. Numbers in parenthesis are t-statistics

Source. Author's calculations, see text

both direct effects become insignificant (not reported). When the direct effects are excluded, the regression with the interaction effect dominates the baseline regression.

Thus, as in the growth regression, ethnic diversity and political rights affect the success of a project only through their interaction. An ethnically diverse society does not have a higher project failure rate as long as it is fully democratic. But an ethnically diverse society that is undemocratic has a dramatically higher rate of project failure—44 percent for an undemocratic diverse society, compared with 26 percent for a democratic diverse society. These results hold when the same regressions are run controlling for nine time periods and 16 project types. In fact, the interaction effect becomes more highly significant (and indeed, becomes the most significant variable in the regression). Thus the lack of political rights that is ruinous for growth in ethnically diverse societies is also ruinous for development projects.

Ethnic Diversity, Political Institutions, and the Incidence of Violent Conflict

To examine the effect of ethnic diversity on violent conflict, I modeled the risk of civil war as a cost-benefit calculation. (This section draws on Collier and Hoeffler 1998a, b and Collier, Hoeffler, and Söderbom 1998). The costs of rebellion are the coordination cost and the opportunity cost. Recall that in the context of public service provision Alesina, Baqir, and Easterly (1997) find that ethnic fractionalization increases the difficulty of cooperation. If cooperation is a necessary input into the initiation and maintenance of rebellion, ethnic diversity among potential rebels may reduce the risk of violent conflict.

This conclusion implies a nonmonotonic relationship between the ethnic diversity of a society and the risk of conflict. A society with 2 ethnic groups would have lower coordination costs for rebellion than a homogeneous society because the rebels could identify themselves as ethnically distinct from government supporters. But a society with 20 ethnic groups would have higher coordination costs for rebellion than a society with 2 ethnic groups (and potentially higher costs than a homogeneous society) because potential rebels would need to cooperate across ethnic boundaries.

The other postulated cost of rebellion is the opportunity cost, for which per capita GDP is a reasonable proxy. At very low income levels, people—especially the teenage males who fight in most rebellions—have little to lose from joining rebel armies.

The benefits of rebellion are conditional on military success, which in turn depends on the financial capacity of government to purchase defense. Actual defense expenditure is endogenous to the risk of war and so cannot be used to predict it. Instead I use the share of primary exports in GDP as a proxy for the taxable base of the economy, because in most societies primary exports are highly taxed. Conditional on success, the benefits of rebellion depend on its objectives. Some rebel groups aspire to secession. A simple proxy for the desire for secession is the size of the population: governments in countries with larger populations are less likely to satisfy the needs of their peripheral citizens (see Alesina and Spolaore 1997). Other rebel groups aspire to capture the state. In this case a proxy for the

gains from rebel victory is the taxable base being captured. The taxable base thus has an ambiguous net effect on the risk of rebellion, both reducing the chance of rebel victory and increasing the gain should victory be achieved. Because the net effect need not be monotonic in the taxable base, primary exports are entered as a quadratic in the test of the model.

The probability of civil war increases with the net present value of the utility the rebels expect to gain by war, $E(U_w)$. Formally,

$$E(U_w) = \int_0^{\infty} \int_D^{\infty} \left[\frac{p(T)g(T,P)}{(1+r)^t} \right] dt - \int_0^D \left[\frac{f(Y,C)}{(1+r)^t} \right] dt \Big] h(D) dD$$

where p is the probability of rebel victory, T is the share of primary exports in GDP, P is population (log), Y is per capita GDP, C is the level of ethno-linguistic fractionalization, D is the expected duration of warfare, and r is the discount rate.

The model was tested on a comprehensive set of civil wars since 1960 using country characteristics in 1960 or as close to 1960 as the data permit. The probit attempts to predict from among the 97 states for which data are available the 27 states where civil wars occurred after 1960. The year 1960 is used as a starting date because many countries gained independence around that time; before independence the underlying risks of civil war were largely suppressed by colonial powers. By looking at conflicts since 1960, we avoid the potential endogeneity of income to civil war risk.

In the baseline regression all the variables are significant with the expected signs (table 4). The risk of civil war is strongly related to per capita GDP: poverty sharply

Table 4. Two Probit Regressions of the Occurrence of Civil War

Variable	Baseline regression	Political rights added
Constant	-1.543 (2.14)	-2.61 (-2.66)
GDP	-0.001 (2.70)	-0.0004 (-1.87)
Primary exports/GDP	16.16 (2.56)	13.89 (2.18)
Primary exports/GDP ²	-29.47 (2.28)	-27.71 (-2.05)
Ethno-linguistic fractionalization	0.033 (1.35)	0.027 (1.09)
Ethno-linguistic fractionalization ²	-0.0004 (1.60)	-0.0003 (-1.26)
Population	0.0003 (2.39)	0.0002 (2.26)
Political rights		0.225 (1.80)

Note: Numbers in parenthesis are t-statistics

Source: Baseline regression from Collier and Hoeffler 1998a

increases the risk of war. The effect of ethnic diversity is nonmonotonic: moderately diverse societies are at greater risk of civil war than homogeneous societies, a result consistent with the fact that the coordination costs of rebellion fall if potential rebels are ethnically differentiated from government supporters. But highly ethnically diverse societies are at even less risk of civil war than homogeneous societies, a result consistent with the fact that diversity among potential rebels increases the coordination costs of rebellion.

These results suggest that the high incidence of civil war in Africa is explained entirely by the continent's low income; high ethnic diversity has actually reduced the risk of violent conflict (see Collier and Hoeffler 1998a). The risk of civil war rises with the level of primary exports unless such exports are abundant. My interpretation of this result is that the increased lure of capturing the state is eventually offset by the enhanced capacity of the government to defend itself and buy off opponents.

The second probit in table 4 adds political rights, measured as before. It shows that per capita GDP and democratic rights are correlated, with the introduction of rights reducing the effect of income (which nevertheless remains significant). The effect of ethnic diversity loses significance, but the coefficients are little affected. Political rights substantially reduce the risk of civil war; moving from the lowest to the highest level of rights has an effect equivalent to raising income from the mean to 2.4 times the mean. Thus democratization is worth about half a century of income growth in terms of reducing the risk of civil war.

Although ethnically diverse societies are not directly prone to violent conflict, in the absence of democracy they are prone to poverty, which makes them vulnerable to conflict. Recall that a maximally fractionalized society that lacks democratic institutions grows 3 percentage points less a year than an undemocratic homogeneous society. Given that the mean per capita growth rate during 1960–90 was only 1.7 percentage points, this combination effectively condemns such societies to radically low incomes.

To see the effect of fractionalization combined with democracy, consider two societies—one that has the maximum level of fractionalization and one in which the level of fractionalization maximizes the risk of war. Complete fractionalization reduces the risk of civil war by as much as a 62 percent increase in income above the mean. But if the completely fractionalized society is also completely undemocratic and the partly fractionalized society is fully democratic, the partly fractionalized society will grow by 1.4 percentage points a year more than the completely fractionalized society. If the two societies start at the same income level, after 34 years the higher income level in the partly fractionalized society will fully offset its higher risk of civil war.

Thus democracy reduces the risk of civil war both directly, by enhancing the capacity to mediate disputes, and indirectly, by increasing income. In the most fractionalized societies the main contribution of democracy to peace is through its effect on income. This is partly because democracy's contribution to income growth is much larger in highly fractionalized societies and partly because, for a given income level, the risk of war is lower in such societies. In partly fractionalized societies the

risk of civil war is at its peak, and the contribution of democracy to income growth is reduced. However, democracy also directly reduces the risk of conflict.⁴

Ethnic Diversity, Political Institutions, and the Escalation, Maintenance, and Revival of Violent Conflict

Finally, I turn to the problems posed by the escalation of rebellion, ongoing civil wars, and the preservation of postwar peace. All three problems are subject to common processes that differ from those affecting the risk of civil war in initially peaceful societies.

For the escalation of rebellion I use data from Gurr (1993), who distinguishes six levels of rebellion, ranging from political banditry to civil war. I use an ordered probit model to explain the passage from peace through these six stages. The core innovation in Gurr's dataset is its focus on disadvantaged ethnic minorities. To keep the results comparable with those of the other regressions, however, I ignore this aspect of Gurr's data and instead retain the 0–100 measure of ethno-linguistic fractionalization. Gurr's dataset covers fewer countries than the previous samples but is nevertheless substantial.

To examine the problems of ongoing civil wars and postwar peace, I estimate hazard functions. These are variants on the functions estimated in Collier, Hoeffler, and Söderbom (1998), which provides a fuller discussion of the methods and data used. *The hazard functions model the duration of conflicts and postwar peace. The civil war data are drawn predominantly from Small and Singer (1982, 1994) but have been updated to 1997. The hazard functions are estimated using semiparametric methods rather than the more restrictive Weibull specification. The hazard of peace during a civil war is estimated using the first and (where pertinent) second episode of war for each country in the dataset, beginning in 1960. The hazard of renewed war following peace can be investigated only for countries that have at some stage settled a civil war, which restricts the sample. To increase the sample size, this hazard is estimated for all countries that experienced a civil war since 1945. Since the focus here is not on the structure of the hazard but rather on the significance of the explanatory variables, only the explanatory variables are reported (table 5).*

The three results display a common pattern. First, ethnic diversity is always important and is characterized by a quadratic relationship. The results are consistently more significant than in the probits of the causes of civil war reported in table 4, despite the much smaller sample sizes. Second, per capita GDP, primary exports, and population are no longer remotely significant. Why should the relative importance of per capita GDP and ethnic diversity change so substantially?

Recall that per capita GDP is a proxy for the opportunity cost of joining a rebellion and ethnic diversity is a proxy for the cost of coordination. The decisive factor in determining whether a civil war occurs is whether would-be rebel leaders can persuade others to join them. Potential recruits, who are initially free to choose whether to rejoin a rebellion, weigh the opportunity cost of doing so. But once a rebellion begins, its ability to sustain itself depends on its capacity to sustain rebel cohesion.

Table 5. Escalation, Maintenance, and Revival of Violent Conflict

<i>Variable</i>	<i>Escalation</i>	<i>Maintenance</i>	<i>Revival</i>
GDP	-0.00008 (-1.47)	-0.00017 (-0.92)	-0.00004 (-0.22)
Primary exports/GDP	-5.48 (-0.63)	^a	^a
Primary exports/GDP ²	5.25 (0.20)	^a	^a
Ethno-linguistic fractionalization	0.053 (1.92)	0.0512 (1.71)	-0.05 (-1.59)
Ethno-linguistic fractionalization ²	0.00052 (-1.85)	-0.0006 (1.74)	0.006 (1.72)
Population (log)	0.18 (1.04)	0.0693 (0.32)	-0.2385 (-1.21)
Political rights	0.30 (2.33)	-0.839 (0.72)	-0.1517 (-1.19)
Number of observations	47	45	56

Note. Numbers in parenthesis are t-statistics

a Primary exports/GDP were completely insignificant in the hazard functions and were dropped to ease estimation
 Source Author's calculations using data from Gurr 1993 for escalation, based on Collier, Hoeffler, and Soderbom 1998 for maintenance and revival

Ordinary rebel soldiers are not free to quit. Indeed, all armies classify quitting as desertion and punish it severely. This diminishes the importance of per capita GDP because the calculus of individual choice is less important. But rebel groups easily splinter and fight among themselves, a phenomenon governments exploit to suppress rebellion. I suspect that the quadratic relationship in ethnic fractionalization may reflect the difficulty of maintaining cohesion among rebel officers.

The preservation of peace in postwar societies appears to be determined by the same process (but with opposite signs) that determines whether wars are sustained. Rebel organizations do not voluntarily disband during the early stages of postwar peace, and the inability of rebels to sustain their cohesion helps preserve peace—just as it helps avoid the escalation of conflict and the perpetuation of civil war.

The most significant variable in the escalation probit is political rights. Democracy substantially reduces the risk that terrorist groups will be able to build up their support base sufficiently to escalate violence to full-scale civil war. But political rights appear to be less effective in reducing the duration of a conflict once it has escalated to full-scale war and in reducing the risk of renewed conflict following peace.

Thus the problem of avoiding civil war is to some extent distinct from the problems of preventing the escalation of civil wars, ending current wars, and sustaining peace following wars. In all four processes, however, homogeneous societies are safer than moderately diverse societies, and highly diverse societies are safer than homogeneous societies. Political rights are very important in avoiding and preventing the escalation of a conflict but appear to be less effective in achieving and sustaining settlements once civil wars are under way. Income is very important in

preventing civil war but less effective once violence begins. There is little evidence that income shortens the duration of war or sustains postwar peace.

Notes

1. Telephone density might be a proxy for other factors, notably income. The telephone system, however, is much more significant and powerful than other forms of infrastructure that are equally correlated with income. Because income is endogenous to trust it is not appropriate to control for it directly.

2. As a proxy for political rights, I use the Gastil Index, a subjective judgment of a number of underlying attributes, including the meaningfulness of elections, fairness of electoral laws and campaigning opportunities, voting power of the electorate, existence of political competition, evidence of political power shifting through elections, significant opposition voting, freedom from external or military control of domestic politics, minority self-determination or pluralism, decentralization of political power, and attempt of political agents to reach a consensus on national issues. Although subjective, the Gastil Index is closely correlated with other indexes, such as those of Bollen and Humana (see the discussion in Fedderke and Klitgaard 1998).

3. Even when ethnic diversity and political rights are included separately along with the interaction effect, their t-statistics are only 0.4 and 0.6, while the interaction effect is highly significant.

4. The transition to democracy may involve a temporary phase of increased risk (Gurr 1993). I do not revisit this problem here.

References

- Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112 (November): 1027–56.
- Alesina, Alberto, Reza Baqir, and William Easterly. 1997. "Public Goods and Ethnic Divisions." World Bank, Washington, D.C.
- Collier, Paul, and Ashish Garg. Forthcoming. "On Kin Groups and Wages in Africa." *Oxford Bulletin of Economics and Statistics*.
- Collier, Paul, and Anke Hoeffler. 1998a. "The Coming Anarchy? The Global and Regional Incidence of Civil War." Oxford University, Centre for the Study of African Economies.
- . 1998b. "On Economic Causes of Civil War." *Oxford Economic Papers* 50 (4): 563–73.
- Collier, Paul, Anke Hoeffler, and Mans Söderbom. 1998. "On the Duration of Civil War and Postwar Peace." Oxford University, Centre for the Study of African Economies.
- Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy." *Quarterly Journal of Economics* 112 (4): 1203–50.
- Fedderke, Johannes, and Robert Klitgaard. 1998. "Economic Growth and Social Indicators: An Exploratory Analysis." *Economic Development and Cultural Change* 46 (April): 455–89.
- Gurr, Ted Robert. 1993. "Why Minorities Rebel: A Global Analysis of Communal Mobilisation and Conflict since 1945." *International Political Science Review* 14: 161–201.
- Kaufmann, Daniel, and Yan Wang. 1995. "Macroeconomic Policies and Project Performance in the Social Sectors: A Model of Human Capital Production and Evidence from LDCs" *World Development* 23 (May): 751–65.

- Putnam, Robert. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, N.J.: Princeton University Press.
- Small, Melvin, and Joel David Singer. 1982. *Resort to Arms: International and Civil War, 1816–1980*. Beverly Hills, Calif. Sage.
- . 1994. *Correlates of War Project: International and Civil War Data, 1816–92*. Ann Arbor, Mich.: Inter-University Consortium for Political and Social Research.
- World Bank. 1997. “Social Capital: The Missing Link?” In *Expanding the Measure of Wealth*. Washington, D.C.

Comment on “The Political Economy of Ethnicity,” by Paul Collier

Pauline E. Peters

As one of the few anthropologists at this conference, I would like to borrow a quotation from Clifford Geertz that Donald L. Horowitz cites in his article: progress in anthropology is “marked less by a perfection of consensus than by a refinement of debate” (Geertz 1973, p. 29, cited in Horowitz, this volume). In this comment I offer suggestions on where refinement is needed in concepts of and research on ethnicity. I believe that such refinement is particularly important when an influential organization like the World Bank decides to include ethnicity in its research agenda. Without careful refinement, there is a danger not merely of compromising the quality of research but also of intensifying the ethnicization of political struggles in Africa. This is because ethnicity is not a primordially given essence, but the outcome of complex sociocultural and political processes of labeling and identifying people.

Conceptual Problems

The main problem I have with Paul Collier’s article is that the central concept of ethno-linguistic diversity is far too abstract and underspecified in social terms to have explanatory value in real-world African politics. The article conflates the terms ethno-linguistic diversity, ethnic diversity, ethnicity, ethnic fractionalization, ethnic identity, and ethnic groups, all of which refer to quite different social phenomena. Ethno-linguistic diversity may not indicate the presence of multiple ethnic groups, and multiple ethnic groups may not constitute social fractionalization. Moreover, the data on ethno-linguistic diversity—most of which were compiled by social scientists, including anthropologists, of earlier generations—overstate fractionalization in Africa because many of the distinctions are not socially and politically significant. That is, many groups considered distinct because of various ethno-linguistic markers interact almost seamlessly in many parts of the continent. Finally, the lack of clarity

Pauline E. Peters is development adviser at the Harvard Institute for International Development and lecturer on anthropology at Harvard University

Annual World Bank Conference on Development Economics 1998
©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

on what ethnicity is and how it works in political-economic and social ways fails to help us understand the political economy of ethnicity.

I recognize and appreciate Collier's attempt to counter what may be a dangerously negative view of Africa as doomed by its ethnicity. Chabal (1996, p. 33) criticizes the common view that "Africans are peculiarly ethnic in their sense of identity" and so are peculiarly doomed to violence. Collier's regression results lead him to the upbeat conclusion that African countries can enjoy the best of both worlds because democracy eliminates the potentially negative effects of ethnic diversity on economic growth, while high ethnic diversity reduces the risk of violent conflict. But therein lies Africa's dilemma. The critical condition for this happy future is political, but it is precisely the difficulties in achieving political stability and democracy that are causing the huge outpouring of anguished commentary by Africans and Africa-watchers alike. Collier's conclusion is thus really the start of a discussion on the political economy of ethnicity. Let me indicate some directions in which such a discussion might go.

Ethnicity and History

First, ethnicity must be put in historical context. Ethnicity is never the sole basis of identity; it is relative and situational. The designation of a person as belonging to a particular ethnic group may be more salient in situations where identity based on religious group, clan, occupation, or other social status becomes more relevant. Ethnic identity is not primordial nor is it given for all time. Many cases can be charted in Africa of the emergence, disappearance, fragmentation, or amalgamation of a particular ethnic designation.

O'Brien (1986), for example, describes the emergence of an ethnic group in the Gezira area of Sudan. As irrigated cotton production became established there around 1925, large numbers of seasonal migrant laborers came from places as far afield as northern Nigeria to work and eventually to settle. At first these laborers were referred to as *slave workers*, a term that did not refer to their origin but that clearly placed them at the bottom of the social hierarchy. Over time these newly settled groups began to take on distinctive behaviors, such as establishing more fundamentalist Islamic sects than the people among whom they had settled. Although they had come from various groups in Muslim Hausaland, they came to form a distinctive "ethnic" group of their own. The new name by which they came to be known, one they themselves promoted, derived from a term meaning "religious pilgrims from West Africa."

Peel (1989) shows how the people known in Nigeria as Yoruba represent a category that has expanded since the beginning of this century, when Yoruba referred only to the people of Oyo. From the 1920s on, as migration, education, and cash cropping increased and Christianity and Islam spread, other groups took on the Yoruba identity. Since the 1930s the development of Yoruba political organization has kept pace with organization by opposing groups known as Ibo (Igbo). Peel shows how certain administrative reorganizations under the British contributed to the shaping of this "ethnic" organization.

Ranger (1996, p. 274) argues more explicitly that political, legal, and administrative changes under colonial rule produced "monolithic ethnicities [that] replaced previously much more fluid, multicultural and multilingual networks of interaction and identity." He gives an example from the Republic of Congo, where Ekholm-Friedman and Sundberg (1995) document how the political dynamics of various chiefdoms shifted first during the slave trade and were further dislocated by an oppressive colonial overrule in which newly emerging political alignments took their shape in ethnic form.

These examples and many more show that ethnic identities cannot be read from simple "ethno-linguistic diversity" but are produced by specific historical, political, economic, and social processes, and that their social salience and shape vary. Such ethnic differences do not "cause" economic decline or political violence, whatever the signs in the regressions indicate. What Cruise O'Brien (1996, p. 63) points out with reference to Mali and Senegal can be applied more broadly: in most cases "ethnic identification does not ... take on a sharp adversarial quality." The challenge, then, is to identify the conditions under which ethnic difference is associated with adversarial relations or violence.

Political Tribalism and Transition

Social scientists sharply distinguish ethnicity and ethnic identity from political tribalism. As Lonsdale (quoted in Chabal 1996, p. 49) writes:

Western students of Africa have, until recently, felt defensive about the continent's political ethnicity, anxious to disarm the racist prejudice of our readers. We used to make excuses for Africans: tribalism was not their fault. Now that some European tribes have proved to be more savage than most of Africa's, that timidity is disappearing. But the issue will always be complex. Ethnicity is a world-wide social fact: all human beings make their cultures within communities that define themselves against "others." But we do not always politicise culture; and when we do, it may not necessarily be to pursue a reactionary xenophobia. To imagine the existence of a new "tribe" may be the best way to look outward, to embrace social progress. Students of Africa are now beginning to understand such ambiguity—that while some aspects of ethnicity are indeed inherited and conservative, its meanings are also reinvented every day, to meet new needs. Cultural identity is what people make it rather than what they historically and ineluctably are. And "moral ethnicity"—what I call that contested internal standard of civic virtue against which we measure our personal esteem—is very different from the unprincipled "political tribalism" with which groups compete for public resources.

There are a number of examples of political tribalism in Africa—Kenya, Rwanda, Sierra Leone, and Zimbabwe, to name a few. But in none of these cases can one

argue that ethnic diversity was the cause of violence because people of different ethnic designations—such as the Tutsi and Hutu in Rwanda—lived together peacefully for years. Careful studies of genocide—in Cambodia, Rwanda, Serbia—show clearly the manipulation and orchestration of political violence by political leaders, which they couch in ethnic terms. Once mass violence is set in motion, ethnic definitions tend to sharpen, which escalates fear, violence, and terror. This line of argument suggests that research should focus not on ethno-linguistic diversity but on political tribalism—that is, the orchestration of political conflict and violence in ethnic terms.

I would give much greater emphasis to the point Collier relegates to a footnote, namely, the increased risk of ethnic violence during the transition to democratic rule. All over Africa we are seeing that the more radical are the shifts toward more open, democratic regimes, the greater is the social and political turmoil. When the increased volatility associated with liberalization is added to this turmoil, it creates an incendiary situation.

There are parallels here between countries in Africa and those in Eastern Europe and the former Soviet Union, where civil conflict has increased and the conflict has taken on an ethnic cast. Ethnic conflict is not inevitable; conflict over authority or resources can be based on other social identities. In Albania and Somalia, for example, clashes have occurred between clans that belong to the same ethnic group.

Areas for Future Research

The configuration of political change and economic liberalization in association with an increase in corruption and crime and a breakdown in other forms of regulation and rule desperately demands research, not least because it is in these destabilizing circumstances that the political manipulation of ethnic difference takes shape. It is for this reason that I would like to turn researchers at the World Bank away from mere ethno-linguistic diversity to the political, economic, and social situations in which ethnic difference becomes—or, rather, is made to become—salient.

A particular aspect that warrants research and policy attention is the “lost generation” of African youth. With unemployment high and rising, many young people are being drawn into ragtag armies, participating in mass lootings, and engaging in other dangerous and marginalized activities. Richards (1995, p. 164) suggests that “the youth factor ... may take over from ethnicity” in the study of African politics. In fact, this rising tide of unemployed, undereducated, disenfranchised, and increasingly desperate youth may feed the fires of politically orchestrated ethnic violence.

My reasons for urging much more careful specification and investigation of what is going on under the name of ethnic conflict go beyond the refinement of research. If we do not demystify our key concepts, we are in danger of increasing the ethnization of political conflict. Consider a simple example, cited by Thornton (1996), of how easy it is to mistakenly identify an incident as ethnic or tribal. In 1995 in South Africa a Pedi man killed a Mosotho man in a workers hostel. The incident set off fights in which 15 more people were killed. People who had witnessed the incident or heard about it variously attributed the murder and subsequent fighting to a

case of discovered adultery, political party in-fighting, intrahostel rivalry, and ethnic rivalry. But as Thornton points out, the South African press—as well as the overseas press—usually attributes such episodes to ethnic or tribal violence. In far too many instances Americans and Europeans assume that ethnic difference is the cause of violence or potential violence in Africa. Research that assumes ethno-linguistic diversity to be a social fact risks reinforcing such conventional wisdom rather than exploding it.

Another important topic for research emerging from studies of political competition and conflict in which ethnicity is involved concerns the different specifications of what constitutes political democracy. In most African countries there are not only multiple bases for social identity but also multiple modes of sociopolitical organization. Common examples are forms of chiefly or other traditional or indigenous authority as well as state-organized elective and appointive structures. In my opinion these arrangements are the main reason for Botswana's stability—not its relative ethnic homogeneity.

Some countries also contain ritual and religious loci of authority that have significance for political interactions. The mix varies; the key is the multiplicity of such forms. Perhaps the main problem with the party politics associated with democracy by Europeans and Americans is that it tends to produce a highly divisive style of politics in Africa. Indeed, the English word *politics* has been indigenized in many places in Africa to mean specifically divisive interaction. In rural Botswana, for example, *politics* is likened to *war*. In contrast, political activities in the traditional chief's court are assumed to be contentious but are expected to be worked out through debate. Similar notions are found in other countries.

Thornton (1996) suggests that we should assume that democracy in Africa seeks to produce not a nation-state but rather a country within which there may be several centers of authority along the lines of city-states in medieval Europe. Such a view would lead to a consideration not only of the nostrums of Western liberal democracies—parties, parliaments, individual rights—but also of other political conceptions and forms. One road to preventing the mobilization of ethnic difference for predatory political ends may be the careful specification and fostering of political multiplicity.

It is, then, in these more carefully specified sociopolitical directions that research on the political economy of ethnicity should move. Collier's article is valuable if we interpret it as telling us that we should look much more at the politics of situations in which ethnic difference takes on a conflictual or violent shape, at political and social institutions, and at social justice rather than at ethno-linguistic diversity itself.

References

- Chabal, Patrick. 1996. "The African Crisis: Context and Interpretation." In R. Werbner and Terence Ranger, eds., *Postcolonial Identities in Africa*. London: Zed Books.
- Cruise O'Brien, Donal B. 1996. "A Lost Generation? Youth Identity and State Decay in West Africa." In Richard Werbner and Terence Ranger, eds., *Postcolonial Identities in Africa*. London: Zed Books.

- Eckholm-Friedman, Kajsa, and Anne Sundberg. 1995. "Ethnic War and Ethnic Cleansing in Brazzaville." In Preben Kaarsholm, ed., *From Post-Traditional to Post-Modern? Interpreting the Meaning of Modernity in Third World Urban Societies*. Roskilde, Denmark: IDS.
- Eriksen, Thomas Hylland. 1993. *Ethnicity and Nationalism*. London: Pluto Press.
- Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.
- O'Brien, Jay. 1986. "Towards a Reconstitution of Ethnicity: Capitalist Expansion and Cultural Dynamics in Sudan." *American Anthropologist* 88 (4): 898-906.
- Peel, John D. Y. 1989. "The Cultural Work of Yoruba Ethnogenesis." In Elisabeth Tonkin, Maryon McDonald, and Malcolm Chapman, eds., *History and Ethnicity*. London: Routledge.
- Ranger, Terence. 1996. "Postscript: Colonial and Postcolonial Identities." In Richard Werbner and Terence Ranger, eds., *Postcolonial Identities in Africa*. London: Zed Books.
- Richards, Paul. 1995. "Rebellion in Liberia and Sierra Leone: A Crisis of Youth." In Oliver Furley, ed., *Conflict in Africa*. New York: I.B. Tauris.
- Thornton, Robert. 1996. "The Potentials of Boundaries in South Africa: Steps Towards a Theory of the Social Edge." In Richard Werbner and Terence Ranger, eds., *Postcolonial Identities in Africa*. London: Zed Books.

Comment on “The Political Economy of Ethnicity,” by Paul Collier

Jennifer Widner

Paul Collier’s article cleverly combines research in economics and political science. First I will say some nice things about the article. Then I will quibble with the model. Finally, I will offer suggestions on how future researchers might proceed.

Defining Identities

Analysis relying on an ethnic fractionalization index raises concern about the implicit notion of ethnicity being used. Forty years of research have improved our understanding of identity politics in Africa and in other parts of the world, leading to three main conclusions. First, ethnic identities are socially constructed, highly malleable, and situationally defined. Econometric analyses that use “counts” of ethno-linguistic groups, often borrowed from early research, ignore this fact by equating the existence of a perceived difference in identity at one historical juncture with an enduring, inflexible characteristic. In a rough-hewn way the interaction variables in Collier’s models capture some of what noneconomists mean when they say that ethnicity is situationally defined and varies in salience across settings. In this respect, the underlying understanding of ethnicity in the article is a good deal more sophisticated than the concepts implicit in the work on which it draws.

Second, most people hold not one identity but a portfolio of identities—race, linguistic group, lineage, clan—from which researchers may select, depending on the setting. Third, an ethnic fractionalization index ignores the fact that the intensity of attachment to a particular identity among the many in an individual’s portfolio usually varies across contexts.

Third, we know that politics matters. Collier’s article is a significant improvement over most econometric analyses that use an ethnic fractionalization index because it recognizes that political institutions may reduce the salience of ethnic identity in political life. The introduction of regime variables helps address this concern.

Jennifer Widner is associate professor of political science at the University of Michigan.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

Collier's Findings

Collier's article asks us to give explicit consideration to the role of institutions in shaping patterns of growth and violence. Given what the World Bank and others have learned over the past 10 years, the inclusion of such variables in econometric analyses is overdue. Yet the task is difficult. Good cross-national and longitudinal data on institutional features are very hard to find. Robert Bates (see elsewhere in this volume) is acquiring and organizing such information for Africa, but the effort is expensive and time-consuming. Although one may fault Collier's article for the measures used and the narrow range of institutional features they take into account, this criticism merely highlights the need to invest in this type of data collection exercise.

The article finds that the relationship between ethnicity and civil violence is non-monotonic. Scholars such as Donald L. Horowitz (see elsewhere in this volume) have observed this relationship before. In homogeneous societies and societies with many different communities, ethnicity is usually less salient in political life than it is when two, three, or four groups compete for power. Collier explains this phenomenon through the lower coordination costs of rebellion if potential rebels are ethnically differentiated from government supporters.

Finally, Collier's finding that Africa's civil wars are more the product of resource scarcity than of ethnic diversity corresponds with the observations of many specialists in the region's politics. The concentration of strategic installations in a vulnerable single location, the inability to project government into rural areas, the permeability of borders—all are functions of income level and all contribute to civil violence.

The Link to Political Science

All this said, as a political scientist I am a bit confused by this article, because its models and conclusions are not those that would first come to mind. The main reason for my puzzlement lies with the problem of endogeneity. Collier's models suggest that in resource-poor environments, pluralistic societies experience higher economic growth in the presence of democratic political institutions. The theory behind the equations is presumably that where people of different backgrounds can vote and articulate their views, levels of trust are higher and people are more likely to trade and invest.

Political scientists, however, find that resource scarcity, social pluralism, and stable democracy rarely co-exist. It is difficult to create democratic political systems under the conditions described in Collier's model. Income and subnational pluralism often interact. There are many reasons; let me cite just three.

First, even if not politically salient, kin ties are usually stronger in poor countries than in rich. Why? Kin substitutes for a social security system, a bank, unemployment compensation, and an information network, among other things. Kin ties may be latent, but they can be activated in a pinch. In industrial societies the availability of substitutes through the banking system, social programs, and the media reduce

dependence on kin and ethnic group for vital functions, making these identities less available in political life.

Second, where the high cost of transport makes it hard for political parties to project into the rural areas where most people in developing countries live, the easiest way to activate a base of support in an election is through kin ties.

Third, where resources are limited and concentrated in government, the alternative bases for interest group formation are limited. Ethnic organization has few competitors in the political arena; occupation and class are less important than they are in other regions.

For these reasons and others, it is almost a truism in political science that low income and cultural heterogeneity make democracy difficult to achieve. The recent adoption of multiparty politics in many African countries does no violence to this observation. If one looks carefully at the data on civil and political liberties (the Gastil Index used in Collier's models), the indicators display high volatility that is not reflected in the long-term averages often used in econometric analyses. That suggests that stable democracy is very hard to build quickly in these settings.

The underlying distribution of these data and others is bimodal. We have two clusters, with very few observations of countries that have high pluralism and high income or low pluralism and low income. This distribution suggests that endogeneity tends to produce path dependence. Poverty makes it difficult to establish democracy, and the absence of democracy increases the salience of ethnicity, lowers trust, and depresses income. Thus ours is a world of virtuous and vicious cycles—and breaking out of a vicious cycle is difficult.

In a footnote Collier acknowledges that the transition to democracy may involve a temporary increase in the risk of violence. Certainly, that observation is consistent with recent events. Elections mobilize rather than reduce ethnic conflict in many countries. Rwanda is one example, although what happened there can be explained only partly in this way. In Kenya opposition parties have divided into ethnic blocks. The same is true to greater or lesser degrees in many—but not all—African countries.

For social scientists the patterns in the data suggest two things. The first is that different models may be required to capture causal relationships at different income levels. At one income level the concerns that have long been of interest to political scientists may prove central; at another it becomes possible to model relationships as Collier suggests. The second implication is that use of longitudinal data might be considerably more interesting than cross-sectional analyses. The longitudinal research must proceed within regions or across regions with caution in the choice of appropriate comparison periods.

Modifying Models

Collier's models are exciting and suggestive, but the data do not allow him to test them adequately. Measures of key concepts may yield faint images of reality. For

example, the use of telephone access as a proxy for social capital is deeply flawed, even if it is the best that cross-national datasets can offer.

The findings may be marred by bad data. Still, Collier's article suggests that even *when using available measures we should not succumb to the ethno-pessimism that other econometric models induce*. Using the same data but slightly different models, we get wildly different results. The aim should be to more concretely identify the source of the discrepancies.

Floor Discussion of "The Political Economy of Ethnicity," by Paul Collier

Paul Collier (presenter) noted that as an economist he was in an unenviable position in having to defend himself from the challenges raised by an anthropologist (discussant Pauline E. Peters) and a political scientist (discussant Jennifer Widner). Collier asserted that the relationships between democratization and ethno-linguistic fractionalization are statistically significant, but acknowledged that it may not be clear what those relationships mean. That ethno-linguistic fractionalization is poorly measured is clear. But anthropologists and political scientists created that variable, not economists. And while ethno-linguistic fractionalization may be a poorly measured variable, Collier doubted that it was so poorly measured as to be meaningless. Collier challenged anthropologists to develop better measures of ethnic diversity.

Anthropologists had abandoned the ethno-linguistic fractionalization measure, said Peters, because it did not reflect social and political realities. But her larger concern was that an analysis relating a misspecified or underspecified variable to various economic and political phenomena could be manipulated by the same political leaders promoting ethnic division in Africa. Peters argued that it was more important to focus on the political conditions that defined what ethnic diversity did and to identify the economic determinants of political change that will not support the mobilization of ethnic violence.

A participant from the World Bank asked how education could lower the costs of building institutions and traditions that foster democracy. Collier said that his research had explored the relationship between education and democracy but had not found a statistically significant relationship.

Another participant from the World Bank wondered about the post-traumatic costs of ethnic conflicts even after democracy has been established. In Africa, Eastern Europe, and Latin America the psychological, educational, and health costs of ethnic conflict have been quite high. How can these costs be assessed and dealt with?

Economists can quantify the post-traumatic costs of conflict, said Collier—as with health status, for example. But what Collier had studied was economic growth in

This session was chaired by Jane Armitage, director, External Affairs, at the World Bank.

Annual World Bank Conference on Development Economics 1998

©1999 The International Bank for Reconstruction and Development / THE WORLD BANK

post-conflict situations. He found that countries grow very quickly in the first five years of peace after long civil wars—by an average of about 6 percent a year. After short civil wars the economy continues to decline as if the war were still under way. Collier hypothesized that after a short civil war the notion that a society is prone to war produces a sustained flight of financial and physical capital. But after a long civil war that exodus has already occurred, and the economy recovers.

Peters cited recent research at Yale University that has found that extensive documentation can lower post-traumatic costs and reduce the possibility that atrocities will be perpetuated. In Cambodia and Rwanda, for example, the extensive attention paid to recent instances of genocide will make it difficult for political leaders to mobilize ethnic violence in the same way in the future.

Widner added that social norms suffer when communities are vulnerable to crime by outsiders. It is extremely upsetting when members of a kin group cannot police the activity of fly-by-night criminals. If random and unpredictable crime increases in southern Africa, it will lead to ethnic violence simply because a lot of social norms will break down. Donors do not work with police for obvious reasons, noted Widner, but in her view fighting crime would reduce ethnic violence.

A participant from the Development Assistance Committee of the Organisation for Economic Co-operation and Development noted that Arthur Lewis had reached a conclusion similar to Collier's in 1965, when he wrote that democracy was the only solution to managing ethnically diverse African countries. The Development Assistance Committee has been working with developing countries in Africa to explore their agendas for democratic governance. It was learning that efforts to improve African politics have to happen on the ground, working with the relevant actors. Given that, how should efforts to strengthen democratic governance proceed?

Peters argued that it is essential not to take an ethnocentric view of what democracy is. Party politics, for example, can increase violence in a number of ways. Thus many open political discussions in Africa do not view the U.S. model of party politics as the only form of democracy. That is not to say that party politics should be rejected. But one has to be careful about the form that is promoted. In Peters's view a lot of donor efforts in Africa increase conflict. Multiple, often conflicting objectives are pursued in the same areas. Coordination may be a goal, but it is not being achieved. And uncoordinated aid may contribute to conflict and to competition, which can then take ethnic form.

Widner added that in her research she had found that ethnicity matters in some places but not in others. It often does not matter in villages where many different groups are living in the same place, because people have to find ways to get along with each other. Ethnicity becomes more important in national politics, particularly in democratic electoral politics. One thing the World Bank and other donors could do to break this pattern, said Widner, is to build roads. Health care and education often have limited reach because of transport problems, and democratic politics may suffer from similar constraints.

Distributors of World Bank Group Publications

Prices and credit terms vary from country to country. Consult your local distributor before placing an order.

ARGENTINA

World Publications SA
Av Cordoba 1877
1120 Ciudad de Buenos Aires
Tel (54 11) 4815-8156
Fax (54 11) 4815-8156
E-mail wpbooks@infovia.com.ar

AUSTRALIA, FIJI, PAPUA NEW GUINEA, SOLOMON ISLANDS, VANUATU, AND SAMOA

D A Information Services
648 Whitehorse Road
Mitcham 3132, Victoria
Tel (61 3) 9210 7777
Fax (61 3) 9210 7788
E-mail service@dadirect.com.au
URL www.dadirect.com.au

AUSTRIA

Gerold and Co
Weihburggasse 26
A-1011 Wien
Tel (43 1) 512-47-31-0
Fax (43 1) 512-47-31-29
URL www.gerold.co.at online

BANGLADESH

Micro Industries Development Assistance Society (MIDAS)
House 5, Road 16
Dhanmondi R/Area
Dhaka 1209
Tel (880 2) 326427
Fax (880 2) 811188

BELGIUM

Jean De Lannoy
Av du Roi 202
1060 Brussels
Tel (32 2) 538-5169
Fax (32 2) 538-0841

BRAZIL

Publicações Técnicas Internacionais Ltda
Rua Peixoto Gomide, 209
01409 Sao Paulo, SP
Tel (55 11) 259-6644
Fax (55 11) 258-6990
E-mail postmaster@pti.uol.br
URL www.uol.br

CANADA

Renouf Publishing Co Ltd
5369 Canotek Road
Ottawa, Ontario K1J 9J3
Tel (613) 745-2665
Fax (613) 745-7660
E-mail order_dept@renoufbooks.com
URL www.renoufbooks.com

CHINA

China Financial & Economic Publishing House
8, Da Fo Si Dong Jie
Beijing
Tel (86 10) 6401-7365
Fax (86 10) 6401-7365
China Book Import Centre
P O Box 2825
Beijing

Chinese Corporation for Promotion of Humanities

52, You Fang Hu Tong,
Xuan Nei Da Jie
Beijing
Tel (86 10) 660 72 494
Fax (86 10) 660 72 494

COLOMBIA

Infoenlace Ltda
Carrera 6 No 51-21
Apartado Aereo 34270
Santafé de Bogotá, D C
Tel (57 1) 285-2798
Fax (57 1) 285-2798

COTE D'IVOIRE

Center d'Édition et de Diffusion Africaines (CEDA)
04 B P 541
Abidjan 04
Tel (225) 24 6510, 24 6511
Fax (225) 25 0567

CYPRUS

Center for Applied Research
Cyprus College
6, Diogenes Street, Engomi
P O Box 2006
Nicosia
Tel (357 2) 59-0730
Fax (357 2) 66-2051

CZECH REPUBLIC

USIS, NIS Prodejna
Havelkova 22
130 00 Prague 3
Tel (420 2) 2423 1486
Fax (420 2) 2423 1114
URL www.nis.cz/

DENMARK

Samfundslitteratur
Rosensøerms Alle 11
DK-1970 Frederiksberg C
Tel (45 35) 351942
Fax (45 35) 357822
URL www.sl.cbs.dk

ECUADOR

Libri Mundi
Libreria Internacional
P O Box 17-01-3029
Juan Leon Mera 851
Quito
Tel (593 2) 521-606,
(593 2) 544-185
Fax (593 2) 504-209
E-mail librimu1@librimundi.com.ec
E-mail librimu2@librimundi.com.ec

CODEU

Ruiz de Castilla 763, Edif Expocolor
Primer piso, Of #2
Quito
Tel/Fax (593 2) 507-383, 253-091
E-mail codeu@impsat.net.ec

EGYPT, ARAB REPUBLIC OF

Al Ahram Distribution Agency
Al Galaa Street
Cairo
Tel (20 2) 578-6083
Fax (20 2) 578-6833
The Middle East Observer
41, Sherif Street
Cairo
Tel (20 2) 392-6919
Fax (20 2) 393-9732
E-mail mafouda@meobserver.com.eg
URL www.meobserver.com.eg

FINLAND

Akateeminen Kirjakauppa
P O Box 128
FIN-00101 Helsinki
Tel (358 0) 121 4418
Fax (358 0) 121-4435
E-mail akateemil@stockmann.fi
URL www.akateeminen.com

FRANCE

Editions Eska, DBJ/Offilib
48, rue Gay Lussac
75005 Paris
Tel (33-1) 43-29-21-32
Fax (33-1) 43-29-91-67
E-mail offilib@offilib.fr
URL www.offilib.fr

GERMANY

UNO-Verlag
Poppelsdorfer Allee 55
53115 Bonn
Tel (49 228) 949020
Fax (49 228) 217492
URL www.uno-verlag.de
E-mail unoverlag@aol.com

GHANA

Epp Books Services
P O Box 44
TUC
Accra
Tel 223 21 778843
Fax 223 21 779099

GREECE

Papasotiriou S A
35, Stourmaria Str
106 82 Athens
Tel (30 1) 364-1826
Fax (30 1) 364-8254

HAITI

Culture Diffusion
5, Rue Capois
C P 257
Port-au-Prince
Tel (509) 23 9260
Fax (509) 23 4858

HONG KONG, CHINA, MACAO

Asia 2000 Ltd
Sales & Circulation Department
302 Seabird House
22-28 Wyndham Street, Central
Hong Kong, China
Tel (852) 2530-1409
Fax (852) 2526-1107
E-mail sales@asia2000.com.hk
URL www.asia2000.com.hk

HUNGARY

Euro Info Service
Margitszegeti Europa Haz
H-1138 Budapest
Tel (36 1) 350 80 24, 350 80 25
Fax (36 1) 350 90 32
E-mail euinfo@mail.mata.vu

INDIA

Allied Publishers Ltd
751 Mount Road
Madras - 600 002
Tel (91 44) 852-3938
Fax (91 44) 852-0649

INDONESIA

Pt Indra Limited
Jalan Borobudur 20
P O Box 181
Jakarta 10320
Tel (62 21) 390-4290
Fax (62 21) 390-4289

IRAN

Ketab Sara Co Publishers
Khaled Eslamboli Ave, 6th Street
DeLaFrooz Alley No 8
P O Box 15745-733
Tehran 15117
Tel (98 21) 8717819, 8716104
Fax (98 21) 8712479
E-mail ketab-sara@neda.net.ir

Kowkab Publishers

P O Box 19575-511
Tehran
Tel (98 21) 258-3723
Fax (98 21) 258-3723

IRELAND

Government Supplies Agency
Orrig an tSolathair
4-5 Harcourt Road
Dublin 2
Tel (353 1) 661-3111
Fax (353 1) 475-2670

ISRAEL

Yozmot Literature Ltd
P O Box 56055
3 Yohanan Hasandir Street
Tel Aviv 61560
Tel (972 3) 5285-397
Fax (972 3) 5285-397

R O Y International

P O Box 13056
Tel Aviv 61130
Tel (972 3) 649 9469
Fax (972 3) 648 6039
E-mail royil@netvision.net.il
URL www.royint.co.il

Palestinian Authority/Middle East

Index Information Services
P O B 19502 Jerusalem
Tel (972 2) 6271219
Fax (972 2) 6271634

ITALY

Libreria Licosa Commissionaria
Sansoni SPA
Via Duca Di Calabria, 1/1
Casella Postale 552
50125 Firenze
Tel (39 55) 645-415
Fax (39 55) 641-257
E-mail licosa@ftbcc.it
URL www.ftbcc.it/licosa

JAMAICA

Ian Randle Publishers Ltd
206 Old Hope Road, Kingston 6
Tel 876-927-2085
Fax 876-977-0243
E-mail irpl@colis.com

JAPAN

Eastern Book Service
3-13 Hongo 3-chome, Bunkyo-ku
Tokyo 113
Tel (81 3) 3818-0861
Fax (81 3) 3818-0864
E-mail orders@svt-eps.co.jp
URL www.bekkoame.or.jp/~svt-eps

KENYA

Africa Book Service (E A) Ltd
Quaran House, Mfangano Street
P O Box 45245
Nairobi
Tel (254 2) 223 641
Fax (254 2) 330 272

Legacy Books
Loita House
Mezzanine 1
P.O. Box 68077
Nairobi
Tel (254) 2-330853, 221426
Fax (254) 2-330854, 561654
E-mail Legacy@form-net.com

KOREA, REPUBLIC OF

Dayang Books Trading Co
International Division
783-20, Pangba Bon-Dong,
Socho-ku
Seoul
Tel (82 2) 536-9555
Fax (82 2) 536-0025
E-mail seamap@chollian.net

Eulyoo Publishing Co., Ltd
46-1, Susong-Dong
Jongro-Gu
Seoul
Tel (82 2) 734-3515
Fax (82 2) 732-9154

LEBANON

Librairie du Liban
P.O. Box 11-9232
Beirut
Tel (961 9) 217 944
Fax (961 9) 217 434
E-mail hsayegh@librairie-du-
liban.com.lb
URL www.librairie-du-liban.com.lb

MALAYSIA

University of Malaya Cooperative
Bookshop, Limited
P.O. Box 1127
Jalan Pantai Baru
59700 Kuala Lumpur
Tel (60 3) 756-5000
Fax (60 3) 755-4424
E-mail umkoop@tm.net.my

MEXICO

INFOTEC
Av. San Fernando No. 37
Col. Torrello Guerra
14050 Mexico, D.F.
Tel (52 5) 624-2800
Fax (52 5) 624-2822
E-mail infotec@rtm.net.mx
URL rtm.net.mx
Mundi-Prensa Mexico S.A. de C.V.
c/Rio Panuco, 141-Colonia
Cauahatemoc
06500 Mexico, D.F.
Tel (52 5) 533-5658
Fax (52 5) 514-6799

NEPAL

Everest Media International Services
(P) Ltd.
GPO Box 5443
Kathmandu
Tel (977 1) 416 026
Fax (977 1) 224 431

NETHERLANDS

De Lindeboom/Internationale
Publicaties b.v.
P.O. Box 202, 7480 AE Haaksbergen
Tel (31 53) 574-0004
Fax (31 53) 572-9296
E-mail lndeboo@worldonline.nl
URL www.worldonline.nl/~lndeboo

NEW ZEALAND

EBSCO NZ Ltd
Private Mail Bag 99914
New Market
Auckland
Tel (64 9) 524-8119
Fax (64 9) 524-8067

Oasis Official
P.O. Box 3627
Wellington
Tel (64 4) 499 1551
Fax (64 4) 499 1972
E-mail oasis@actrix.gen.nz
URL www.oasisbooks.co.nz/

NIGERIA

University Press Limited
Three Crowns Building Jericho
Private Mail Bag 5095
Ibadan
Tel (234 22) 41-1356
Fax (234 22) 41-2056

PAKISTAN

Mirza Book Agency
65, Shahrah-e-Quard-e-Azam
Lahore 54000
Tel (92 42) 735 3601
Fax (92 42) 576 3714

Oxford University Press
5 Bangalore Town
Sharae Faisal
PO Box 13033
Karachi-75350
Tel (92 21) 446307
Fax (92 21) 4547640
E-mail ouppak@TheOffice.net
Pak Book Corporation
Aziz Chambers 21, Queen's Road
Lahore
Tel (92 42) 636 3222, 636 0885
Fax (92 42) 636 2328
E-mail pbc@brain.net.pk

PERU

Editorial Desarrollo SA
Apartado 3824, Ica 242 OF 106
Lima 1
Tel (51 14) 285380
Fax (51 14) 286628

PHILIPPINES

International Booksource Center
Inc
1127-A Antipolo St, Barangay,
Venezuela
Makati City
Tel (63 2) 896 6501, 6505, 6507
Fax (63 2) 896 1741

POLAND

International Publishing Service
Ul. Piekna 31/37
00-677 Warszawa
Tel (48 2) 628-6089
Fax (48 2) 621-7255
E-mail books%ips@ikp.atm.com.pl
URL www.ipscg.waw.pl/ips/export

PORTUGAL

Livraria Portugal
Apartado 2681, Rua Do Carmo
70-74
1200 Lisbon
Tel (1) 347-4982
Fax (1) 347-0264

ROMANIA

Compani De Librarii Bucuresti S.A.
Str. Lipscani no. 26, sector 3
Bucharest
Tel (40 1) 313 9645
Fax (40 1) 312 4000

RUSSIAN FEDERATION

Isdateilstvo <Ves Mir>
9a, Kolpachniy Pereulok
Moscow 101831
Tel (7 095) 917 87 49
Fax (7 095) 917 92 59
ozimarin@glasnet.ru

SINGAPORE; TAIWAN, CHINA

MYANMAR, BRUNEI
Hemisphere Publication Services
41 Kallang Pudding Road #04-03
Golden Wheel Building
Singapore 349316
Tel (65) 741-5166
Fax (65) 742-9356
E-mail info@hemisphere.com.sg

SLOVENIA

Gospodarski vestnik Publishing
Group
Dunajska cesta 5
1000 Ljubljana
Tel (386 61) 133 83 47, 132 12 30
Fax (386 61) 133 80 30
E-mail repansej@gvestnik.si

SOUTH AFRICA, BOTSWANA

For single titles
Oxford University Press Southern
Africa
Vasco Boulevard, Goodwood
PO Box 12119, N1 City 7463
Cape Town
Tel (27 21) 595 4400
Fax (27 21) 595 4430
E-mail oxford@oup.co.za
For subscription orders
International Subscription Service
PO Box 41095
Craighall
Johannesburg 2024
Tel (27 11) 880-1448
Fax (27 11) 880-6248
E-mail iss@is.co.za

SPAIN

Mundi-Prensa Libros, S.A.
Castello 37
28001 Madrid
Tel (34 91) 4 363700
Fax (34 91) 5 753998
E-mail libreria@mundiprensa.es
URL www.mundiprensa.com/
Mundi-Prensa Barcelona
Consell de Cent, 391
08009 Barcelona
Tel (34 3) 488-3492
Fax (34 3) 487-7659
E-mail barcelona@mundiprensa.es

SRI LANKA, THE MALDIVES

Lake House Bookshop
100, Sir Chittampalam Gardiner
Mawatha
Colombo 2
Tel (94 1) 32105
Fax (94 1) 432104
E-mail LHL@sri.lanka.net

SWEDEN

Wennergren-Williams AB
P.O. Box 1305
S-171 25 Solna
Tel (46 8) 705-97-50
Fax (46 8) 27-00-71
E-mail mail@wv.se

SWITZERLAND

Librairie Payot Service Institutionnel
C(ites)-de-Montbenon 30
1002 Lausanne
Tel (41 21) 341-3229
Fax (41 21) 341-3235
ADECO Van Diermen
Editions Techniques
Ch. de Lacuz 41
CH1807 Blonay
Tel (41 21) 943 2673
Fax (41 21) 943 3605

THAILAND

Central Books Distribution
306 Silom Road
Bangkok 10500
Tel (66 2) 2336930-9
Fax (66 2) 237-8321

TRINIDAD & TOBAGO AND THE CARIBBEAN

Systematics Studies Ltd
St. Augustine Shopping Center
Eastern Main Road, St. Augustine
Trinidad & Tobago, West Indies
Tel (868) 645-8466
Fax (868) 645-8467
E-mail tobe@trinidad.net

UGANDA

Gustro Ltd
PO Box 9997 Madhvani Building
Plot 16/4 Jinja Rd
Kampala
Tel (256 41) 251 467
Fax (256 41) 251 468
E-mail gus@swiftuganda.com

UNITED KINGDOM

Microinfo Ltd
P.O. Box 3, Omega Park, Alton,
Hampshire GU34 2PG
England
Tel (44 1420) 86848
Fax (44 1420) 89889
E-mail wbank@microinfo.co.uk
URL www.microinfo.co.uk
The Stationery Office
51 Nine Elms Lane
London SW8 5DR
Tel (44 171) 873-8372
Fax (44 171) 873-8242
URL www.the-stationery-
office.co.uk/

VENEZUELA

Tecni-Ciencia Libros, S.A.
Centro Ciudad Comercial Tamanaco
Nivel C2, Caracas
Tel (58 2) 959 5547, 5035, 0016
Fax (58 2) 959 5636

ZAMBIA

University Bookshop, University of
Zambia
Great East Road Campus
P.O. Box 32379
Lusaka
Tel (260 1) 252 576
Fax (260 1) 253 952

ZIMBABWE

Academic and Baobab Books (Pvt)
Ltd
4 Conald Road Graniteside
P.O. Box 567
Harare
Tel 263 4 755035
Fax 263 4 781913

THEMES AND PARTICIPANTS FOR THE
11TH ANNUAL WORLD BANK CONFERENCE ON
DEVELOPMENT ECONOMICS (ABCDE)

APRIL 28–30, 1999

Economic Architecture

Short-Term Capital Movements
Issues in Global Economic Architecture
Corporate Governance
Openness and Transparency



Social Issues

Crimes and Violence
Concept and Measurement of Exclusion
Land Reform
Inequality and Poverty



Economics of Transition

Macroeconomic Stabilization and Financial Reform
Privatization, Competition, and Regulation Policy
China
Russia



Philippe Aghion • Kenneth Arrow • Anders Åslund • Abhijit Banerjee

Erik Berglof • Amar Bhattacharya • Francois J. Bourguignon • Ann Florini

Kenneth A. Froot • Alan Gelb • Leonid Grigoriev • Karla Hoff • Ravi Kanbur

David S. Landes • Florencio Lopez-De-Silanes • Glenn C. Loury • Nora Lustig

Yingyi Qian • Dani Rodrik • Mauricio Rubio • Nicholas Stern

Joseph E. Stiglitz • Andres Velasco • James D. Wolfensohn • Charles Wyplosz



THE WORLD BANK

1818 H Street N W
Washington, DC 20433, U S A
Telephone (202) 477-1234
Facsimile (202) 477-6391
Telex MCI 6415 WORLDBANK
or MCI 248423 WORLDBANK

World Wide Web <http://www.worldbank.org/>
E-mail books@worldbank.org

The Annual World Bank Conference on Development Economics (ABCDE) brings together the world's leading scholars and development practitioners for a lively debate on state-of-the-art thinking in development policy and the implications for the global economy. The tenth conference dealt with four difficult topics in development: the role of geography in countries' success, with Paul Krugman, John Luke Gallup and Jeffrey Sachs, the role and design of regulation and competition policy, with Paul Joskow and Jean-Jacques Laffont, the causes of financial crises and ways to prevent them, with Bruce Greenwald, Asli Demirguc-Kunt and Enrica Detragiache, and how ethnic conflict affects democracy and growth, with Paul Collier and Donald L. Horowitz. This volume also includes insights from Nobel Prize winner James Tobin, Stanley Fischer, Joseph E. Stiglitz, and World Bank President James D. Wolfensohn.

ALSO IN THIS VOLUME*

COMMENTS BY

Isher Judge Ahluwalia • Robert Bates • Michael Chege • Gerard Caprio Jr
Paul A. David • Charles A. E. Goodhart • Ana Julia Jatar-Hausmann • Thomas Hellmann
J. Vernon Henderson • Ronald McKinnon • Pauline Peters • James E. Rauch
Richard Schmalensee • Kotaro Suzumura • Anthony Venables • Jennifer A. Widner



COVER DESIGN BY COMMUNICATIONS DEVELOPMENT INCORPORATED