Annual World Bank Conference on Development Economics 2001/2002

Boris Pleskovic
Nicholas Stern

**WORLD BANK AND
OXFORD UNIVERSITY PRESS**

Annual

World Bank

Conference

on Development

Economics

2001/2002

Edited by **Boris Pleskovic and Nicholas Stern**

*A copublication of the World Bank
and Oxford University Press*

Contents

Health and Development

About This Book

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. Participants' affiliations identified in this volume are as of the time of the conference, May 1–2, 2001.

Boris Pleskovic and Nicholas Stern edited this volume.

Developmental and copyediting services were provided by Meta de Coquereaumont and Allison Strong of Communications Development, Inc. Cover design by Communications Development, Inc. Book production and dissemination were coordinated by the World Bank Publications team.

This Page Intentionally Left Blank

Introduction

Boris Pleskovic, Nicholas Stern, and F. Desmond McCarthy

The Annual World Bank Conference on Development Economics is aimed at expanding the flow of ideas among development researchers and policymakers from around the world. By fostering a better understanding of development and the problems developing countries face, the conference leads to more informed policymaking at the World Bank. It also provides a forum for exposition by academics and practitioners as they seek to identify and elaborate on new trends and issues in development.

The 13th annual conference, held at the World Bank on 1–2 May 2001, brought together a broad spectrum of people concerned with development—participants from developing countries, within and outside government, and participants from universities, think tanks, nongovernmental organizations, and international financial institutions. The conference focused on two broad themes: globalization and health. It opened on the first day with remarks by James Wolfensohn, followed by four papers on globalization and inequality. On the second day it opened with a keynote address by Nicholas Stern, followed by four papers on health and development. Each afternoon a variety of workshops provided an opportunity for discussing more specialized topics.

James D. Wolfensohn, in his opening remarks, outlines the objectives of the World Bank and the ways in which it seeks to contribute to development. Addressing poverty is at the core of the Bank's work, and there is growing recognition that this work must include many dimensions, such as social justice and issues relating to gender, opportunity, and safety. While the Bank seeks to tackle many of these issues directly, the debate suggests that economic growth is a prerequisite of poverty reduction.

The expected growth in the global population over the next quarter century increases the urgency of addressing questions of poverty and development. Wolfensohn sees the need to address those issues over the next 25 years in a way that leads to greater equity, a greater sense of peace, and a greater sense of opportunity.

Boris Pleskovic is research administrator at the World Bank. Nicholas Stern is senior vice president, Development Economics, and chief economist at the World Bank. F. Desmond McCarthy is a consultant at the World Bank.

That leads to considerations of globalization—the fear of globalization and the challenges of globalization, including trade, communications, the environment, health, migration, drugs, and crime. A key question is how to alleviate the burdens of globalization while taking advantage of its opportunities. Another major concern is resolving conflict, ranging from issues of fundamentalism in Asia to civil wars in Africa to serious social unrest in Latin America.

For the Bank a key concern is how to choose what to do while maintaining a sense of direction. Decisions cannot be made in Washington about what programs to offer countries. Programs have to be country based and country owned. To define roles and responsibilities for all concerned, the Bank must take a broad perspective and exercise selectivity at the country level through a participatory process involving government, regional development banks, civil society, and the private sector.

Nicholas Stern, in his keynote address, details a strategy for development that draws on a wide range of perspectives. He suggests two pillars as the basis for this strategy: building an investment climate that facilitates investment and growth and empowering poor people to participate in that growth. Unlike older stories of capital accumulation, the notion of investment climate tries to be specific and quantitative about the environment for entrepreneurship, productivity and investment—governance, institutions, learning, and interaction are crucial to the story. Stern emphasizes three types of investment that are important in empowering poor people: investment in education, in social protection, and in participatory processes and social and economic inclusion. Both pillars are essentially processes where learning and participation are central. He outlines a research program in support of the strategy for development. The investment climate can be understood in terms of a combination of macroeconomic and trade policies, governance and institutions, and the functioning of infrastructure. Empirical work at the firm level is crucial to a deeper understanding of the problems faced by enterprises in developing countries. We also need to probe more deeply into the challenge of changing institutions and governance. There is also a large and potentially fruitful theoretical agenda in developing genuinely dynamic analysis of the processes described. The concept of empowerment suggests a research agenda focusing on participation, social inclusion, education, health, and social protection. The role of international financial institutions in resource transfer to developing countries has become much smaller. They must work to accelerate development and poverty reduction by serving as *agents of change*, helping developing countries to change the way processes work and things are done in both public and private sectors.

Globalization and Inequality

Kevin O'Rourke poses a timely question: Does globalization lead to the world becoming a more equal place, or does it lead to the rich getting richer and the poor getting poorer? He analyzes this question by exploring the historical links between trade, migration, and capital flows, on the one hand, and inequality, on the other. He distinguishes between two dimensions of inequality, that between countries and that within countries.

In the 19th century globalization was positively associated with between-country convergence, while the evidence from the late 20th century is still inconclusive. If globalization is associated with convergence, why was there so much between-country divergence in the late 19th and 20th centuries? The evidence points not to globalization, but to superior technological progress in rich countries.

O'Rourke concludes with a note of optimism. The trend of rising inequality over the past 200 years, primarily between countries, now appears to have been reversed, and the experience of the 19th century suggests that increased globalization will accelerate this decline.

Daniel Cohen also examines inequality but in a somewhat different context. He focuses on returns to human capital. Noting that inequality rose in the 1980s, Cohen concedes that globalization alters the distribution of income across agents and sectors. But he argues that other factors warrant analysis, such as skill-biased technological progress, labor market deregulation, and changes in the organization of production.

Contrary to conventional views about the effects of globalization, Cohen finds that the social returns to education do not appear to have risen in rich countries and, if anything, may have risen in poor countries. Among poor countries, the rewards to human capital differ between closed and open economies. In countries where the constraints to trade are binding, trade appears to raise the returns to education, while the returns appear to be uniformly high in open countries. In the countries that shifted to more open trade regimes, inequality rose—more so in the countries that initially had low levels of education. Being open or closed does not in itself help bridge the gap between rich and poor countries.

Richard N. Cooper examines the effect of foreign trade and investment on the distribution of income, focusing on developing countries. He suggests that a major disturbance can either reduce or increase average income and can either reduce or increase income inequality—and that these two effects need not be closely related. He argues that there are no compelling reasons, theoretical or empirical, to conclude that, in general, trade promotes growth. There is some aggregate evidence, however, that foreign direct investment has a significant growth-enhancing effect, especially where adequate skills are locally available.

In the second half of the 20th century the world economy turned in one of its strongest growth performances, lifting more people out of poverty than ever before. Cooper contends that while trade liberalization was a major contributor, the improvement in growth cannot be attributed to trade alone. Many other factors played a role, including the relatively peaceful period, improved macroeconomic management, and better international institutions. Cooper suggests that rather than focusing our interest and research efforts on global inequality, we should focus on the most efficacious ways to avoid or mitigate its undesirable consequences.

Anthony J. Venables, exploring the relationship between geography and international inequality, evaluates the prediction that new technologies mean the death of distance. He concludes that many of the effects of technological innovation will be similar to those of the transport revolutions of the 19th century, which led to a con-

centration of manufacturing activities in a few countries to benefit from increasing returns and clustering. New technologies will allow some "weightless" activities to relocate, but other activities may become more entrenched in established locations, which have the advantage of dense networks of related activities.

Thus new technologies will change economic geography—by allowing some countries to benefit from the relocation of some activities while leaving many other countries out. Activities that are generally more complex—knowledge intensive and requiring face-to-face communication—will become more deeply entrenched in high-income countries, typically in cities. By contrast, activities that are more readily transportable and less dependent on face-to-face communication may relocate to lower-wage countries. Because of clustering, however, this relocation of activities may result in rapid development for only a small number of countries.

Health and Development

Morten Rostrup, of Médecins sans Frontières, reviews the role of nongovernmental organizations in providing health care. Although nongovernmental organizations are heterogeneous, many receive substantial funding from governments. This, he suggests, raises questions about whether they can maintain operational independence.

Moreover, nongovernmental organizations increasingly are expected to work within the long-term development perspectives and peace-building activities of various "strategic frameworks." Médecins sans Frontières tries to distance itself from this way of thinking, believing that humanitarian action should be independent of initiatives that advocate models for development or social change. Rostrup argues that using humanitarian action as a first step to promoting peace and democracy or a free market–neoliberal agenda threatens a fundamental principle. Instead, he contends, humanitarian groups should provide aid solely on the basis of need.

Médecins sans Frontières rejects a neoliberal order that risks the lives and dignity of millions of people in the name of some future economic benefit that, given enough time, will "trickle down" to the poor. And it seeks to avoid being manipulated to support the system that gave rise to the misery in the first place.

Finally, Rostrup poses a question: Who should be responsible for delivering health care? Should the task be left to the politicians responsible for failure in the first place, or to multinational drug companies whose research and sales are driven by market concerns? Humanitarian groups, he argues, need to stimulate a sense of responsibility among governments and the international community so that health care delivery addresses the needs of all people.

Jean O. Lanjouw focuses on a more specialized topic—the implications of patent protection for pharmaceutical prices and research and development. She distinguishes between two types of global diseases, those specific to developing countries, such as malaria, and those that affect people everywhere but are generally thought of as rich country diseases, such as cancer and heart disease.

Lanjouw outlines salient features of the global pharmaceutical market. Private pharmaceutical companies devote most of their attention to rich country diseases.

Under today's patent arrangements these companies seek to sell their products in both rich and poor countries at prices often too high for the poor. They show little interest in investing in therapies for poor country diseases, expecting to face difficulties in recouping their investment.

Lanjouw proposes a patent policy that addresses both these concerns. The basic concept is requiring companies requesting a license to make foreign filings, but not use that permission to restrict the sale of those drugs in a "poor" country. This would have the effect of lowering prices of pharmaceuticals in developing countries on diseases such as cancer and heart disease, while at the same time maintaining incentives for R&D into diseases such as malaria. The proposed policy, she contends, would have the effect of lowering the prices in developing countries of pharmaceuticals for rich country diseases while also maintaining incentives for research and development for therapies for poor country diseases.

Anne Case analyzes the links between income and health. It is generally assumed that there is a strong relationship between the two throughout the world. But the direction of causality is an important question for those involved with health policy. Research on this issue has been limited by the lack of data on the health status of individuals, as most surveys tend to be based on the household. In addition, to study the causal effect of income on health, one needs to identify a source of income that is not determined by an individual's health status.

Case resolves the causality issue by considering the role of the state old age pension, a function only of age, in South Africa, using data from an integrated survey of individual health and economic well-being. Her empirical analysis shows that income has a causal effect on health status through several channels, including nutritional status, sanitation and living standards, and reduction of psychosocial stress. Case concludes that governments interested in improving health status may find the provision of cash benefits to be one of the most effective policy tools.

Tomas Philipson and Rodrigo Soares note the common concern that while some countries are getting richer, poor developing countries are falling behind. Per capita growth rates over the period 1962–97 seem to confirm that poor countries are not catching up in income. But in other dimensions of welfare, such as life expectancy, different results emerge. Countries beginning the period with low life expectancy tended to gain more in longevity than those starting with high life expectancy. Thus looking only at income to draw conclusions about changes in overall economic welfare across nations may be misleading.

The United Nations Development Programme has sought to account for nonmaterial dimensions of welfare through the Human Development Index, which includes education and longevity as measures of human welfare. Philipson and Soares argue that this index has limitations because of the "arbitrary way" in which it is constructed. They propose an alternative, an income-equivalent measure of welfare, based on the value the population attaches to gains in different dimensions of development. They argue that their proposed measure is less arbitrary than the Human Development Index.

* * *

Opening Address

James D. Wolfensohn

I would like to welcome you all to this year's Annual World Bank Conference on Development Economics. It is a time when several issues of global consequence have emerged. Of these issues, two in particular—extraordinary poverty and the fear of globalization generated by global inequality—pose critical challenges today and important questions for your research. These challenges also emerged at two recent events that I would like to talk about briefly: the Spring Meetings of the World Bank and the International Monetary Fund (IMF) and the Summit of heads of state of 35 countries of the Americas in Quebec City, Canada.

At this year's Spring Meetings of the Bank and the IMF, held in Washington, D.C., there was a recognition that addressing poverty involves not just dollars, but also voice, freedom, a sense of social justice, and concerns of gender, opportunity, and security. This view of poverty emerged from the remarkable work by some Bank economists and sociologists in the study *Voices of the Poor* (Narayan and others 2000). There is a much more humane sense today of what goes into the assessment of poverty and the issues that must be confronted.

We also had the classic debate—made weaker this year by the absence of Larry Summers—about growth and its role in reducing poverty, that the rising tide of overall economic growth will lift all boats. A minister from Brazil noted that growth lifts all boats except those that have already sunk and those that stand around worrying about navigation, about getting the boat just right, before they too sink.

We reaffirmed our commitment to poverty reduction and all agreed that balanced growth was needed to raise many poor people from living on a dollar day. We all agree that growth is a necessary condition for poverty reduction, a view not inconsistent with the things that the Bank is doing. I also think that we are at a point of change—a very positive one—in moving toward a more balanced relationship between macroeconomic and social issues.

Whether the focus is on developing countries or transition economies, there is now a much greater recognition that social and economic issues are linked.

Policymakers recognize that macroeconomic policy cannot be sound if it does not have a positive impact on people. And they recognize that while the notion of all boats rising is true to a degree, other issues also need to be focused on—equity, social justice, the efficiency of programs.

We are operating in a very challenging environment, if not an atmosphere of near crisis. Of the 6 billion people alive in the world today, 4.8 billion live in the developing and transition economies. These 4.8 billion people receive roughly $6 trillion, about 20 percent, of global GNP. So 80 percent of the world's people receive only 20 percent of the world's income. Moreover, within these developing and transition economies there is significant inequality, and it is increasing in just about every country.

That in itself is a significant challenge, anticipated in the international development goals that the international community has agreed on. But consider the demographics of the next 25 years. Another 2 billion people will be added to the planet, of which just 60 million will be in the developed world. The 4.8 billion in developing and transition economies will become nearly 6.8 billion. This disequilibrium between the developed and developing countries is a challenge that must be faced today, not in 25 years' time. There is no doubt: poverty is the critical global challenge of our time.

Another phenomenon of recent times is the fear of globalization; demonstrations and protests have abounded since the World Trade Organization talks in Seattle in 1999 and are getting bigger. We have to confront the challenges of globalization. Trade, communications, the environment, health, migration, drugs, and crime all pose their own special challenges. And globalization cannot be stopped, notwithstanding the demonstrations and protests.

Nor is it anything new. Globalization has been going on for millennia. The question for us is, how can globalization be managed? How can we alleviate the burdens globalization imposes—and more important, how can we take advantage of its opportunities? At the Spring Meetings that question translated into a slightly different one: how do we deal with the issues that confront us in that new world of 1.2 billion people in developed countries and 6.8 billion people in developing countries? How do we address those issues over the next 25 years in a way that allows us to achieve greater equity, a greater sense of peace, a greater sense of opportunity? And here we are looking to your deliberations over the next two days, hoping they will educate us.

I see that you are going to deal with the question of trade, a central focus at the Spring Meetings. Perhaps we at the Bank have given trade less attention than we should have. But we are now increasing our focus on trade issues in a very significant way, trying to see where the Bank fits in and looking at the impact of trade on globalization and on individual countries.

I was just in Canada for the Quebec City meetings in April. As the discussion on trade among the 35 presidents of the Americas and other participants ensued, it became clear that there were some realities that we had to deal with. If you take the United States, Canada, Brazil, Mexico, and Argentina, you find that those 5 coun-

tries have 95 percent of the combined GDP of the 35 countries. We were dealing with 35 countries, yet in terms of capacity and strength to deal with trade issues we realized that there were enormous built-in inequities. If the United States were to negotiate a hemispheric trade agreement, there would be 50 people doing it. If it were Costa Rica, there might be three. If it were Trinidad and Tobago, there might be half a person. The story for Africa is similar. If you exclude Nigeria and South Africa, you have 40 states with a minority share in the GDP of the continent.

Thus there are inequities not just in incomes, but also in the capacity to deal with global issues (like trade). But building the capacity to address global issues is by itself an enormous challenge. For the World Bank, dealing with 140 individual countries—helping them set up proper legislation, proper judicial systems, and reasonably efficient financial systems and helping them confront corruption and all the elements in their development strategies—is an enormous task. Clearly, the issues go beyond just economics and impinge on local politics and demographics.

How do we deal with conflict? How do we deal with fundamentalism, whether it be in Afghanistan, Chechnya, or elsewhere? In Africa more than 30 countries are involved in conflict; in a third of the continent there are real or nascent conflicts. In many other parts of the world, notably in the Middle East and in places like Colombia, the problems are really serious. Even in Mexico social unrest enters into the overall social and economic equation.

Within all that, how does the Bank operate? One thing is certain: we cannot decide in Washington on the programs we offer countries. They have to be country based and, more important, country owned. We have decentralized considerably in line with our thinking on this. The Bank does not have all the special knowledge on every aspect of the development strategies that countries want to pursue. We are more like a general practitioner, an institution with an overall view of the economic health of a country or region. We have discussions with specialists within countries to help determine what needs to be done and who will do it. Clearly, we will not do everything.

And while retaining this overall capability, the Bank will exercise selectivity at the country level through a participatory process involving government, regional development banks, civil society, and the private sector. This process will also define roles and responsibilities for all concerned, including the Bank. And there has been general agreement on this as an approach.

I would like to know what you think of what I have outlined above; I would also like to hear what you have to say on how that relates to what have been called global public goods. These are issues that affect everybody—issues that from time to time superimpose themselves on national issues. Here you are discussing two of them, trade and communicable diseases. Three others are the environment, the international financial architecture, and knowledge. And there may be still others.

The current manifestation of the issue of communicable diseases is, of course, the debate on AIDS. But the issue is not limited to AIDS. It also incorporates malaria and tuberculosis and measles. Whatever the local diagnosis is, there is also a global diagnosis—this is an issue that affects us all. You could say the same about the envi-

ronment, the international financial architecture, and knowledge—and about the impact of modern technology, which I regard as absolutely central to the next 20 years. I am delighted that Morten Rostrup will be here. He is the head of Médecins sans Frontières, whose global contribution was recognized with a Nobel Prize.

I hope that I have given you some sense of what it is we are trying to do. You will discuss two very important global public goods. The agenda also includes issues such as civil war and community-driven development; I am interested in your ideas on those subjects too. So don't limit yourselves. We need all the help we can get. I certainly need all the help I can get, and I look forward to the results of your deliberations. Thank you.

Reference

Narayan, Deepa, with Raj Patel, Kai Schafft, Anne Rademacher, and Sarah Koch-Schulte. 2000. *Voices of the Poor: Can Anyone Hear Us?* New York: Oxford University Press.

Keynote Address
A Strategy for Development

Nicholas Stern

My purpose here is to outline a strategy for development and to draw some implications for the development research agenda. The title I have chosen deliberately echoes Albert Hirschman's *The Strategy of Economic Development* (1958) and his emphasis on processes rather than prices and quantities. By leaving out "Economic," I do not pretend to be broader or more catholic than Hirschman. But development research has unfolded in many directions since his pioneering 1958 book, and today the adjective *economic* might be interpreted too narrowly for the strategy I outline, one that draws heavily on Joseph Schumpeter's (1934) majestic vision of the dynamics of a market economy. I must mention, too, Amartya Sen, whose work has been central to the broadening of perspective on the meaning of *development* and *poverty reduction.* The title of his book *Development as Freedom* (1999) embodies the key idea of development as the enhancement of individuals' abilities to shape their own lives. The direct and indirect influences of Hirschman's, Schumpeter's, and Sen's understanding of the processes and objectives of development will be clear in the strategy I outline here.

Two pillars form the basis for this strategy for development: building an investment climate that facilitates investment and growth and empowering poor people to participate in that growth. The strategy is, in essence, a strategy for pro-poor growth. I examine the key links between investment climate and empowerment, as the strategic pillars, and the goal of poverty-reducing development. The strategy for development will, I hope, sound very plausible—indeed, almost obvious. But this description is fairly new and is not universally accepted.

The word *investment* in the phrase *investment climate* will evoke memories, for some, of the development philosophies of the 1950s and 1960s, when the emphasis was on growth through capital accumulation. The private sector was mistrusted, and there was little mention of entrepreneurship or social inclusion. Development assistance was seen primarily as the transfer of capital to the countries that had emerged

from colonialism and aspired to join the ranks of industrial countries. Since those early days of development economics, we have, I hope, learned much.

My emphasis here is very different from those of earlier models. It is, first, on growth driven by the private sector and, second, on the empowerment of poor people so that they can participate strongly in the process of growth. Both elements are crucial for fighting poverty.

Consider the role of the private sector. Not only is the private sector the main engine of aggregate growth, it is also the main provider of economic activity for poor people. Of the 1.2 billion people in the world who live on less than $1 a day, only a minute fraction work in the public sector. The growth of the private sector is therefore vital to the reduction of poverty. But that growth will not take place as a simple translation of investment into increased output. What the notion of *investment climate* captures instead is a Schumpeterian view of investment and growth— a view that what matters is not just how much investment takes place but what drives it, what its effects are (in particular, its dynamic effects on productivity and further opportunities), and what happens to the other factors of production and their productivity. In other words, it tries to capture the entire process of investment and growth, not just a simple snapshot at the point of investment.

Just as our view of production processes has been usefully broadened by the concept of investment climate, so too has our understanding of poverty expanded. We have moved to an ex ante rather than an ex post notion. Rather than simply asking whether a person's current income level classifies that person as poor, we ask whether, or to what extent, he or she has the capabilities and environment for action and success. Put another way, in the language of the World Bank's *World Development Report 2000/2001* (2000e), do people enjoy opportunity, empowerment, and security? This trilogy captures both the ex ante "freedom" notion of poverty and the idea that poverty has more dimensions than lack of income. To attack poverty involves, in large measure, empowering poor people to shape their own lives—through opportunities to obtain education and health care, through risk reduction and mitigation, and through participation in the key decisions that affect them and their families.

The Investment Climate

By *investment climate* I mean the policy, institutional and behavioral environment, both present and expected, that influences the risks and returns associated with investment. The notion of investment climate focuses on questions of institutions, governance, policies, stability, and infrastructure that affect not just the level of capital investment but also the productivity of existing investments—indeed, of all factors of production—and the willingness to make productive investments for the longer term.

Seen in this broad way, the investment climate clearly depends on many different aspects of public and private action. It is useful to group these factors under three broad headings. First are *macroeconomic stability and openness*. These embody the usual sound and sensible, if standard, prescriptions for macroeconomic and trade

policy. Second are issues that can be grouped under *good governance and strong institutions*. These issues, my main focus here, include the following:

- Government institutions and behavior (including limits on bureaucratic harassment, especially in administering taxes and regulations); the strength of financial institutions; the rule of law, including law enforcement; and the control of corruption and crime.
- The effectiveness of the government in providing sound regulatory structures for promoting a competitive private sector.
- The effective provision of public services or of the framework for such services, and the quality of the labor force.

Third is the *quality of infrastructure*, including power, water, transport, and telecommunications.

That is a long list, and it looks fairly comprehensive. Let us try to get more concrete by exploring what it means for economic activity at the microeconomic level.

Why Is the Investment Climate So Important?

The key to improving the investment climate can be simply stated as improving the connection between sowing and reaping. This is not just a point about multinational corporations and foreign direct investors; it is, even more important, also a story of the local level, of the microentrepreneur, small businessperson, or farmer. Every day as they work, these people have to make decisions about the investment of their efforts and resources. A key to these investment decisions is the investors' sense of efficacy—their ability to get on with the job and see the rewards. Can investors carry out their efforts and reap the benefits, or will their investment be frustrated by uncertainty, instability, and predation? They will reasonably ask, "Why should I sow if I will not be able to reap a harvest?"

Government cannot guarantee the harvest against all the natural uncertainties of investment (although the social organization of insurance can mitigate the risks of natural adverse events). But government has a fundamental obligation to see that those who sow are not arbitrarily disrupted in their daily activities or robbed of their harvest by opportunistic human intervention. This obligation is even more important than that to protect existing property, because only where the connection between investment and return is protected will investors create new property.

Too often we see societies that have stagnated at a low-level equilibrium because of a dearth of opportunities or incentives for investment. The government may protect the static accumulated wealth of the past (large landed property, for example), and in some countries it has acted as the employer of last resort. But unless it takes steps to encourage entrepreneurial investment in the private sector, or at a minimum avoids stifling such investment, strong growth and poverty reduction are unlikely to ensue.

Consider the Middle East and North Africa, a region that in the 1980s and 1990s suffered the paradox of high investment in human and physical capital juxtaposed with stagnation. Gross domestic product (GDP) per capita declined by 0.8 percent a

year in the 1980s and increased by only 1.1 percent a year in the 1990s. And yet rates of investment in both human and physical capital in the region have been impressively high. One of the primary reasons that such high rates of investment generated so little additional output is that the institutional structure of the labor market systematically misallocated labor. High government wages attracted the most qualified personnel to the public sector, and measures designed to protect existing employment made it difficult for entrepreneurs to start and sustain dynamic businesses.

In many countries that have exhibited slow growth, those who might make entrepreneurial investments of energy and resources are left to be the prey of "bandits," both real and metaphorical. A never-ending stream of arbitrarily imposed rents, taxes, fees, and outright bribes and confiscations will drown even the most dedicated entrepreneurs, washing away all the energy and dynamism that power economic growth and lift people out of poverty. Creating a climate in which entrepreneurs and firms can do good business is crucial to encouraging the types of investment and economic activity that lead to long-term, sustainable economic growth.

This is especially true for small and medium-size enterprises. Anyone who emphasizes the investment climate risks being labeled a narrow-minded advocate for big business. But it is small and medium-size enterprises—and I include microenterprises under this rubric—that account for the majority of firms and a large share of employment in most developing countries, including (although to a lesser extent, on average) the transition economies. Furthermore, it is in these enterprises, including farms, that most of the world's poor people work. By enabling a dynamic small enterprise sector in both rural and urban areas, governments can strengthen income-generating opportunities for poor people while reducing their vulnerability to economic risks.

There are several reasons for this emphasis on small and medium-size enterprises. First, in rural areas off-farm employment in such enterprises can play a vital role in income growth and stability. In India, for example, World Bank research has shown that about a third to a half of rural households' income comes from nonfarm sources (Lanjouw and Shariff forthcoming). Much of this income is from micro, small, or medium-size firms. Nonfarm sources account for a similar share of income in the village of Palanpur, which my research collaborators and I have been following closely since 1974 and where our data go back to 1957 (Lanjouw and Stern 1998). This income comes from a variety of sectors—including commerce, manufacturing, and services—and from regular and part-time wage employment as well as self-employment. Village studies reveal that rural households value nonfarm sources of income highly, not only because they contribute significantly to total income but also because they can reduce households' exposure to potentially devastating income fluctuations associated with bad harvests. A strong investment climate, then, is central to the off-farm employment that can raise incomes for the poor and diversify economic risks.

Second, small-scale agricultural activities are themselves vulnerable to an adverse investment climate. They can suffer as much as, or more than, other activities do from weak governance, malfunctioning infrastructure, and instability.

Third, urban populations in developing countries are likely to increase by some 2 billion in the next 30 years. Small and medium-size enterprises will need to continue to provide employment opportunities for these rapidly growing urban populations.

Fourth, the experiences of the East Asian countries, notably China and Japan, and of the successful transitions in Hungary and Poland have shown us the great importance of urban and rural small and medium-size enterprises in economic development. In most development success stories the growth of such enterprises has been central, particularly in providing employment for poor people.

Moreover, the benefits of the growth of small and medium-size enterprises go beyond economic opportunity. Small firms give breadth and depth to public voice. They have a stake in sound economic and political governance and thus can generate real forces for reform.

The Dynamics of the Investment Climate

The new focus on investment climate differs in important respects from an emphasis on the ordinary notion of investment. In conventional theory investment is expected to lead to diminishing returns. As more investments are made in one place, marginal returns decline and new investment seeks other outlets. The concept of diminishing returns thus implies that investment is self-limiting in each use, as long as other key inputs, including technology, remain unchanged.

The investment climate, by contrast, can be positively or negatively self-reinforcing and thus can generate either prosperity or stagnation. Unlike the older stories of capital accumulation, the notion of investment climate tries to capture some of the external spillover effects, the complementarities, and the nonrivalrous public nature of improvements in governance and institutions. As the investment climate improves, the frontier of opportunity expands: existing investment becomes more productive, the rewards to productive behavior rise, the "animal spirits" (to use a term of which Keynes was fond) of entrepreneurs are encouraged, and the economy tends to attract more investment rather than less. These successful examples of entrepreneurship and investment show other investors what is possible. But they also foster a greater understanding of and commitment to a sound investment climate, thus strengthening the political and economic forces that work toward improving the investment climate. Because there will always be vested interests that benefit from the status quo, strengthening the forces for change is a key part of the process.

A sound investment climate leads to the kind of sustained productivity improvements and vibrant entrepreneurship that induce a virtuous spiral of investment, growth, and poverty reduction. In contrast, where the climate for productive investment deteriorates, these processes work in reverse, and both replacement and new investments can suffer as a pernicious downward spiral sets in. The kinds of dynamic reinforcement or increasing returns I am describing may not apply as strongly in richer countries, where a greater part of the framework of a well-functioning market economy is in place, but they are unlikely to be entirely absent.

From this perspective we may ask whether the reforms implemented by transition economies have enhanced or diminished people's sense of being able to determine their own lives. In too many cases governments used the rhetoric of reform to justify all sorts of half measures and misguided policies that only deepened people's cynical view that the more things change, the more they remain the same or deteriorate. The chipping away at public trust made later efforts to adopt real reforms—which rely so critically on credibility and consensus—that much more difficult. Much of the challenge for those whose job it is to promote development is to understand how to break free of a downward spiral.

Hirschman strongly emphasized the positive feedback dynamics of induced demand and learning in the investment process. For example, he wrote that his "formulation of the development problem . . . calls particular attention to the fact that the use of different economic resources has very different repercussions or 'feedback' effects on the available stocks of these resources" (Hirschman 1958, p. 7). As he put it, investment in the extraction of nonrenewable resources leads to the depletion rather than the augmentation of those resources and to little feedback elsewhere in the economy. By contrast, the investment of capital in a satisfactory investment climate can have significant positive feedback effects. Profits generate the possibility of more investments, linkages to upstream or downstream industries will call forth complementary investments, and success will snowball by breeding confidence in investors and encouraging them to flock with other investors. Hirschman noted the crucial self-augmenting learning effects: "entrepreneurial and managerial ability . . . are resources that increase directly with and through use (much as the ability to play the piano or to speak a foreign language improves with exercise)" (1958, p. 7). Finally, Hirschman (1958, p. 43) drew the implications for policy:

> The complementarity effect of investment is therefore the essential mechanism by which new energies are channeled toward the development process and through which the vicious circle that seems to confine it can be broken. To give maximum play to this effect must therefore be a primary objective of development policy.

This effect has indeed played an important role in more recent models of growth with positive externalities from capital accumulation. It is part of the story of investment that I am telling—but it is not the whole story. Also central are the forces for change in building a stronger investment climate. We can see the investment climate as another and crucial public good for all investors that is distinct from, although compatible with, the kind of investment externalities and complementarities emphasized by Hirschman and modern growth theory.

We can also see a clear difference between the earlier strategy of the international financial institutions, which emphasized the transfer of capital to developing countries, and the strategy described here, which focuses on the investment climate. The volume of investment by these institutions is typically a small part of the total invest-

ment in a country, and viewed in isolation from complementary external effects, it might exhibit diminishing returns. But the purpose of investment projects by international financial institutions is not simply to transfer capital but also to create powerful demonstration effects (promoting "learning by watching" as well as "learning by doing") and to enhance the forces for changes in governance. These effects work on both productivity and the climate of expectations to help crowd in other investments. The emphasis on advancing the transition from centrally planned to market economies and on creating demonstration effects has particularly characterized the approach of the European Bank for Reconstruction and Development (EBRD), where I was chief economist for six years before coming to the World Bank. The EBRD has played a pioneering role both in analyzing these ideas and in putting them into practice.

The other broad approach, which is particularly appropriate for the World Bank, is to focus directly on improving governance. Through its programs, the Bank works to promote the institutional adjustments, anticorruption measures, and policy changes that will alter the rules and their enforcement, together with the way in which individuals and organizations behave and function, with the goal of directly improving the investment climate. As investors, whether domestic or foreign, come forward, they tend to demand more effective institutions, greater security, and constant improvements in the provision of public goods, which further enhances the quality of the investment climate. Without such changes in governance and institutions, economic decisions may continue to be dominated or blocked by those who benefit from (and helped to create) the status quo.

Moving Forward

Mark Twain reportedly said that "everyone talks about the weather but no one does anything about it." Today he might well ask, "Everyone is talking about the investment climate, but who is doing anything about it?"

The first step is to analyze the investment climate, notably through surveys. Strong analysis and identification of problems not only guide reform but also act as powerful spurs to action. Striking figures can be cited to build public support for reform and help motivate a hesitant government. For example, research carried out by the World Bank and the Confederation of Indian Industry (CII) showed a large "tax burden" effect in the Indian states with a poor investment climate in comparison with the states with a good climate (CII and World Bank 2001). This kind of survey research is crucial for ranking key problems and identifying their nature, both of which will vary greatly from one location to another. And this information is crucial for setting priorities, as governments will not be able to tackle everything at once.

Good data and analysis, then, can help to pressure, motivate, and inform governments. The next question is, what exactly do we want governments to be motivated to do? The response to this question will vary dramatically by country. But one constant is that reforming the investment climate will require leadership and powerful advocacy, primarily from governments.

Leaders promoting reforms always face a fundamental problem, one given its classic expression by Machiavelli in *The Prince* ([1513] 1940, chapter 6):

> It must be considered that there is nothing more difficult to carry out, nor more doubtful of success, nor more dangerous to handle, than to initiate a new order of things. For the reformer has enemies in all those who profit by the old order, and only lukewarm defenders in all those who would profit by the new order. . . . Thus it arises that on every opportunity for attacking the reformer, his opponents do so with the zeal of partisans, the others only defend him half-heartedly, so that between them he runs great danger.

The reformer's leadership lies in selecting and strengthening the best initiatives for reform (drawing on the type of analysis discussed above) and in taking positive action to overcome the obstacles erected by vested interests. Chandrababu Naidu, the chief minister of the Indian state of Andhra Pradesh, is fond of quoting this observation by Machiavelli. As a committed and successful reformer, he has a deep understanding of how tough reform can be when opposed by vested interests.

But people who are not especially privileged may also be among those disrupted by the reallocations that are an integral part of reform. Social protection can play an important role here, but ultimately the best social protection for the great mass of society is a growing economy. Furthermore, changing the investment climate takes time; it is not a stroke-of-the-pen reform like eliminating a tax or fee. Thus reformers not only have to argue their case strongly, they also have to prepare their constituents for the long haul.

The international financial institutions have an important role in supporting improvements in the investment climate. (I return to their role as "agents of change" when I take up the research agenda in this area.) The World Bank's focus on the structural agenda has increasingly led it to address the factors that together describe the investment climate, particularly as the Bank moves in the direction of long-term programmatic lending rather than discrete projects or short-term adjustment loans. Programs addressing macroeconomic stability and openness are levers for changing the investment climate. Programs focusing on governance and institutional development are, or should be, concerned largely with the investment climate. They can influence how difficult or easy it is to register and start firms, to move goods in and out of the country through customs, and to pay taxes in a fair, predictable, and transparent way. The quality of different types of infrastructure—rail, ports, roads, power, and telecommunications, all of which call for far-sighted, market-oriented action—is another crucial part of the story. And programs to improve training and education systems enhance the quality of the labor force, in addition to being valuable in their own right. So, in short, there is a great deal that governments and international financial institutions, equipped with serious analysis of the issues in a country, can do about the investment climate.

Investment Climate ➔ Growth ➔ Poverty Reduction

I would like to elaborate on the strong links leading from investment climate to growth and poverty reduction—not because the links are so surprising but because they are sometimes questioned and because the accumulated evidence is compelling.

It is widely agreed that the World Bank should help clients both with a structural agenda and with a social agenda. Some may be tempted to think of the structural agenda as for (hard-headed) growth and the social agenda as for (soft-hearted) poverty reduction, but this is the wrong way to view these issues. The social agenda of delivering public services, providing effective social protection, and empowering the poor to participate can contribute strongly to growth. And the structural agenda—although aimed directly at improving the investment climate—is vital for poverty reduction. We are not interested in investment or business development solely for its own sake. Instead, there is powerful evidence that the investment climate is a crucial ingredient for growth *and* for poverty reduction.

What is the evidence? First, consider the connection between the investment climate and growth. Developing countries that are improving their investment climates as part of efforts to integrate with the global economy are clearly doing well. The top tier of developing countries (about a third, including China, India, and Mexico), as measured by moves to liberalize and expand investment and foreign trade, did well in the 1990s, while the rest of the developing world as a whole did poorly. The expansion in investment and trade stemmed in part from liberalization but also from investing in infrastructure and streamlining government regulations. The factors contributing to success varied among the countries in this group, as did the style of reform, but all the countries showed real movement on some dimensions. By contrast, a number of studies have shown that Africa's low involvement in international trade has much to do with an unattractive investment climate—notably, with conflict and with poor and deteriorating infrastructure and governance. Elsewhere, among large countries that have performed less well—including, for example, Pakistan and the Russian Federation—severe problems with governance, property rights, and the rule of law have been prominent.

Second, consider the link between growth and poverty reduction. On average, there is a close relationship between growth in per capita income and growth in the incomes of the poor. Among low-income countries, Vietnam provides a clear example of this link. It made strong progress in its investment climate in 1992–97, a period during which growth accelerated. Survey evidence shows that poverty dropped sharply: of the poorest 5 percent of households in 1992, 98 percent were better off five years later.

In some cases members of poor households obtain employment in the formal sector firms that expand in a good investment environment. But benefits for the poor go well beyond this. A good investment climate is also beneficial for the informal sector, which usually employs far more people. Furthermore, formal sector investment and employment generation create new demand for informal sector goods and

services and for more farm output at better prices. In the successful reformers growth in agricultural productivity and farm income has gone hand in hand with growth in off-farm employment opportunities. China provides a striking example: the movement from communes to household operations led to dramatic increases in agricultural productivity in the early 1980s, with immediate consequences for poverty reduction. Very soon afterward, driven by the township and village enterprises, off-farm employment began to grow rapidly. Its growth has averaged more than 12 percent annually over the past 15 years.

I have emphasized small and medium-size enterprises in my arguments here, but in fast-growing economies, large and small firms tend to prosper and cluster together. The task is to create a good investment climate for both.

We are steadily learning more about all these links, not only by comparing countries and periods but also by comparing regions within countries. Much of what goes into the investment climate has to do with local institutions and policies, and in large countries these can vary substantially across regions. I briefly mentioned the CII–World Bank research in India. Because the investment climate varies greatly across states, the impact of macroeconomic reform on poverty also varies. Our recent survey of manufacturing firms ranked 10 states by their investment climate, and found a similar ranking for poverty reduction by state (CII and World Bank 2001).

Empowerment → Poverty Reduction

At the outset I outlined a strategy for development and poverty reduction that is based on two pillars: investment climate and empowerment. A favorable investment climate can generate strong growth, and where there is growth, poverty reduction is likely. But that should not be taken for granted, because not all growth is equally pro-poor. The second pillar, empowerment and investment in people, focuses on such questions as how health and education services and social inclusion can enable poor people to take part in growth and development.

Empowerment is both an inherent part of and a means for poverty reduction. Empowerment can indeed be an instrument for increasing income and for enhancing the assets, human and physical, of poor people. But it also refers to the ability of people to shape their own lives. Thus it is an integral part of the standard of living—a message that we heard very clearly through the World Bank study *Voices of the Poor* (Narayan and others 2000), which drew on surveys of more than 60,000 poor people in more than 60 countries. This perspective has also moved to center stage in the literature on the meaning of development, notably in the work of Amartya Sen. These broader notions of development and the role of empowerment were taken up in *World Development Report 2000/2001: Attacking Poverty* (World Bank 2000e).

Three types of investment are of particular importance in empowering poor people: investment in education, in social protection, and in participatory processes and social and economic inclusion.

Equipping People through Education

Basic education is critical to participation and productivity in economic life. A healthy, literate labor force will both increase the amount of growth realized from establishing a sound investment climate and strongly reinforce the poverty reduction benefit from that growth. The example of education clearly shows that the two pillars of investment climate and empowerment are closely connected and support each other.

We know, of course, that promoting education usually goes far beyond supplying resources. Just as important is organizing the delivery of education and other public services. This is an area in which communities around the world are innovating, and there are many exciting new initiatives. Often these are more decentralized than past approaches, and local control and parental involvement are emerging as prominent themes. In these cases the means of supplying education demonstrates the role of empowerment.

Consider just one example. El Salvador's Community-Managed Schools Program has been expanding education in rural areas by enlisting and financing community management teams to operate schools. These teams, made up of parents and elected by the community, are responsible for hiring and firing teachers and for equipping and maintaining schools. Their experience demonstrates that community-based incentives can encourage teachers to perform better. In particular, the program schools have lower absenteeism among both teachers and students than do traditional schools. Similar effects have been observed in India's District Primary Education Programme (also supported by the World Bank, but conceived in India), which features strong community involvement and incentives for enrolling girls.

The role of education in increasing both empowerment and economic growth is most striking when we look at the evidence on educating women and girls. Raising women's schooling levels allows them to participate more broadly and more effectively in the economy and in policymaking, administration, and government. The effects are deep and wide ranging. First and foremost are the direct implications for women's standard of living. But the broader effects are also profound, and they influence the functioning of the entire economy and society.

The evidence on these effects is extensive and powerful, as is illustrated by the examples and literature cited in the World Bank's recent report *Engendering Development* (2001a). Recent research reveals that countries with more women in parliament typically have significantly less corruption, even after the research controlled for national income and other relevant factors, such as the extent of civil liberties and the degree of trade openness. This finding suggests that women are an effective force both for good government and for business trust. But if women's participation in political decisionmaking is to increase, their educational and literacy levels must be raised. In brief, more girls need to go to school and to stay in school longer.

Women's empowerment and education yield significant benefits virtually everywhere. There is a powerful relationship between female education and overall

health—for example, mothers' education improves children's nutritional status, life expectancy, and general welfare. As data from longitudinal studies accumulate, we see that many of these effects last into later life; mothers' education is an important factor in the age-specific mortality rates of their children well into adulthood.

Greater investments in women's education, then, yield a healthier, more literate, more productive, and better-governed society in the long run. However one looks at the issue of development and poverty reduction, and certainly from the perspective of the investment climate and empowerment, the education of women and girls is an investment with outstanding returns.

Protecting People from Shocks

Social protection can be seen as a dimension of empowerment, because it enables people to adjust to new conditions and to bounce back from economic shocks. During the adjustment after an opening to foreign trade, some formerly protected activities will cease to be viable, and some workers and firm owners will lose their employment and their incomes, even as new productive activities open up. Social protection measures can smooth this adjustment. Such measures have to be tailored to country circumstances. While unemployment insurance can be important for formal sector workers, other approaches, such as public work schemes providing cash or food for work, are much more likely to reach the very poor.

But we should see social protection as much more than a short-run palliative. It is an essential underpinning of a market economy, one that helps it to function well and to involve poor people in the opportunities it creates. Without good social protection, poor people may be unable to take some of the risks that are part of participating in a market economy—even when they stand to gain strongly and to build their assets in the medium term. From this perspective, social protection is indeed a crucial element of empowerment.

Including People in Social Organizations

The third dimension of empowerment is participation and inclusion in social organizations—from self-help credit groups to water users associations, from health services to the governance of schools. Examples of the benefits of participation stretch across sectors and countries, and the powerful effects of community participation on public service delivery are increasingly recognized. I have already given examples from education, but the effects are seen everywhere. For example, in some countries where enforcement of environmental pollution regulation is weak, governments have provided local communities with reliable pollution data (see World Bank 2000b). Poor people living in the vicinity of industrial polluters have then been able to negotiate better arrangements for compensation and cleanup.

* * * * *

It is these dimensions of empowerment—equipping, protecting, and including people—that strengthen the connection between growth and poverty reduction and turn growth into pro-poor growth.

We can now see the deeper connection between the two pillars of investment climate and empowerment. To use still more metaphors, investment climate as a lens brings into focus the reforms that enhance people's sense of opportunity; if they sow, they will then be able to cultivate and reap. But these reforms also empower farmers, workers, and local entrepreneurs and managers to build assets and to take control of their own lives.

Empowerment should be a self-reinforcing process, as changes in the investment climate are. But we now recognize that in some countries many actions seen as reforms have had an adverse effect. In a number of cases they took away the old way of doing things without providing any replacement, leaving people disempowered and demoralized. The people did not participate in reforms; instead, the reform process was captured by elites, and most people were left to glean what they could from the leftovers. In those countries the investment climate has tended to spiral downward.

A Research Program in Support of the Strategy

I hope that the pillars of investment climate and empowerment provide a plausible—indeed, convincing—basis for a strategy for pro-poor growth. But they are not yet standard, and they constitute a fairly new way of looking at the challenge of development. This means that they also suggest a program of research that needs to be done if we are to better understand each of the elements and how they interact.

Analyzing and Measuring the Investment Climate

The hard-headed economists among you will be saying, "All these notions of investment climate and empowerment are splendid, but where are the analytics and data?"

Can we be analytical in assessing or measuring the investment climate? We can, and researchers have already gone a long way toward doing so. The World Bank is helping its clients develop the skills to evaluate the quality of the investment climate in their own countries through systematic surveys of private firms, with a particular focus on small and medium-size enterprises. This research identifies the key problems in the investment climate. It has already demonstrated dramatically that a hostile investment climate hits small and medium-size enterprises the hardest.

An important example of this work is the joint EBRD–World Bank Business Environment and Enterprise Performance Survey, which polled nearly 4,000 firms in 22 transition economies (Hellman, Jones, and Kaufmann 2000). The survey broke new ground in quantifying the adverse impact of different forms of corruption on the performance of firms. It also systematically demonstrated—beyond the flashy media headlines—the pernicious influence of powerful vested interests, including the so-called oligarchs, in distorting the investment climate in transition

economies and undermining growth. Such surveys are a valuable tool, not for lecturing other countries about the virtues of Western systems (which have their own problems), but for giving domestic firms a way of letting their governments know about the day-to-day obstacles they face in running a business.

Let me also share with you the results of the survey of 1,000 manufacturing and software firms in India that I mentioned. Through the World Bank's FACS (Firm Analysis and Competitiveness Surveys) program, we worked with the CII to investigate differences in the investment climate in 10 Indian states. We found that the costs to businesses of a poor investment climate—for example, an unreliable power supply, onerous regulations, and intrusive and disruptive visits from government officials—are high. For firms in poor-climate states, such as Uttar Pradesh (which has a population of 166 million, larger than that of the Russian Federation, Pakistan, Bangladesh, or Nigeria), those costs are analogous to an additional tax burden of up to 30 percent, conservatively estimated, compared with costs in Maharashtra State.

On examination, it is not difficult to see how one could arrive at such a large cost estimate. A functioning power grid is a key requirement for a vibrant small business sector. In India, because of the poor quality of grid-based electricity, most small and medium-size enterprises that we surveyed have their own power generators. In Uttar Pradesh 98 percent of firms surveyed had their own generators (they had little alternative), while in Maharashtra the share was much smaller, 44 percent. For the economic theorist, it is striking that firms surveyed in Uttar Pradesh—one of the poorest states in India, with very low wage rates—showed higher capital-output ratios on average than those in some of the richest states. The intrusiveness of government regulation also varied sharply across states; for example, firms in states with poor investment climates were visited twice as often by government officials as were firms in states with good climates.

Thus the variation in investment climate from states like Maharashtra and Karnataka, at the high end, to Uttar Pradesh and West Bengal, at the low end, is real and measurable. Not surprisingly, the states with poor investment climates enjoy less growth and have seen less poverty reduction than those with good climates. Variations in the investment climate are a key reason that some states remain poor and struggling while others are beginning to win the fight against poverty.

This type of firm survey has several advantages: it is systematic, it covers a broad section of the economy, it is amenable to fairly structured analysis, and it can yield reasonably robust general conclusions. Another source of information is the direct experience of international financial institutions as participant-investors in the private sector. The International Finance Corporation (IFC) and the EBRD are examples of public institutions that work as private-oriented investment banks. One of the great benefits of their activities is that people actually involved in financing and making investments are also thinking about how to improve the environment in which the institution is functioning—and their hands-on analysis is from the perspective of public policy, not the narrow self-interest of business. This basis for economic policy is invaluable.

A major item on the research agenda, then, must be to expand the effort to understand and, to the extent possible, quantify the different elements of the investment climate. Surveys of firms should stand next to surveys of households as a centerpiece of microeconomic research and statistics in developing countries.

Focusing on Other Areas of Research on the Investment Climate

The investment climate lens brings into focus other areas where research is needed, on both the theoretical and the empirical front. In a standard neoclassical growth framework, investment in physical capital is subject to diminishing returns. Yet the examples of successful reformers, from China to Chile, suggest that the returns to reforms in the investment climate can help power robust growth for years and even decades. Understanding what is going on in such cases is a challenge for research on the investment climate.

One strong possibility is that when countries launch credible reforms in the investment climate, they in effect propel themselves into a virtuous circle characterized by increasing returns. A better investment climate increases the number of profitable investments. These success stories serve as evidence of the government's reform-mindedness, and they motivate other entrepreneurs to take a chance and commit themselves to investing. That activity in turn builds constituencies for change and deepens the society's commitment to reform, and the society accordingly adopts additional reforms that further improve the investment climate.

A second, and more direct, effect of reforms in the investment climate is an increase in the (quality-adjusted) supply of other inputs into production, such as skilled and unskilled labor, entrepreneurship, and technology.

A third effect is the learning externality story of productivity enhancement that has become familiar from modern theories of growth. (Of course, the theories of learning-by-doing and learning-by-watching have been with us for a long time and are central to Hirschman's argument.)

Combining these effects, we see that the observed productivity of capital will not necessarily fall over time as its supply increases, and the country will thus seem to have overcome the problem of diminishing returns. Each effect would be welcome in a country that has stagnated in the past, and each probably plays its part in most actual cases. It would be useful to disentangle the stories theoretically and conceptually and explore their implications for policy and measurement.

On the empirical side, an initial part of the agenda for research on the investment climate is to identify more clearly the effects of reforms in this area. In doing so, it is useful to consider more broadly what metrics we will use in assessing the success of reforms. Time series of firm survey data will tell us whether investors are becoming less hassled and more optimistic as policies change and will even allow us to quantify how reforms are affecting obstacles and production. They will not, however, allow us to gauge the full effects of improvements in the investment climate on the larger economy. For example, a regulatory framework that prevented new firms from entering an industry might increase the satisfaction of the owners and man-

agers of existing firms—even as it reduced the dynamism of the economy as a whole. But there is, of course, a broader constituency of potential investors and of consumers, now and in the future. We need other metrics to measure the effects of investment climate on this broader group and on the wider economy.

An obvious measure is the medium-term growth of the economy and the increase in productivity of all the factors. This GDP metric was used in the research showing that the top tier of developing country reformers had grown much faster than other developing countries in the 1990s and faster even than the industrial countries. Growth in total factor productivity, though hard to calculate with precision (and a notion that has its conceptual problems), can also give us some insight into how the investment climate is changing over longer periods, or at least point to its effects. Indeed, it helps to differentiate the notion of the investment *climate* from the measurement of quantities of investment only. Growth in total factor productivity will not capture the induced increases in capital and labor supply, but it will indicate the extent to which the investment climate supports or hinders the process of combining factors into final products.

A second measure, though a crude one, is the incremental capital-output ratio (ICOR), which focuses more directly on the effects of investment itself. An earlier generation of development economists tended to view the ICOR as a well-specified function; although it increased with capital accumulation, it was assumed to remain constant for any given level of the capital-output ratio (at least with Harrod-neutral technical progress). But we know that in fact there are large variations in the ICOR, depending on the degree to which the investment climate is conducive to productivity and growth. For example, in recent years Poland, with a relatively modest investment ratio, has grown fairly rapidly, while the Russian Federation has grown much more slowly despite higher investment ratios. (Recall that the growth rate of output is the investment ratio divided by the ICOR.) Put simply, the climate in the Russian Federation was not conducive to productivity either of capital or of other factors.

So these measures, when combined with the results of firm surveys, can suggest whether the investment climate is making a difference. While we should recognize that this analysis is one of those that "labels the residual" from growth analysis, at the same time it does try to link the residual to the direct measurement of underlying phenomena that we have good reason to think are genuine drivers of growth.

Now let us turn to research on each of the broad components of the investment climate: macroeconomic stability and openness, good governance and strong institutions, and quality of infrastructure. In each case a great deal of "drilling down" into these concepts and ideas is required.

MACROECONOMIC STABILITY AND OPENNESS. We can be reasonably confident that improved macroeconomic stability and greater openness will generate more growth. But in view of the crisis in East Asia in 1997–98 and its reverberations in the Russian Federation and Latin America, not to mention more recent trouble spots, we have much to learn from further research on the best policy mixes and time paths for

domestic macroeconomic policies, exchange rate systems, and capital account management. Furthermore, successful countries have found their own way, with their own time paths, in generating growth from more open trade and more stable macro-economic policies. The adjustment process is not always easy, and it has to be adapted to local conditions and constructed and led by the country itself.

All this relates to the larger issue of making globalization more pro-poor. Many fear that the forces of globalization will sweep through countries, carrying new riches to the already well off while leaving poor people in stagnant backwaters. Our job is to find and foster pro-poor growth policies that empower poor people to lift their own sails and catch this new wind. The World Bank's Development Economics group is now preparing a policy research report on globalization and how it can be harnessed to propel people out of poverty (World Bank 2001b).

GOOD GOVERNANCE AND STRONG INSTITUTIONS. One of the newest and most active areas of research deals with the role of good governance and strong institutions and how to build them. In recent years the Bank has moved from barely being able to mention corruption to undertaking a major effort in research and programmatic lending on this issue. Surely, little or no research is needed to learn that corruption is a major impediment to both domestic and foreign investment. In many countries the age-old problem is the herculean task of cleaning out the public stables. In this, surely research can help. Tolstoy tells us that every unhappy family is unhappy in its own way. Each country with a corruption problem, too, seems to have its own story to tell. Research is needed to understand how corruption works, who benefits the most, and who suffers the most. This knowledge will allow governments to devise more effective anticorruption policies.

Closely related to the problems of corruption and governance is the issue of institutional development. Recent experience in the transition economies has taught us hard lessons about the importance of institutions. Fair, competitive, and stable markets will not arise spontaneously out of the confusion of the transition; they require supporting institutions to administer and enforce property rights, legal codes, economic regulations, and taxation of individuals and companies. Those who are benefiting from—or have benefited from—the existing postsocialist system will not necessarily push to complete the transition to a full market economy. Laws can be changed overnight, but the bigger challenge is to establish reasonable and responsible behavior that is consistent with a well-functioning market economy. Changing behavioral habits takes understanding, positive examples, leadership, and time.

Where the organic processes of building institutions have been torn apart by civil conflict, the ground is poisoned for future investment. The Bank is undertaking a new research program, led by Paul Collier, on the economic roots of civil conflict (including lootable natural resources and land pressures), in accordance with the old wisdom that an ounce of prevention is worth a pound of cure.

No research in Washington or elsewhere can substitute for public resolve to fight corruption, or invent a way to "install" new institutions. Yet much can be learned from the experience of the industrial countries, and perhaps even more from suc-

cessful reforms in developing countries. The Bank's early experience was summarized in Arturo Israel's 1987 book, *Institutional Development*. More recently *World Development Report 1997* (World Bank 1997), on the role of the state, set forth the initial results of research on corruption and on matching the complexity of proposed institutional reforms to the implementation capacity of the state. Several research studies published before the annual meetings of the World Bank and International Monetary Fund in Prague in 2000 summarized the Bank's progress in anticorruption and public governance programs since *World Development Report 1997,* focusing on the particularly vexing problem of corruption in the transition economies. These include *Anticorruption in Transition* (World Bank 2000a), *Helping Countries Combat Corruption* (World Bank 2000c), and *Reforming Public Institutions and Strengthening Governance* (World Bank 2000d).

During the entire post–World War II period the World Bank and other international financial institutions have been accumulating experience, successful and otherwise, in building institutions. Gathering, analyzing, and distilling policy recommendations from this experience now constitute an important research effort in the Bank. Indeed, *World Development Report 2002* (World Bank 2001c) focuses on the institutional foundations of a market economy.

QUALITY OF INFRASTRUCTURE. Investors, both foreign and domestic, look to the quality of the power, transport, communications, and even financial infrastructure as an important complement to investment. Regulatory economics is now enjoying something of a renaissance as an outgrowth of the new information economics of moral hazard and adverse selection—pioneered by, among others, my predecessor as chief economist, Joseph Stiglitz—and also because of the new focus on increasing-returns phenomena such as network externalities. As Joe always explained, the objective should not necessarily be deregulation but sound regulation. Despite this new birth of theory, political economy problems still have to be solved in practice, as we were rudely reminded by the recent blackouts in California. Thus the Bank's regulatory research program has had to walk on two legs. Bank researchers have mined the new regulatory economics for relevant theoretical and empirical insights, but they have also helped to set up think tanks on practical regulatory issues in Africa, Latin America, South Asia, and China. Perhaps these new think tanks can be persuaded to give a seminar in California.

Understanding Empowerment

The concept of empowerment also carries with it a rich research agenda. We have seen an extraordinary amount of experimentation with programs to empower poor people in different countries and at different levels. Think of the overwhelming variety of approaches to educating children and adults, to reducing and mitigating risk, and to building social cohesion and inclusion. Our goal as researchers should be to understand how people have succeeded in empowering themselves—it is not for us to instruct people on their own empowerment. Rigorous evaluation has a large role

to play here because it can help us cut through the thicket of promising approaches and disparate examples to discover which work best and under what circumstances. In an important sense this research parallels the work on investment climate. In both cases our goal is to understand how governments can help the individual—or family, or firm—acquire the tools needed for success and how they can reduce the impediments to the fulfillment of the individual's (or family's, or firm's) potential.

More specifically, the research agenda for this second pillar can be derived from three important factors that affect empowerment and that I examined earlier: education and the delivery of other public services, social protection, and social inclusion and participation.

It should come as no surprise that *education* has been a strong area of research in the Development Economics Group, the Human Development Network, the World Bank Institute, and other parts of the Bank. For example, the Bank has taken the lead in evaluating the effects of some recent innovations in education, including the programs that have increased community involvement in managing schools in Central America and elsewhere. But a great deal remains to be done toward understanding the effectiveness of different approaches to schooling. And the problem is not limited to the education sector; many of the difficulties with education in developing countries are similar to those that impede the delivery of other public services, such as health care. Indeed, the provision of basic services has such an important impact on poor people that it is likely to be the topic of a future *World Development Report,* which will try to structure, summarize, and disseminate the research findings. Where we find successful practices, we will analyze and publicize them so that developing countries can learn from one another's experiences.

A crucial and central topic in current education research is girls' and women's education. As I noted, educating girls and women has effects far beyond the expected gains in labor force productivity; for example, it improves the health outcomes of families, and it directly attacks the illiteracy and ignorance that constitute a key dimension of poverty. Some of the evidence has been brought together in *Engendering Development* (World Bank 2001a), which demonstrates the crucial role of women's and girls' education across the economy. But much of the evidence, although wide ranging, is patchy, new, and tentative. *Engendering Development* is as much a research agenda as a research report.

Social protection is sometimes seen, mistakenly, as a charity program for alleviating the symptoms of poverty. We need a more active perception of social protection as a springboard to help poor people bounce back from economic shocks. The dynamism of a market economy is based on risk taking. Social protection is not only a safety net for those who drop out of the market; it should function as a part of a broader market economy that allows market participants who are less well off to take more entrepreneurial risks and to make the adjustments necessary for advancement. This perception of social protection is the topic of active research in the Bank. But societies develop their own approaches to social protection, and here, as in any question of institutional change, we should be careful not to urge people to abandon the old ways when the path to the new ways is blocked or obscured.

Participation and social inclusion are among the principles of the Comprehensive Development Framework (CDF), which have guided the Bank in reappraising and recasting its relationship with client countries. The basic CDF theme of "the country in the driver's seat" entails participation not simply by the government but also by the social organizations that reach down to the roots of society. Without that participation and inclusion, the new policies that might emerge from the CDF and the Poverty Reduction Strategy Papers would lack the broader ownership necessary for implementation. At the same time, country ownership does not imply automatic support for whatever is proposed. If we are serious about poverty reduction, we must concentrate our investments on the countries and programs that are generating the changes that can overcome poverty.

How the International Financial Institutions Can Act as Agents of Change

This leads directly to the next key area of research: how can the international financial institutions most effectively help countries implement the two-pillar strategy? Our answer to this question is different from the one that would have been given 50, or perhaps even 10, years ago. Much has changed since the Bretton Woods conference of 1944. The world economy has moved decisively toward greater integration, dramatically increasing the flow of private capital to developing countries. At the same time, the international financial institutions' understanding of development has broadened and deepened. We not only look beyond aggregate income to its distribution but also recognize that standards of living and thus development have key dimensions beyond income. And along with this broader perspective, the international financial institutions have also developed an appreciation of the crucial role of reforming policies and institutions.

The strategy for development outlined here can help clarify the role of the international financial institutions in this new environment: they must serve as *agents of change*. We recognize that the private sector is the driver of development. The international financial institutions must play a complementary and catalytic role by working to create new opportunities—that is, helping developing countries extend the frontier of what is possible. This should be the constant test that they apply to their activities, and it has at least three implications. First, development assistance should have the effect of crowding in private investment—for example, by building a sound regulatory environment or upgrading the skills of the labor force—rather than substituting for such investment. Second, development assistance must be designed to help build public resources and raise their productivity rather than merely replace such resources. Third, international financial institutions should support projects that have powerful demonstration effects. These can be private sector projects (funded, for example, by the IFC or the EBRD) that are at the cutting edge of what the private sector can do or is willing to do and that thus demonstrate new opportunities. Or they can be public sector projects that can be replicated by other institutions and by other provinces or countries. (Indeed, the World Bank works both to find such

projects and to encourage the learning process.) A fundamental principle should run through everything we do: *the international financial institutions should help finance the costs of change—and should not cover the costs of not changing.*

The international financial institutions should orient their work toward helping countries improve their investment climate while at the same time supporting country ownership of and commitment to reform initiatives. How do we do it? Serious research on this topic begins by recognizing that there is a conundrum in the idea of being an agent for change, or "helping people help themselves," to quote the World Bank's mission statement. This conundrum runs through all forms of assistance. Being at once an agent of change, an external catalyst of empowerment, and a helper who is actually helpful is a subtle matter. If the helping hand is too heavy, the agent for change can spoil, distort, or suppress initiatives for and commitment to change.

Those offering development advice to a country should not employ, implicitly or explicitly, a monolithic model of the country and its polity; there are always contending groups with differing views, agendas, and interests. International development agencies should, as much as possible, resist the temptation to take sides in domestic politics. They should offer a range of ideas and good examples, lay out the evidence as scientifically as possible, and outline the rationales and the costs of the options. Intellectually self-confident decisionmakers may ask for our opinion, and we have an obligation to give it, professionally and clearly. A country can be "in the driver's seat" and still look for advice from those who are experienced, impartial, and committed to poverty reduction. Local politics may prevent leaders from acknowledging the help, but it is nevertheless generally welcome.

As private capital flows have increased and economies have grown, the international financial institutions have focused their resources still more on activities that push the boundaries on both the investment climate and governance and on projects with powerful demonstration effects. In each case, precisely because expanding the frontier is of the essence, careful analysis and preparation are vital. Hence the ever-increasing importance of the Bank's analytical work, both in supporting broad diagnostic analysis of countries and sectors and in preparing projects. The growing importance of the "frontier" dimension of our work implies ever-stronger emphasis on the Bank as an ideas or knowledge institution, and thus on research.

The political economy of assistance to reform is a complex art. The agent for change gains skill by learning from others through case studies, "war stories," and apprenticeship, as well as through reflection. Rather than trying to discover or impose a fixed sequence or time path, the agent for change must be open to supporting travel on many possible roads. Research on this topic should not search for a blueprint but should seek a "description" of the local landscape, so that dead-end paths and pitfalls can be avoided and so that we can give encouragement in the general direction of the more promising paths to reform.

There will be times when, in pursuit of the goal of poverty reduction, an international financial institution has to take a chance and support a reform program that is serious but is under threat. Indeed, it is under these circumstances that conditionality can be most effective. A reformer who is trying to limit the effectiveness of

the opposition to reforms may seek conditionality as a way of tying a country more closely into the reforms. Such support has to be supplied judiciously; we have to be aware that it can backfire and that it may be ineffective where divisions are too deep. We have to be realistic and assess whether the program has a chance of success. Conditionality is no substitute for real commitment to reform. But since we hear often from reforming ministers who quietly ask for conditionality on key issues, I have no doubt that conditionality can be a substantial help at crucial times in taking reforms forward.

The international financial institutions can help to strengthen and broaden reform coalitions. Reform A might be modified to yield a reform A' that would be more complementary and conducive to reform B, and the beneficiaries of reform B might therefore be willing to join the coalition. Here conditionalities backed by resources can play a similar role in helping to build a reform coalition. If reform A unlocks resources that would allow reform B (or take away the threat to an existing program B), it may be possible to broaden the reform coalition. The resources supplied by international financial institutions might also cover the costs of adjustments required by the reform, thus bringing on board some who might otherwise be in opposition. Eventually the reform coalition might be broad enough to reach the tipping point, so that even indifferent bystanders will want to join the winning team. By anticipating problems, by being creative in suggesting possible solutions, and by adding its integrity in support, an international financial institution can help to foster a "deal for reform" that might not otherwise materialize. If its integrity is to be maintained and to have a real effect over time, the institution should focus its support on programs that have a real chance of success and are likely to benefit poor people.

The effectiveness of an international financial institution as an adviser and agent of change is based on open, respectful, and long-term relationships with countries. If the institution asks a country to be open to new learning and experimentation, it should ask the same of itself. There is no expert quick fix for the problems of social change, no royal road to learning what works in a country, and no instantaneous solution or cheap shortcut to institutional development. To be an effective agent of change, the institution must have a relationship with the country that is a genuine partnership and that is based on an understanding of local conditions. Because such an understanding requires local presence, the research to build it should eventually be generated as much as possible by the knowledge institutions of the country itself.

The concept of partnership applies to middle-income countries as well as to low-income countries. If we take seriously the common humanity that binds us all, we must recognize that significant absolute poverty in a middle-income country such as Brazil should not be seen solely as Brazil's problem. Moreover, as a country grows into the ranks of the middle-income countries, the partnership should change and mature. The access to capital markets that comes with rising income opens new opportunities. But while middle-income countries and other emerging markets do gain access to international private capital flows, that access is intermittent, often at arm's length, and sometimes very costly. By contrast, what the international finan-

cial institutions can provide is a steady, responsive partnership and a steady, flexible flow of capital in support of needed reforms. The fight for poverty reduction is a long haul, and the partnership is especially important when the situation is at its most difficult.

Let me close this discussion on international financial institutions as agents of change with a word or two to the theorists. Being an effective agent of change requires some understanding of the dynamic process being influenced. There are several interesting theoretical questions. First, we have to have some notion of how to convince and encourage. That can be done in part through information and incentives, for which we have a fairly standard economic theory. But some of the convincing—indeed, in development, much of it—has to do with changing preferences. Second, we have to think about how to act as an external player in a repeated game if we are to understand how to influence outcomes, especially when the domestic polity is divided. We have to recognize, too, that the external role inevitably becomes internal to the game. Third, it would be good to know whether we should be using a model with high and low equilibria (where a fairly crude shove in roughly the right direction might achieve the desired result) or a more complex dynamic model, either equilibrium or disequilibrium. When we try to model explicitly, there is always a danger that some misguided or simplistic clot will take us too literally. But that is a risk that theorists always have to run, and one that is well worth running if we can gain greater insight into these issues.

Conclusion

I have laid out a strategy for development, based on the twin pillars of investment climate and empowerment. What are the prospects that developing countries will implement this strategy and that the industrial world will provide real support for it? My answer is more positive than it might have been even a few years ago. One cause for hope is the significant improvement in the quality of macroeconomic and trade policies in many developing countries. I have mentioned some high-profile reformers, but the phenomenon is more general. In the developing world as a whole, many of the policies that make a difference in growth and poverty reduction—such as controlling inflation and lowering trade barriers—are better than ever before. For this reason, notwithstanding the recent slowdown in the United States and the continued economic troubles in Japan, the medium-term outlook for the developing countries is stronger than it has been in many years.

Reforms in the developing world thus create opportunities. But are the rich countries ready to do their part to help developing countries seize those opportunities? Again, recent trends have strengthened the grounds for optimism.

First, consider trade policy. The returns to improving the investment climate depend in significant part on whether reforming countries have access to foreign markets for their goods. Rich country barriers in key sectors and industries, such as agriculture and textiles, discourage reforms in the investment climate. But recent movements to increase market access for developing countries—such as the

European Union's Everything but Arms initiative, which eliminates all barriers on imports from the very poorest countries—are cause for hope that the rich countries have recognized their responsibility.

Second, the rich countries have increasingly shown that they recognize the need to reduce the administrative burden of aid on developing countries. One of the best ways to move in this direction would be for all countries to follow the lead of the United Kingdom by dropping the practice of tying aid. Another is to improve donor coordination.

Third, the most recent data show a small increase in aid volumes from a number of countries (although, in aggregate, aid as a share of donor country GNP is not increasing and remains far below its level of a decade ago).

Finally, both industrial and developing countries have declared their commitment to the international development goals for progress against poverty over the next decade and a half.

Together, these recent trends signal significant movement toward support for more rapid poverty reduction.

Thus those making political decisions and formulating economic policies have generated real opportunities for the international financial institutions to help accelerate development and poverty reduction. Research has a vital role to play in helping to take advantage of those opportunities. We have learned much about what works and what does not in economic development. We have learned much, too, in recent years about how to catalyze and support reforms most effectively. We have already begun to build this knowledge into our work. But there remains much that we still have to learn about the investment climate and empowerment and about how to foster change.

What I have described is, in many ways, only a structuring of the agenda. Our work should build on the understanding of the two-pillar approach that itself has emerged from the research of the past few years. Of course, we have to act on policy in real time; we cannot afford Hamlet's agonized approach to decisionmaking. But today's research underpins tomorrow's crucial decisions. The research agenda I have described not only provides deep intellectual challenge and fascination, it also can yield powerful returns in our common fight against poverty.

References

Goswami, Omkar, David Dollar, A. K.. Arun, C. Srivatsava, Vishal More, Arindam Mookherjee, Taye Mengistae, Mary Hallward-Driemier, Guiseppe Iarossi. 2002. "Competitiveness of Indian Manufacturing: Results from a Firm-Level Survey." Confederation of Indian Industry (CII), New Delhi; and World Bank, Washington, D.C.

Hellman, Joel, Geraint Jones, and Daniel Kaufmann. 2000. "'Seize the State, Seize the Day': State Capture, Corruption, and Influence in Transition." Policy Research Working Paper 2444. World Bank Institute, Governance, Regulation, and Finance Division; World Bank, Europe and Central Asia Region, Public Sector Group; and European Bank of Reconstruction and Development, Office of the Chief Economist, Washington, D.C.

Hirschman, Albert O. 1958. *The Strategy of Economic Development*. New Haven, Conn.: Yale University Press.

Israel, Arturo. 1987. *Institutional Development: Incentives to Performance*. Baltimore, Md.: Johns Hopkins University Press.

Lanjouw, Peter, and A. Sharif Shariff. Forthcoming. "Rural Non-Farm Employment in India: Access, Income, and Poverty Impact." Working paper. National Council of Applied Economic Research, New Delhi.

Lanjouw, Peter, and Nicholas Stern, eds. 1998. *Economic Development in Palanpur over Five Decades.* New York: Oxford University Press.

Machiavelli, Niccolò. [1513] 1940. *The Prince and the Discourses*. New York: Random House.

Narayan, Deepa, with Raj Patel, Kai Schafft, Anne Rademacher, and Sarah Koch-Schulte. 2000. *Voices of the Poor: Can Anyone Hear Us?* New York: Oxford University Press.

Schumpeter, J. A. [1912] 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle.* Cambridge, Mass.: Harvard University Press. (Originally published in German in 1912 as *Theorie der wirtschaftlichen Entwicklung.*)

Sen, Amartya. 1999. *Development as Freedom*. Oxford: Oxford University Press.

World Bank. 1997. *World Development Report 1997: The State in a Changing World*. New York: Oxford University Press.

———. 2000a. *Anticorruption in Transition: A Contribution to the Policy Debate.* Washington, D.C.

———. 2000b. *Greening Industry: New Roles for Communities, Markets, and Governments*. World Bank Policy Research Report. New York: Oxford University Press.

———. 2000c. *Helping Countries Combat Corruption: Progress at the World Bank since 1997.* Washington, D.C.

———. 2000d. *Reforming Public Institutions and Strengthening Governance*. Washington, D.C.

———. 2000e. *World Development Report 2000/2001: Attacking Poverty*. New York: Oxford University Press.

———. 2001a. *Engendering Development: Through Gender Equality in Rights, Resources, and Voice*. World Bank Policy Research Report. New York: Oxford University Press.

———. 2001b. *Globalization, Growth, and Poverty: Building an Inclusive World Economy.* World Bank Policy Research Report. New York: Oxford University Press.

———. 2001c. *World Development Report 2002: Building Institutions for Markets*. New York: Oxford University Press.

This Page Intentionally Left Blank

Globalization and Inequality

This Page Intentionally Left Blank

Globalization and Inequality: Historical Trends

Kevin H. O'Rourke

This article surveys trends in international economic integration and inequality over the past 150 years, as well as the links between them. In doing so, it distinguishes among the different dimensions of globalization (trade, migration, and capital flows) and between inequality between countries and that within countries. International economic integration has largely followed a U-shaped path during this period, with late-19th-century integration being followed by interwar disintegration and post-1945 integration. The post-1945 recovery has been uneven, however, with some dimensions of integration remaining undeveloped relative to 1913 and with some countries participating less fully in the recovery than others. Total world inequality has been on the rise since 1820, driven entirely by a rise in between-country inequality. However, recent evidence suggests that there may have been a historic turning point around 1980, with between-country divergence being replaced by convergence.

Broadly speaking, in the late 19th century both trade and migration (but not capital flows) made the rich New World more unequal and the (less rich) Old World more equal. The evidence on the links between within-country inequality and globalization in the late 20th century is mixed. The balance of evidence suggests that globalization was a force for between-country convergence in both the late 19th and the late 20th centuries, with long-run patterns of divergence due to other factors.

Does globalization lead to the world becoming a more equal place, or does it lead to the rich getting richer and the poor getting poorer? This question has assumed ever-greater importance with the emergence of the World Trade Organization as a force for trade liberalization throughout the world, with the

increased economic integration of Europe, with the collapse of Communism and the opening of previously autarkic economies, and with the renewed speculation about the formation of a pan-American free trade area.

The question is increasingly being raised by opponents of globalization, but public debate on the issue can be frustratingly confused. Protestors are often vague about what globalization is, and fail to recognize that globalization has different dimensions, which may have different effects on inequality. Most seriously, they often define globalization as encompassing many different phenomena, some of which have little or nothing to do with globalization as economists define it (Rodrik 2000). Globalization as economists define it encompasses declining barriers to trade, migration, capital flows, technology transfers, and foreign direct investment. This article explores the historical links between trade, migration, and capital flows, on the one hand, and inequality, on the other.

The article distinguishes between two separate dimensions of inequality: between-country inequality and within-country inequality.[1] If we take the appropriate unit of observation to be the individual citizen anywhere in the world, total world inequality will clearly depend on both between- and within-country inequality, with globalization potentially affecting both through quite different channels. What these channels might be is the subject of a theoretical overview in the following section.

The article then traces the evolution of globalization during the 19th and 20th centuries, distinguishing between the different dimensions involved. Next, it briefly documents trends in both types of inequality over the past two centuries. Finally, the article explores in greater detail the experiences of inequality during the two most dramatic episodes of globalization, in the late 19th and late 20th centuries, and documents what is known about the links between globalization and inequality during these two periods. It concludes with suggestions for further research.

Globalization and Inequality: Theoretical Connections

What does theory have to say about the links between globalization and inequality? It turns out that globalization should affect within- and between-country inequality through quite different channels, so these are discussed separately below.

Globalization and Within-Country Inequality

Our intuitions about the links between international economic integration and income distribution arise for the most part from the static neoclassical trade theory developed by Eli Heckscher and Bertil Ohlin in the early years of the 20th century. Their basic insight was that trade patterns reflect differences in the distribution of endowments across countries and that countries export goods embodying the factors of production with which they are well endowed. Commodity market integration therefore leads to an increase in the demand for abundant (and cheap) factors of production, thus raising their price, while trade

leads to a decline in the demand for scarce (and expensive) factors of production, thus lowering their price.

In a simple two-country, two-factor, two-good setting, with identical technology in the two countries, trade and factor mobility should have identical effects on factor prices and thus on income distribution. If the United States is abundant in skilled labor, and Mexico in unskilled labor, trade will increase U.S. skilled wages and Mexican unskilled wages, and it will lower U.S. unskilled wages and Mexican skilled wages. Thus trade leads to greater wage inequality in the United States and lower wage inequality in Mexico. It is also true, of course, that the migration of unskilled workers from Mexico to the United States, or of skilled workers from the United States to Mexico, will have identical effects on factor prices.

Things get more complicated once we move away from this simple 2x2x2 framework. Consider this example: the U.S. economy has superior technology or superior endowments of a third factor of production, so that U.S. skilled wages are higher than Mexican skilled wages, rather than lower. Trade and unskilled migration may raise wage inequality in the United States and lower it in Mexico, as before, but skilled migration will be from South to North and will have the opposite effect on wage inequality in the two countries. In this case some dimensions of globalization (trade and unskilled migration) have very different implications for inequality than others (skilled migration).

Alternatively, it is important to recognize that not all developing countries are identical: instead, they differ greatly in their endowments of skills, labor, and capital (Davis 1996). Middle-income countries such as Mexico might be abundant in skilled labor relative to countries such as China and India. Middle-income countries might therefore protect sectors abundant in unskilled labor and as a result might see skill premiums rising with liberalization. There is in fact evidence that sectors intensive in unskilled labor received the most protection in countries such as Mexico and Morocco before liberalization (Currie and Harrison 1997; Hanson and Harrison 1999).

A third possibility is that liberalization and foreign direct investment might lead to new skill-intensive activities being introduced in developing countries (Feenstra and Hanson 1996). Under such circumstances the relative demand for skilled labor would rise in the South. Finally, capital inflows to the developing countries might increase the demand for skilled labor, if skilled labor is complementary to capital, and thus raise wage inequality. Alternatively, if skilled labor and capital are complementary to some natural resource (such as minerals), liberalization in a resource-rich developing country might increase skill premiums and overall inequality (Kanbur 1999).

Thus far the discussion has assumed that the only factors influencing the impact of globalization on inequality are countries' endowments and their technology. Another crucial factor should be mentioned: the distribution of those endowments among a country's citizens. Take, for example, a positive trade shock in a land-abundant country, which raises the returns to land. Clearly, if landholdings are concentrated among a few large landowners, this shock is likely to be a force for greater inequality. But if the land belonged exclusively to poor peasant proprietors, such a

shock might well imply greater equality. The fact that trade theory implies a link between globalization and factor prices, rather than inequality, with the distribution of endowments intervening between the two, is one of the problems facing applied researchers seeking to explain the behavior of summary inequality measures such as the Gini coefficient.

In conclusion, the links between globalization and within-country income distribution are ambiguous. First, globalization affects factor prices differently in different countries, for standard Heckscher-Ohlin reasons. Second, different dimensions of globalization (such as trade and factor flows) may have different implications for factor prices in a given country. Third, a given dimension of globalization (such as capital flows) may have ambiguous effects on factor prices in a given country, depending on, among other things, patterns of complementarity or substitutability between factors of production. Finally, a given impact on factor prices can have different effects on inequality, depending on the distribution of endowments across individuals.

Ultimately, these are issues that can only be resolved empirically.

Globalization and Between-Country Inequality

Arguments based on static trade theory suggest that globalization affects factor prices in the first instance. However, these arguments also have implications for between-country inequality: other things equal, factor price convergence should bring per capita incomes closer together. Typically, however, between-country inequality is discussed in the context of dynamic growth theory rather than static trade theory. Models developed in the past decade that endogenize the long-run growth rate are capable of deriving long-run growth effects of a number of policies, including trade policy. Their conclusion is that the implications of trade liberalization for convergence are theoretically ambiguous.

Many examples could be cited, but two will suffice. Both Stokey (1991) and Grossman and Helpman (1991, chapters 6, 9) assume that North-South trade is driven by differences in relative endowments of skilled and unskilled labor. Both also assume that trade drives factor price convergence, with wage inequality rising in the North and falling in the South. In Stokey's model the growth mechanism is individuals investing in human capital: when trade leads to a decline in the returns to skill in a small developing country, this can reduce the incentive to acquire skills and thus slow growth in that country. Trade may thus lead to divergence.

Grossman and Helpman take the endowment of human capital as exogenous and assume that human capital is useful in that it is an input into research and development (R&D), which drives growth. In such a scenario trade that lowers skilled wages in a developing country, and increases them in a developed country, boosts technical progress in the developing country and slows it in the developed country. The reason is that the cost of innovation declines in the developing country (and increases in the developed country). Thus trade leads to convergence.

Once again, these issues can only be resolved empirically.

Trends in International Economic Integration

To casual commentators it seems obvious that today's globalization is unprecedented. However, the late 19th century was also a period of dramatic globalization, and the world economy was extremely well integrated in 1914, even by the standards of the late 20th century (O'Rourke and Williamson 1999). War brought all this to an end, and despite the efforts of politicians and organizations such as the League of Nations, the interwar period saw further retreats into protectionism, the erection of barriers to immigration, the final breakdown of the gold standard, and a wave of international defaults. To a large extent post-1945 integration can be seen as an attempt to recoup the losses of the interwar period. At what point were these losses finally recouped, and when did integration start to progress beyond the levels achieved in 1914?

International Integration of Commodity Markets

The 19th century saw a series of dramatic technological developments, chiefly the steamship and the railroad, that were to have a profound impact on international trade. The impact on transport costs was substantial: Harley's (1988) index of British ocean freight rates remains relatively constant between 1740 and 1840, then drops by about 70 percent between 1840 and 1910. Until the 1870s trade policy reinforced these trends. Britain liberalized from 1815 to 1846, when the country took a decisive step toward free trade. The years after 1860 saw significant tariff cutting in Europe. For example, by 1877 Germany "had virtually become a free trade country" (Bairoch 1989, p. 41). In the late 1870s, however, cheap New World and Russian grain began depressing European land values, sparking a powerful protectionist response. In the United States the Northern victory in the Civil War ensured high levels of protection for the rest of the century. By contrast, in Asia declining transport costs did not have to contend with rising tariffs: China, India, Indonesia, Japan, Korea, and Thailand all moved toward free trade, most forced to do so by colonial dominance or gunboat diplomacy.

What were the combined effects of transport cost and trade policy developments in the late 19th century? To answer this question we need to focus on international commodity price gaps, and the evidence is striking (O'Rourke and Williamson 1999, chapter 3). Trend estimates based on Harley's (1980) data show that Liverpool wheat prices exceeded Chicago prices by 58 percent in 1870, by 18 percent in 1895, and by 16 percent in 1913. Nor was this Anglo-American price convergence limited to the grain market: it can also be documented for tin, coal, wool, hides, iron, copper, bacon, coffee, and cotton textiles. On the European continent tariffs impeded international price convergence, but in Asia trade policy strengthened the impact of technological developments: the cotton price spread between London and Bombay fell from 57 percent in 1873 to 20 percent in 1913, while the jute price spread between London and Calcutta fell from 35 percent to 4 percent, and the rice price spread between London and Rangoon fell

from 93 percent to 26 percent (Collins 1996). Commodity market integration in the late 19th century was both impressive in scale and global in scope. Indeed, developing economies were becoming more rapidly integrated with the rest of the world than were their Atlantic economy counterparts during this period (Williamson 2000).

Transport costs continued to fall during the 20th century, but at a slower rate. Isserlis (1938) provides an index of British tramp freight rates from 1869 to 1936. Between 1869 and 1914 real trend freight rates fell by 34 percent. They increased sharply during the war, remaining abnormally high until 1920. While they fell up to 1925, they never attained their prewar levels, and they rose thereafter, with the trend between 1921 and 1936 being broadly flat (at a level roughly equal to the 1869 level; Findlay and O'Rourke 2001).

What of the post-1945 period? In the most careful study of the subject to date, Hummels (1999) concludes that ocean freight rates have actually increased. By contrast, air freight rates declined dramatically in the 1950s, 1960s, and 1980s, declined more slowly in the 1990s, and rose in the 1970s. The result, predictably enough, has been a more than tenfold increase in the ratio of air to ocean shipments in the years since 1962.

It thus follows that trade liberalization played a much greater role in commodity market integration in the late 20th century than in the 19th century. Data on average tariffs on manufactured products are available for a number of countries back to 1913 (table 1). These data show the interwar rise in protection and the decline in tariff barriers since 1950. They also show that for most of these countries tariffs are much lower today than in 1913. There are exceptions, however, notably the United Kingdom and certain Asian countries. Both China and India, for example, have much higher tariffs now than in 1913—an extremely important qualification, given these countries' share of world population. Tariffs are much higher now in developing than in developed countries, while the opposite was true in the late 19th century (although there have been substantial declines in Latin American tariffs since the early 1980s, and smaller declines in East Asia; Rodrik 1999).

Moreover, emphasizing industrial tariffs overstates the extent to which industrial countries today have moved toward free trade, for two reasons. First, agricultural protection is very high in many rich countries (and higher than in 1913). Coppel and Durand (1999) report that it raises prices received by farmers by about 60 percent in Japan, 40 percent in the European Union, 20 percent in the United States, and 15 percent in Canada. Second, nontariff protection (such as quotas, voluntary export restraints, and technical barriers to trade) is much more prevalent today than a hundred years ago.

What has been the combined impact of the transport cost and trade policy developments documented above? Price gaps for identical commodities in different markets remain the best measure of market integration, but little work has been done to document these for the 20th century. Moreover, obvious international sources of price data (such as the commodity price data in the World Bank's *World Development Indicators* or the International Monetary Fund's *International*

Table 1. Average Tariffs on Manufactured Goods, Selected Economies and Years, 1913–99

Economy	1913	1931	1950	1980	1998–99
Austria	18	24	18	14.6	n.a.
Belgium	9	14	11	n.a.	n.a.
Denmark	14	—	3	n.a.	n.a.
France	20	30	18	n.a.	n.a.
Germany	13	21	26	n.a.	n.a.
Italy	18	46	25	n.a.	n.a.
Netherlands	4	—	11	n.a.	n.a.
Spain	41	63	—	8.3	n.a.
Sweden	20	21	9	6.2	n.a.
United Kingdom	0	—	23	n.a.	n.a.
European Union	n.a.	n.a.	n.a.	8.3	4.1
Russia	84	n.a.[a]	n.a.[a]	n.a.[a]	13.4[b]
Switzerland	9	19	—	3.3	3.2[c]
Australia	16	—	—	—	6.0
Canada	26	—	—	—	4.9
Japan	25–30	—	—	9.9	5.5
New Zealand	15–20	—	—	—	4.4
United States	44	48	14	7	4.5
Argentina	28	—	—	—	14.0
Brazil	50–70	—	—	—	15.2
China	4–5	—	—	—	17.4
Colombia	40–60	—	—	—	11.4
India	5[d]	—	—	—	34.2
Iran, Islamic Rep. of	3–4	—	—	—	—
Mexico	40–50	—	—	—	12.6
Thailand	2–3	—	—	—	47.2[e]
Turkey	5–10	—	—	—	0.3

— Not available.
n.a. Not applicable.
a. Refers to the Soviet Union, which ran such a restrictive trade policy that average tariffs were irrelevant.
b. Refers to 1997.
c. Refers to 1996.
d. Approximation.
e. Refers to 1993.
Source: Bairoch 1989, 1993; World Bank, *World Development Indicators 2000.*

Financial Statistics) reveal no discernible general trend toward commodity price convergence during the past four decades (O'Rourke 2002).

Overall, what can we conclude about commodity market integration over the past 150 years? First, the late 19th century probably saw more dramatic progress toward commodity market integration than did the late 20th century.[2] Second, commodity markets are probably even better integrated today, but we lack the empirical evidence to document this. Clearly, we need further research on this important issue.

International Integration of Capital Markets

Capital exports from the center to the periphery were enormous in the late 19th century (O'Rourke and Williamson 1999, chapter 11). The share of British wealth overseas was 17 percent in 1870, and it increased to an impressive 33 percent in

1913. The flows were extremely large during peak years: as a share of British GDP they were 7.7 percent in 1872, 6.9 percent in 1888, and 8.7 percent in 1911. No Organisation for Economic Co-operation and Development (OECD) country, including the United States, exported capital to that extent in the late 20th century. For example, German and Japanese current account surpluses in the mid- and late 1980s peaked at around 4 or 5 percent of GDP.

Foreign capital flows were equally important at the receiving end. To give just one example, net inward foreign investment ranged from 10 to 20 percent of gross fixed capital formation among the major developing country importers in the decade before 1984, though it was less than 10 percent of investment in developing countries in the early 1990s (World Bank 2000, p. 121). The same statistic for the four decades between 1870 and 1910 was 37 percent for Canada (Jones and Obstfeld 1997), about 70 percent for Argentina,[3] and perhaps as much as 75 percent for Mexico. By some measures, international capital flows have never been as important as they were in the late 19th century, despite all the rhetoric about the unprecedented nature of today's globalization.

Capital flows diminished during the 1920s, but things would soon get worse. The onset of the Great Depression was followed by a wave of default in developing countries, where capital flows remained limited for decades thereafter. Between 1945 and 1972 most of the limited capital flows that occurred took the form of direct government and multinational institutional investment abroad. Since 1972 the global capital market has become increasingly important. But historical data for 14 countries place current trends into their proper context. While the average current account (absolute value) for these countries has been on the rise as a share of GDP since the early 1970s, it is still only half what it was in the late 1880s (figure 1).

The sample of countries biases the picture somewhat, however; it consists of 13 OECD countries plus Argentina. In the late 20th century current account deficits were much greater in some developing countries. Between 1991 and 1997 they averaged 5.1 percent of GDP in Romania, 6.0 percent in Thailand, 6.1 percent in Malaysia, 13.5 percent in Singapore, and 18.9 percent in Sudan. By contrast, the equivalent figure for Argentina between 1910 and 1913 was 11.0 percent, while it was 14.5 percent in Canada. Clearly, capital flows involving certain developing countries were very high in both periods.[4]

So much for the size of flows. What about the extent of integration? Standard measures tell a consistent story: capital markets were highly integrated in the late 19th century, disintegrated during the interwar period, and are only now recovering the levels of integration experienced in 1913. This U-shaped pattern is apparent in data on real and nominal interest rate differentials (Obstfeld and Taylor 1998, 2001). Taylor (1996), applying the Feldstein-Horioka (1980) test to historical data, finds that global capital markets were better integrated between 1870 and 1924 than they were between 1970 and 1989.

What of the composition of the capital flows? Dunning (1993) estimates that about 35 percent of the stock of international long-term debt in 1914 consisted of foreign direct investment. By contrast, foreign direct investment accounted for only

Figure 1. Capital Flows, 1870–1992
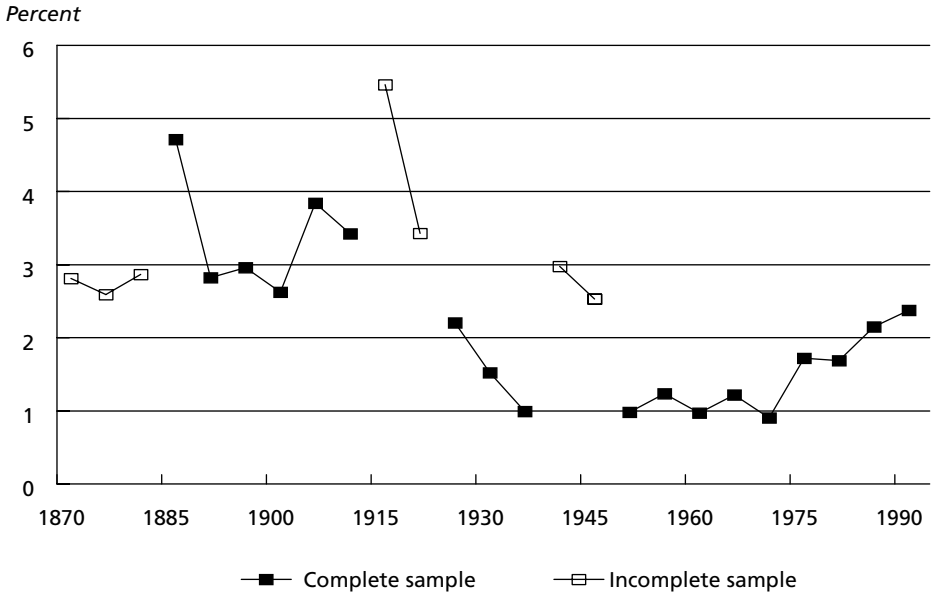
Percent

Note: Sample consists of 13 OECD countries and Argentina.
Source: O'Rourke and Williamson 1999 (figure 11.2, p. 213).

16.8 percent of private capital flows in 1973–81, but 50.3 percent of private capital flows in 1990–97 (World Bank 2000, p. 126). According to both Jones (1996, p. 32) and Bairoch and Kozul-Wright (1998, p. 46), the stock of foreign direct investment reached more than 9 percent of world GDP in 1913, a figure exceeded only in the early 1990s (the figure stood at 16 percent in 1999; UNCTAD 2000, p. xvi). Foreign direct investment is more important now, but we are talking about a quantitative, not a qualitative shift.

What proportion of these flows went to developing countries? Between 1907 and 1913 British overseas lending was split evenly between what we would now consider to be rich countries (in North America, Australasia, and Europe, including Eastern Europe) and the rest of the world (Taylor and Williamson 1994). In 1914, 62.8 percent of foreign direct investment was in developing countries, with 32.7 percent in Latin America, 20.9 percent in Asia (including 7.8 percent in China), and 6.4 percent in Africa (table 2). These figures might suggest that 19th-century capital flows were largely directed toward poor countries and were thus a force for convergence. Looks can be deceiving, however: late-19th-century capital flows were predominantly toward the resource-abundant New World and were thus a source of divergence (Clemens and Williamson 2001).[5]

Still, the late 19th century saw a greater share of foreign direct investment going to developing countries than did the 20th century. Developing countries still hosted

Table 2. Regional Distribution of Stock of Foreign Direct Investment, Selected Years, 1914–99

(percentage of total stock of foreign direct investment)

Country group or country	Foreign direct investment by origin			
	1914	*1938*	*1960*	*1999*
Developed countries	100.0	100.0	98.9	89.9
North America	19.2	30.4	52.0	27.5
Canada	1.0	2.7	3.8	3.7
United States	18.2	27.7	48.3	23.8
Western Europe	75.4	63.9	42.1	54.1
France	12.0	9.5	6.2	6.3
Germany	10.3	1.3	1.2	8.8
United Kingdom	44.6	39.8	16.3	14.0
Other Western Europe	8.6	13.3	18.3	25.0
Japan	2.1	2.8	0.8	6.2
Russia	2.1	1.7[a]	0.0[a]	0.2
Developing countries	0.0	0.0	1.1	9.8
Total	100	100	100	100

Country group or country	Foreign direct investment by host			
	1914	*1938*	*1960*	*1999*
Developed countries	37.2	34.3	67.3	67.7
North America	16.0	16.8	37.6	26.3
Canada	5.7	9.4	23.7	3.5
United States	10.3	7.4	13.9	22.8
Western Europe	7.8	7.4	22.9	36.8
United Kingdom	1.4	2.9	9.2	8.3
Other Europe	9.9	1.6	0.9	2.4
Australasia and South Africa	3.2	8.0	6.6	3.5
Japan	0.2	0.4	0.2	0.8
Russia	7.1	0.0	0.0	0.3
Developing countries	62.8	65.7	32.3	30.1
Latin America	32.7	30.8	15.6	10.2
Africa	6.4	7.4	5.5	2.0
Asia	20.9	25.0	7.5	17.7
China	7.8	5.8	0.0	6.4
India and Sri Lanka	3.2	5.6	2.0	0.4
Middle East	2.8	2.6	2.8	1.3

a. Refers to the Soviet Union.
Source: Dunning 1993; UNCTAD 2000.

almost two-thirds of investment in 1938, but their share collapsed to less than one-third by 1960. By 1999 the figure stood at 30.1 percent (see table 2). Nor is this pattern limited to direct investment: according to Obstfeld and Taylor (2001), a smaller proportion of internationally mobile capital was located in poor countries in 1997

than in 1913, illustrating "an important dimension in which the globalization of capital markets remains behind the level attained under the classical gold standard" (p. 51).

Most investment in the late 19th century was being used for social overhead investment, as is true in the developing world today. About 70 percent went into railroads, municipal sewerage, telephones, and other social overhead investments. Railroads alone accounted for about 41 percent of the total in 1913 (Feis 1930, p. 27). The debt was issued largely by governments, as was also true of the developing world during most of the postwar years. But governments' share of the borrowing was not much larger then than now: they accounted for maybe 40 percent of the investment flows in the late 19th century. By 1930 or so, 62 percent of London-based lending, and 80 percent of New York lending, was going to governments. Governments got 80 percent of the flows as late as 1980, but only 33 percent in 1997 (World Bank 2000, p. 127). In this respect, once again, the late 20th century returned to late-19th-century patterns.

The sectoral distribution of capital flows broadened, however, with far more going into industry and finance in the late 20th century than in the earlier period. Of the accumulated stock of foreign direct investment in 1914, about 55 percent was in the primary product sector, 20 percent in railroads, 15 percent in manufacturing, and 10 percent in trade, distribution, public utilities, and banking (Dunning 1993, p. 116). By contrast, in the 1990s only about 6 percent of foreign direct investment from the European Union went to the primary sector, but 31 percent went to manufacturing, and 63 percent to services (Baldwin and Martin 1999, p. 19). About half of U.S. foreign direct investment is in services, and about 35 percent in manufacturing (Bordo, Eichengreen, and Irwin 1999).

The issue is important, since foreign direct investment can serve as a vehicle for technology transfer and thus hasten international convergence, as it did in Ireland during the 1990s. The changing sectoral composition of foreign direct investment over time suggests that such investment probably played a more important role in this regard in the late 20th century than it did in the late 19th (although this is mere speculation).

The composition of portfolio flows has changed dramatically. In the late 19th century bonds accounted for an overwhelming share of such flows. During the lending boom of the 1970s, by contrast, bank lending accounted for almost two-thirds of total flows, with both bond issues and portfolio equity flows being minimal. During the 1990s the composition of portfolio flows became far more balanced, with a fairly equal division among bank lending, bond issues, and equity finance (World Bank 2000, p. 126).

This broader range of financial assets traded clearly distinguishes the late 19th and late 20th centuries, as does the greater share of investment today in manufacturing and services. Other differences, highlighted by Bordo, Eichengreen, and Kim (1998), include the huge volume of gross capital flows today. Although clear evidence on the late 19th century is lacking, it seems certain that the ratio of gross to net capital flows is much greater now than then, reflecting greater volumes of short-

run capital flows. Presumably, however, net long-run flows matter more than gross short-run flows for growth and income distribution. Finally, while much foreign direct investment in the late 19th century was undertaken by "free-standing companies," incorporated in the core to carry out business in the periphery, foreign direct investment today occurs mostly within multinational corporations that do business in both home and host countries.

International Migration

It is in migration that the late 19th century seems most clearly to have seen more globalization than today. Although barriers to immigration were being erected by the end of the period (O'Rourke and Williamson 1999, chapter 10), the late 19th century stands out as a relatively liberal interlude in terms of migration policy. Once transport costs had fallen sufficiently relative to the average wage, the inevitable consequence was a huge intercontinental flow of people. Between 1820 and 1914 roughly 60 million Europeans emigrated to the New World. European emigrants averaged 300,000 a year in the three decades after 1846, more than doubled in the next two decades, and exceeded a million after the turn of the century (Hatton and Williamson 1998).

Some of the country-specific migration rates were enormous. During the 1880s the decadal emigration rate per thousand was 141.7 in Ireland and 95.2 in Norway, while the immigration rate per thousand was 221.7 in Argentina and 85.8 in the United States. In the first decade of the 20th century decadal emigration rates of 107.7 per thousand were recorded in Italy, while immigration rates per thousand were 291.8 in Argentina, 167.6 in Canada, 118.4 in Cuba, and 102.0 in the United States. There were also significant migrations within Europe and the New World, and emigration from Asia.

The United Nations has estimated that the world stock of migrants was 2.3 percent of the world population in both 1965 and 1990 (Zlotnik 1999). In Western Europe the share of migrants in the total population increased from 3.6 percent to 6.1 percent in 1965–90, while in North America the share increased from 6.0 percent to 8.6 percent (Zlotnik 1999, table 1a, p. 47). By contrast, in 1911 the foreign born accounted for 14.7 percent of the U.S. population and 22.0 percent of the Canadian population (U.S. Department of Commerce 1975, p. 14; Urquhart and Buckley 1965, p. 19).

Annual immigration to the United States averaged 770,000 in 1990–94 and 814,000 in 1995–96, implying decadal immigration rates of roughly 30 per thousand. Immigration in the early 1990s was proportionally higher in Canada, which saw decadal immigration rates of 70–80 per thousand in the early 1990s. Rates of around 80 per thousand were also recorded in Germany in 1990–94, while rates of around 50 per thousand were recorded in Australia and Germany in 1995–96. These flows, although fairly substantial, are dwarfed by those of the late 19th and early 20th centuries.

Once again, however, a caveat is in order. Labor market integration cannot be measured by the size of flows alone; the responsiveness of migration to wage dif-

ferentials is a better measure. Several papers have compared the responsiveness of interregional migration across countries in the context of the debate on the European Monetary Union (such as Eichengreen 1993; and Obstfeld and Peri 1998). But hardly any have calculated the changing responsiveness of migration to incentives over time. An exception is Hatton and Williamson (2001), which compares African (intracontinental) migration elasticities between 1977 and 1995 with the (intercontinental) elasticities in late-19th-century Europe and finds little difference between the two periods. More work along these lines is needed before we can make definitive statements about trends in labor market integration.

Mass migration will have the greatest impact on between-country inequality if it transfers population from poor to rich countries. In the late 19th century migration was clearly of this form, since Europe was significantly poorer than the New World; however, emigration was initially higher from the richer European regions, with the poorer southern and eastern regions becoming involved only with a lag. Something similar appears to have taken place in the late 20th century (Chiswick and Hatton 2001). For example, the share of developing country migrants in total U.S. immigration rose from 50 percent in the 1960s to 63 percent in the 1970s, 86 percent in the 1980s, and 80 percent in the early 1990s (Zlotnik 1999, table 3). Similar trends are apparent in Australia, Canada, and Europe (where immigration from Eastern Europe has increased over time). In both periods, then, mass migration was increasingly involving poorer countries and thus potentially making a bigger contribution to convergence. The big question for the 21st century is to what extent Africa will begin participating in mass, intercontinental migration (Hatton and Williamson 2001).

The impact of migration on within-country inequality depends largely on the skill mix, as suggested above. In the late 19th century migration involved predominantly young, unskilled adults with very high labor force participation rates; it thus had a large potential impact on inequality, lowering it in Europe and raising it in the New World. As the late 20th century progressed, the picture became increasingly similar, at least for the United States: the skill profile of immigrants, relative to that of the native born, has declined dramatically since the mid-1960s (Borjas 1999, chapter 2).[6] In several countries, however, policy has responded by encouraging more immigration of skilled workers, often through temporary work permit programs. In principle, this could lead to greater inequality in emigrant economies and greater equality in immigrant countries—the opposite of what occurred in the late 19th century.

Trends in Inequality

The previous section documented a U-shaped trend in international economic integration, with integration in the late 19th century followed by disintegration in the interwar period and a recovery since World War II. The recovery has been uneven, however, with some dimensions of integration remaining undeveloped relative to 1913 (such as migration) and with some countries participating less fully in the recovery than others (in particular, certain developing countries). The question now

arises, what has been the experience of inequality over the past two centuries, and to what extent can this experience be related to trends in globalization?

The benchmark study of trends in world income inequality over the past two centuries is Bourguignon and Morrisson (1999), which uses data on population, real GDP per capita, and income shares by vintile (a 20th of the population) for 33 groups of countries between 1820 and 1992. The resulting Theil coefficients for world inequality, and a decomposition into the part explained by between-country inequality and the part explained by within-country inequality, point to several key lessons (figure 2).

First, world inequality has increased substantially since 1820. Between 1820 and 1910, a period of rapid growth and globalization, the Theil coefficient rose from 0.533 to 0.799, a rise of 0.266 (0.030 per decade), or 50 percent. Inequality remained fairly stable between 1910 and 1960, dipping during the 1950s before resuming its rise after 1960, another period of rapid growth and globalization. Between 1960 and 1992 the Theil coefficient rose by 0.093 (0.029 per decade, roughly the same rate as in the 19th century), or by 12 percent.

Second, the rise in total inequality in 1820–1992 was driven entirely by a rise in inequality between countries; indeed, within-country inequality declined over the period. Between-country inequality rose continuously from 1820 to 1950; it fell

Figure 2. World Inequality, 1820–1992

Theil inequality index
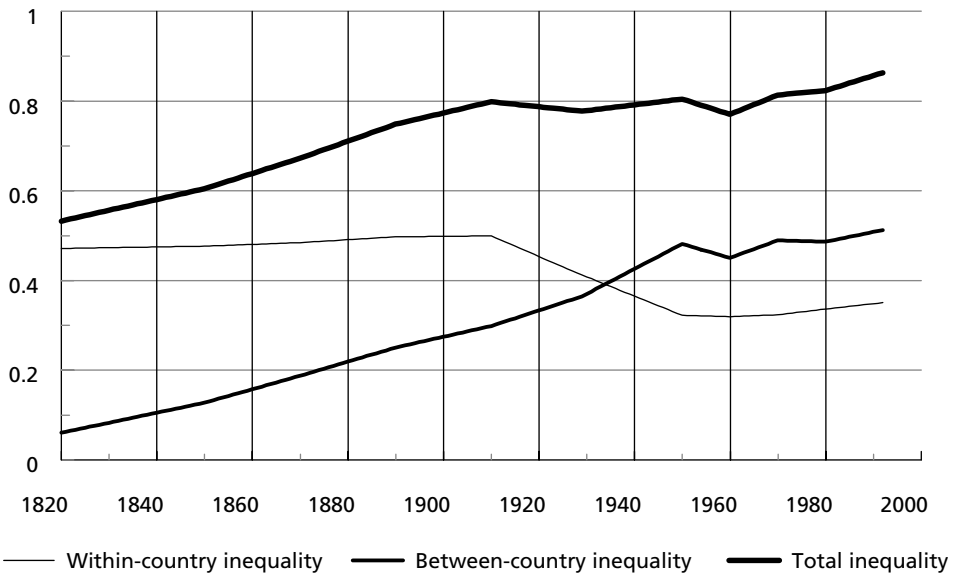
Source: Bourguignon and Morrisson (1999).

during the 1950s and 1970s, but rose during the 1960s and 1980s and rose over the post-1960 period as a whole. This confirms Pritchett's (1997) finding that divergence, rather than convergence, characterizes the long-run aggregate growth record.

Third, the cessation of the trend toward greater overall inequality during the interwar period was not due to more favorable trends in between-country inequality, since between-country inequality continued to rise rapidly. It was due instead to a dramatic decline in within-country inequality, which fell from 0.500 in 1910 to 0.323 in 1950, a decline of 0.177 (35 percent) in four decades, or 0.044 per decade. This interwar experience seems to have been an aberration, since within-country inequality trended gently and continuously upward both before 1910 and after 1950. Over the period 1820–1992 as a whole, within-country inequality declined sharply, as a result of this apparent step decline during the interwar period. This is in clear contrast to between-country inequality, which increased sharply.

Fourth, while within-country inequality was the dominant force driving total world inequality in 1820, its relative importance has declined over time. Since World War II between-country inequality has been the most important factor in total inequality. For those concerned about world inequality, this suggests that while their focus should have been on domestic redistribution in the early 19th century, now it should be firmly focused on policies designed to help poor countries achieve macroeconomic convergence with the rich.

Such convergence may at last be happening. While the long-run evidence clearly points to between-country divergence, recent papers suggest that this trend was replaced by convergence at the end of the 20th century. Melchior, Telle, and Wiig (2000) calculate (population-weighted) Gini coefficients for world income distribution, using per capita income data for 115 countries. They find that the Gini coefficient fell from 0.59 in 1965 to 0.52 in 1997, with China's catch-up being crucial. Schultz (1998) finds that between-country inequality started falling in the mid-1970s; Boltho and Toniolo (1999) find that it started falling in 1980. A common feature of these papers is their use of income data adjusted for purchasing power parity; the implication is that there has been what Lindert and Williamson (2001) refer to as an "epochal turning point." According to Schultz, the convergence since the mid-1970s has been strong enough so that overall world inequality (both between- and within-country) has also started to decline.[7]

To what extent has globalization been responsible for any of these trends? I turn first to the globalization boom of the late 19th century.

Globalization and Inequality in the Late 19th Century

As we have seen, the late 19th century was characterized by dramatically declining transport costs, by mass migration from the Old World to the New, and by large transfers of capital from the Old World to the New. How did each of these dimensions of globalization influence income distribution within and between countries? Let's look first at the evidence on factor prices, and begin with the within-country evidence.

Trends in Factor Prices

We have abundant evidence on relative factor prices in the late 19th century, compiled by O'Rourke, Taylor, and Williamson (1996) and by Jeffrey Williamson in a series of papers summarized in Williamson (1998, 2000). These papers present data on w/r, the ratio of the unskilled, urban wage to the returns to agricultural land. This ratio was the key relative factor price in an era in which agriculture was still an important part of the economy and in which intercontinental trade was dominated by the exchange of resource- and land-intensive products for labor-intensive products such as manufactured goods.

The metaphor that motivated Heckscher and Ohlin in the first place was one of the land-abundant New World exchanging food for European manufactured goods. Their logic suggests that in times of globalization w/r should have converged internationally. In land-abundant New World economies, where w/r was high, it should have declined; and in land-scarce European economies, where w/r was low, it should have increased. Moreover, in absolute terms low European wages should have caught up with high New World wages, while low New World land prices should have caught up with high European land prices. By and large, these suppositions hold true for the late 19th century.

Between 1870 and 1910 real land prices increased by more than 400 percent in Australia and by more than 250 percent in the United States. Moreover, in three European countries—Britain, France, and Sweden—land prices fell, in Britain by more than 50 percent. There was certainly absolute convergence in the returns to land during this period. Meanwhile, wages in Europe were converging on New World wages (Williamson 1995), although this absolute wage convergence was more modest in scale and there were poor countries that did not participate in the convergence.

The net result was that the 40 years after 1870 saw substantial convergence in relative factor prices, with wage-rental ratios rising in Europe and falling in the New World, just as Heckscher and Ohlin would have predicted (Williamson 2000, table 3). By 1910 the Argentine ratio had fallen to a fifth of its mid-1880 level, the Australian ratio to a quarter of its 1870 level, and the U.S. ratio to half of its 1870 level. In Europe by 1910, the Irish ratio had increased by a factor of 5.6 over its 1870 level, the Danish ratio by a factor of 3.1, the British ratio by a factor of 2.7, and the Swedish ratio by a factor of 2.6. The increase was less pronounced in protectionist economies: the ratio increased by a factor of 2.0 in France, 1.4 in Germany, and not at all in Spain.

The Heckscher-Ohlin logic was also well borne out by the experience of the developing countries that participated in the late-19th-century global economy (Williamson 2000, table 4). In land-scarce economies such as Japan, Korea, and Taiwan the wage-rental ratio increased substantially, while it plummeted in land-abundant food exporting economies such as Argentina, Burma, Egypt, the Punjab, Siam, and Uruguay. Relative factor price convergence was not limited to the present-day OECD countries, it appears.

What was responsible for these trends, particularly for the impressive convergence in wage-rental ratios? O'Rourke, Taylor, and Williamson (1996) explore this issue econometrically, using data for seven countries between 1875 and 1914: Australia, Britain, Denmark, France, Germany, Sweden, and the United States. The results support Heckscher and Ohlin: the ratio of agricultural to manufacturing prices has the expected negative effect on wage-rental ratios for five of the seven countries, the exceptions being Australia and Denmark. It appears that commodity market integration *was* important in driving factor price convergence during this period.

Other dimensions of globalization also contributed to the convergence in wage-rental ratios. In particular, migration increased wage-rental ratios in Europe and lowered them in the New World. However, international capital flows were probably a force for divergence rather than convergence during this period. Capital flowed from low-wage Europe to the high-wage New World, exacerbating rather than moderating differences in wage-rental ratios.

Factor Price Convergence and Inequality in the Late 19th Century

What were the implications of these factor price movements for inequality? To a large extent they depended on who owned the land. Landowners were typically at the top of the income distribution, so the globalization forces that raised wage-rental ratios in Europe should have made Europe more equal, while the same forces that lowered wage-rental ratios in the New World should have made those societies more unequal. Presumably the inequality impact was greater where landholding was more concentrated, such as in Latin America, than where the family farm predominated, such as in the northern United States or in Burma. Meanwhile, intercontinental migrants were typically unskilled. Thus immigration would have lowered unskilled wages and raised inequality in the New World, but raised unskilled wages and lowered inequality in Europe. Broadly speaking, globalization should have made the relatively poor Old World more equal, but at the expense of higher inequality in the affluent New World—precisely the correlation that Heckscher-Ohlin thinking would predict today.

Complete income distributions are typically unavailable for the late 19th century, but Williamson (1997) constructed an alternative measure of inequality: the ratio of the unskilled wage to GDP per worker hour, w/y. This measure compares the income of those at the bottom of the distribution with a weighted average of all other relevant factor prices—skilled wages, but also returns to such factors as land and capital, which were highly relevant both 100 years ago and in the developing world of the late 20th century. Williamson found that inequality fell dramatically (w/y increased, from 100 in 1870 to 153 or 154 in 1913) in poor European countries like Denmark and Sweden, where w was initially low. But it rose substantially in rich New World economies like the United States and Australia, where w was initially high (w/y fell, from 100 in 1870 to 53 or 58 in 1913). Inequality remained more stable in the richer European economies and in peripheral economies that remained aloof from globalization (such as Iberia).

The OECD evidence is thus consistent with what the Heckscher-Ohlin model would have predicted: globalization should have raised inequality in rich, equal societies and lowered it in poor, unequal societies. The developing world evidence is more mixed, however. Where data are available, they show w/y falling during the late 19th century—in Argentina, Brazil, Burma, India, Indonesia, Japan, Mexico, the Philippines, Taiwan, Thailand, and Uruguay (Williamson 2000, tables 4, 5). For Latin America there is an explanation consistent with that given for other New World economies—globalization lowered wage-rental ratios, which should boost inequality. Furthermore, countries such as Argentina were not low wage by late-19th-century standards.

The case of Asia is more puzzling. First, inequality rose everywhere, even in indisputably poor countries. Second, inequality rose not only in land-abundant areas such as Burma and Thailand, where wage-rental ratios fell, but also in land-scarce economies such as Japan and Taiwan, where wage-rental ratios rose. Presumably this anomaly is due to the fact that inequality trends were determined not just by globalization (and wage-rental ratios), but also by other factors, such as demographic and technological change.

To summarize, there appears to be a causal relationship between globalization and within-country inequality in the late 19th century. Trade did have an impact on wage-rental ratios, just as theory says it should have. Moreover, Williamson (1997) shows that there is a strong relationship between migration flows and movements in w/y, with w/y rising more (falling less) in countries that experienced more emigration (less immigration). However, the episode shows that the links between globalization and distribution are subtle and varied, just as was suggested earlier.

First, globalization had different effects on factor prices and inequality on different continents: trade raised w/r in Europe and lowered it in the New World, and migration raised w/y in Europe and lowered it in the New World. This has to be borne in mind when inspecting the average inequality trends in figure 2: the dramatic egalitarian trends in some European countries during this period, and the equally dramatic inegalitarian trends in some New World countries at the same time, all of which were closely linked to globalization, largely canceled each other out in the aggregate. Thus a regression of inequality on some measure of globalization that failed to take account of the very different links between the two variables on different continents might well incorrectly conclude that on balance there was no link between globalization and inequality.

Second, different dimensions of globalization had different effects on distribution, with migration raising European wages, for example, and capital flows lowering them. And third, the impact of a given factor price change on inequality (for example, a rise in the returns to land in land-abundant countries) depended on the distribution of endowments, as the contrast between Burmese peasant landowners and Argentine *estancia* owners makes clear.

Globalization and Between-Country Inequality

Between-country inequality rose substantially during the late 19th century, a period of rapid globalization (see figure 2). Moreover, while relative factor prices were con-

verging internationally, and while there was absolute factor price convergence within the Atlantic economy, globally there was absolute factor price divergence, at least for real wages. Real wages in Burma, India, Indonesia, Japan, Taiwan, and Thailand all fell further behind British wages during this period, reflecting superior European growth (Williamson 1998, tables 1, 2). Was globalization responsible?

O'Rourke and Williamson (1997, 1999) explore the links between globalization and convergence for a sample of European and New World economies between 1870 and 1913. Their first finding mirrors a late-20th-century one: while there may have been divergence for the world as a whole, there was convergence for this smaller sample of rich countries, although the convergence was weak. However, it is the variety of country experiences that stands out, rather than this general pattern: while some countries, like Ireland and Italy, converged on the leaders of the day (like Britain and the United States) at about the expected rate, others (like the Scandinavian countries) converged much faster than expected, and still others, like the Iberian countries, did not converge at all.

O'Rourke and Williamson then quantify the trade, migration, and capital flow shocks that hit these economies during the late 19th century and calculate the contribution of each of these forces to the patterns of convergence and divergence that the data reveal. In each case the methodology is the same: Calculate the impact of trade, or migration, or capital flows on real wages in a peripheral country and in Britain and the United States. Infer the impact of the shock on the real wage gap between the core and the periphery, and express this change as a percentage of the total change in the wage gap. Mass migration and international capital flows explain a third to a half of the Scandinavian countries' catch-up with Britain and about a half to nine-tenths of their catch-up with the United States. These forces explain more than two-thirds of Ireland's and Italy's catch-up with Britain and all their catch-up with the United States. The Iberian countries' failure to converge on the leaders can be attributed in large part to their failure to import enough capital and send out enough people.

Globalization thus helped several peripheral European countries converge on the core, while insufficient globalization helps to explain Iberia's failure to converge. The crucial factor was migration, which accounted for some 70 percent of the total convergence in the Atlantic economy during the period (Taylor and Williamson 1997). Trade may have been important for within-country distribution, but it played a much smaller role in between-country distribution. (And, as mentioned, capital flows were a force for divergence.) The rising between-country inequality of the late 19th century seems not to have been due to globalization.

Globalization and Inequality in the Late 20th Century

The Heckscher-Ohlin model provides a good guide to the experience of the late 19th century, in which trade was dominated by the exchange of food for manufactures, the two key regions were the Old and New Worlds, and the two key factors of production were land and labor. In the context of the late 20th century the

debate has often assumed that the two key factors are skilled and unskilled labor, and the two key regions the North and South. Heckscher-Ohlin logic implies that under these conditions globalization should lead to rising skill premiums and inequality in the North and falling skill premiums and inequality in the South (Wood 1994).

These predictions have not been borne out. For example, although the Heckscher-Ohlin predictions were largely vindicated by the East Asian experience of the 1960s and 1970s, skill premiums seem to have widened in several Latin American countries following liberalization in the 1980s (Robbins 1996; Wood 1997). Beyond this small, often-studied group of countries, the most comprehensive evidence on wage dispersion comes from the International Labour Organisation's October Surveys, summarized in Freeman and Oostendorp (2000). Consistent with many other studies and with Heckscher-Ohlin theory, Freeman and Oostendorp find sharply rising wage inequality in the United Kingdom and the United States in the 1980s and 1990s. However, just as many advanced countries saw wage dispersion falling as saw it rising during this period. Moreover, through regression analysis, Freeman and Oostendorp find wage differentials falling in high-income and upper-middle-income countries in the 1980s and 1990s, and rising in lower-middle-income countries (as well as formerly Communist countries)—a correlation opposite to what theory would predict.

Inequality has been increasing in most OECD countries since the 1970s, particularly since the mid-1980s (Burniaux and others 1998). Here again, however, there are exceptions, such as Canada, Denmark, and France. The increased inequality is due mostly to widening gaps in labor earnings. As Lindert and Williamson (2001) point out, if labor earnings gaps are widening but wage differentials are not, unemployment and reductions in hours worked must be playing a large role in driving inequality trends in the OECD countries.

Among developing countries the picture is mixed. Inequality has been steadily declining in Latin America since the 1960s, despite what happened to skill premiums in the 1980s (table 3). The patterns in Africa and the Pacific Rim are erratic, with inequality rising between the 1960s and 1970s, falling through the 1980s, and rising again between the 1980s and 1990s. Inequality has been rising in China and India since the mid-1980s, and this should dominate any population-weighted inequality index for developing countries (Lindert and Williamson 2001). Again, this trend of rising within-country inequality in the South is not what simple 2x2 Heckscher-Ohlin models would lead us to expect.

Of course, these trends do not on their own disprove simple trade theory, since distribution is driven by many factors other than globalization. For example, political developments disfavoring unions, or the entry of China with its vast reserves of unskilled workers into the world market, or the simultaneous and unrelated introduction of new technology disfavoring unskilled workers might account for the increased wage inequality in Latin America (Wood 1997). Alternatively, such factors as democratization, demographic changes, educational developments, or the collapse of Communism might have been the most important factors behind inequality

Table 3. Inequality in the Late 20th Century

Country group	1960s	1970s	1980s	1990s
OECD				
Gini coefficient	0.347	0.336	0.326	0.330
GAP	6.9	6.6	6.2	6.5
Number of countries	12	19	20	13
Africa				
Gini coefficient	0.453	0.498	0.416	0.464
GAP	12.2	17.5	9.6	12.9
Number of countries	4	4	11	15
Latin America				
Gini coefficient	0.536	0.504	0.501	0.500
GAP	21.2	17.0	16.2	13.3
Number of countries	6	12	12	10
Pacific Rim				
Gini coefficient	0.374	0.390	0.385	0.392
GAP	8.3	9.0	7.9	8.1
Number of countries	6	9	10	7

Note: GAP is the ratio of the income earned by the top income quintile to the income earned by the bottom quintile.
Source: Higgins and Williamson (1999), based on Deininger and Squire (1996).

trends. As always, we need multivariate analysis to disentangle these effects, and it is to such studies that we now turn.

Within-Country Inequality and Globalization: Cross-Country Studies

Since the publication of Deininger and Squire's (1996) data set on income inequality, there has been a proliferation of articles exploring the determinants of late-20th-century inequality across countries and over time. Many ask whether there is support for Kuznets's (1955) prediction that in the initial stages of growth inequality will increase as a result of structural change, while in the later stages it will moderate.

It is the relationship between inequality and openness that concerns us here, however, and the literature provides ambiguous answers. For example, when Higgins and Williamson (1999) regress inequality on openness (using the Sachs and Warner [1995] dummy variable), together with the Kuznets variables and cohort size, the coefficient has a negative but insignificant sign. When, in addition, they interact openness with dummy variables indicating whether a country is in the bottom or top third of the international labor productivity distribution, the coefficient on openness becomes negative and significant, while the interaction terms are insignificant. This finding suggests that the standard Stolper-Samuelson theory is not applicable; the negative effect of openness on inequality suggests that globalization has a benign effect on income inequality. This effect turns out to be modest, however.

By contrast, Barro (2000)—using an expanded version of the Deininger-Squire data set and a synthetic openness measure based on regressions of trade share on population, land area, and trade policy—finds that openness is positively related to inequality. Moreover, an interaction term between openness and GDP per capita is

negative and significant: openness raises inequality below per capita incomes of about $13,000 (1985 U.S. dollars) and lowers it at per capita incomes above that amount. This finding is of course at odds with the Heckscher-Ohlin prediction relating to trade and skill premiums.

Spilimbergo, Londoño, and Székely (1999) provide the empirical analysis of the Deininger-Squire data set most closely related to Heckscher-Ohlin thinking. Factor prices are related to endowments and prices of traded goods, and these prices depend on world endowments and trade policies. In the empirical specification inequality is related to country endowments of capital per worker, arable land per worker, and skill intensity, relative to the "effective" world endowment of the factor in question, to these endowment variables interacted with a synthetic trade openness measure, to openness, and to income and income squared.

Spilimbergo, Londoño, and Székely find that openness is positively correlated with inequality, although this finding is not robust to the choice of openness measure. Openness increases inequality in skill-abundant countries, consistent with the Anglo-American evidence and with the Heckscher-Ohlin model. However, openness reduces inequality in land- and capital-abundant countries, inconsistent with the Heckscher-Ohlin intuition that trade should increase the returns to land and capital in such countries and thus raise inequality if these resources are unequally distributed.

The existing cross-country studies leave unanswered many questions about the links between openness and inequality. Spilimbergo, Londoño, and Székely find that the sign of the effect depends on the choice of openness indicator. And while Higgins and Williamson find that openness is negatively associated with inequality, Barro finds a positive association. The results do not seem to be robust to the specification of the equation or to the choice of openness indicator. Further research clearly seems warranted, although the finding that openness has at most a modest impact on inequality (in either direction) seems fairly robust.

Moreover, as O'Rourke and Williamson (1999) emphasize, cross-country regressions can conceal as much as they reveal, and the diversity of country experience is as interesting as average correlations. As we have already seen, differing relationships between openness and inequality in different countries may cancel one another out in the aggregate. Furthermore, there are many dimensions of openness, as we saw above, and each may have different effects on income distribution. We need to distinguish among these dimensions of openness, see to what extent globalization has affected countries along each of these dimensions, and compute the effects of each of these shocks on inequality using well-specified economic models.

Between-Country Inequality and Globalization

An earlier section argued that there was no causal link between late-19th-century globalization and between-country divergence; instead, globalization—in particular, migration—was a force for convergence. Is the same true of the late 20th century? If convergence has indeed replaced divergence since the mid-1980s, could globalization be responsible? In an often-cited study Ben-David (1993) shows that there

was substantial convergence among the original six members of the European Economic Community (EEC) after 1950 and argues that there is a link between intra-EEC trade liberalization and this convergence. For example, dispersion among Denmark, Ireland, and the United Kingdom increased until the mid-1960s, when these countries started to liberalize their trade with one another; dispersion among the three declined after 1973, when they joined the EEC. Turning to the world as a whole, Sachs and Warner (1995) find a strong tendency toward convergence among countries pursuing open trade policies, but not among more closed economies.

However, these arguments are entirely *post hoc, ergo propter hoc.* Moreover, Ben-David's argument that postwar convergence must have been due to liberalization, since there was no convergence before 1945, is incorrect, as O'Rourke and Williamson (1999, chapter 2) show and as Rodríguez and Rodrik (2001) point out. These correlations are fascinating and suggestive. But we need rigorous model-based analysis if Ben-David's argument is to be made convincingly, especially since Slaughter (2001), using a more sophisticated difference-in-differences analysis, finds no relationship between trade liberalization and convergence.[8]

The dimension of globalization that had the greatest impact on convergence a hundred years ago was not trade, but labor flows—and political reality suggests that mass intercontinental migration will not be allowed to play the same role in the future. Capital flows and technology transfers, operating through trade or foreign direct investment, will have to take up the slack (see, for example, Coe and Helpman 1995; and Keller 2001). Work by Alan Taylor on Latin America (for example, Taylor 2000) suggests that capital flows, where policy allows them, can have a major impact on peripheral economies. Unfortunately, capital markets of the late 20th century seem to have done less well in channeling savings toward developing countries than those of the late 19th century. Ensuring that developing countries benefit from these flows will be crucial for convergence in the future.

Conclusion

What have we learned thus far? We have learned that globalization can have an important impact on within-country income distribution, and that it did so in the late 19th century. However, for the reasons outlined at the start of the article, the relationship between globalization and inequality will vary depending on the country being considered, the dimension of globalization involved, and the distribution of endowments. The experience in the late 19th century illustrates these points clearly, and there is no reason to suppose that things were different in the 20th century. It follows that searching for average relationships between summary measures of globalization and inequality across countries is a hazardous enterprise: results are likely to be sensitive to the countries chosen and the openness indicator used. In this light, the failure of the cross-country regressions surveyed above to yield robust correlations should come as no surprise.

We have learned that in the 19th century globalization was positively associated with between-country convergence. But a generally liberal world environment does

not ensure that poor countries will catch up with the rich: Iberia remained relatively isolated from international factor markets during this period and paid the price. The evidence from the late 20th century remains inconclusive. If globalization is indeed associated with convergence, why was there so much between-country divergence in both the late 19th century and the late 20th (at least until the 1980s)? The answer, presumably, is superior technological progress in rich countries. Globalization was not the culprit.

Lucas (2000) describes, in the context of a simple general model, how divergence might coincide with globalization. The model assumes that countries randomly leave behind a preindustrial past and embark on modern industrial growth, with the probability that this occurs depending on world income. The model further assumes that when countries do embark on such growth, they catch up with the economic leaders—a phenomenon that may, as Lucas suggests, be due to the diffusion of resources or ideas across borders (and, by implication, be positively related to international economic integration). The model predicts a worldwide Kuznets curve, and the recent papers surveyed above suggest that while we were on the upward slope of the curve until recently, we may now have reached the turning point and be experiencing convergence.

We have learned that in many respects we simply do not know enough about globalization and convergence and the relationship between them. First, we lack the commodity price data required to document the evolution of international commodity market integration across the 20th century. If the 19th century can yield such information to diligent scholars, surely the 20th century can do the same.

Second, we need more information on factor prices, which, according to theory, are what should be linked to globalization in the first place. Moreover, we need data not just on different categories of wages, but on the returns to land and capital. It is the availability of such factor price data that has allowed economic historians to uncover the links between international integration and income distribution in the 19th century. Aggregate measures of inequality, such as Gini coefficients, are too crude and influenced by too many different factors for relationships to emerge cleanly from the data.

Third, we need to carry out cross-country growth regressions in ways more consistent with trade theory. We need to interact our openness indicators with more country-level characteristics, as Spilimbergo, Londoño, and Székely (1999) do. Perhaps even more important, we need to enter different openness measures into our regressions and see whether different dimensions of globalization have different effects. And it would be helpful to run regressions with factor price ratios, rather than aggregate inequality indicators, on the left-hand side.

Fourth, we need to supplement cross-country exercises with more country studies, since the impact of globalization on inequality differs so greatly across countries.

Finally, we have learned that the evidence may provide grounds for some optimism about future trends in inequality. The rise in inequality over the past 200 years has been due mostly to a rise in between-country inequality, which is now the dominant cause of overall inequality. This trend may now have been reversed, and the

experience of the 19th century (and arguably of the late 20th as well) suggests that increased globalization will accelerate this decline.

Notes

1. In so doing the article follows in the footsteps of Lindert and Williamson (2001).

2. Indeed, Baier and Bergstrand (2001) report that two-thirds of the trade boom in the late 20th century can be accounted for by income growth, suggesting that commodity market integration had little to do with it. O'Rourke and Williamson (2001) find something similar for the period 1500–1800, which saw little or no commodity market integration. There is no similar accounting for the rapid globalization of the 19th century, however.

3. This statistic, based on data kindly supplied by Alan Taylor, refers to 1885–1910.

4. These data are based on the World Bank's *World Development Indicators 1999,* and on data kindly supplied by Alan Taylor.

5. Crucially, Latin American countries such as Argentina were relatively rich (and resource rich) then.

6. Indeed, a higher proportion of immigrants were high school dropouts in 1998 than in 1990, reversing a steady downward trend in this figure since 1960 (Borjas 1999, p. 21).

7. But see Milanovic (1999) for a contrary view, though one based on data for only two years (1988 and 1993). See also Dowrick and Akmal (2001).

8. Furthermore, Rodríguez and Rodrik (2001) have cast doubt on the Sachs-Warner dummy variable, while it is always possible that countries that pursued open policies shared some other characteristic that made them more likely to converge on the core.

References

Baier, Scott L., and Jeffrey H. Bergstrand. 2001. "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity." *Journal of International Economics* 53(1): 1–27.

Bairoch, Paul. 1989. "European Trade Policy, 1815–1914." In Peter Mathias and Sidney Pollard, eds., *The Cambridge Economic History of Europe.* Vol. 8. Cambridge: Cambridge University Press.

———. 1993. *Economics and World History: Myths and Paradoxes.* Chicago: University of Chicago Press.

Bairoch, Paul, and Richard Kozul-Wright. 1998. "Globalization Myths: Some Reflections on Integration, Industrialization and Growth in the World Economy." In R. Kozul-Wright and R. Rowthorn, eds., *Transnational Corporations and the Global Economy.* New York: St. Martin's Press.

Baldwin, Richard E., and Philippe Martin. 1999. "Two Waves of Globalization: Superficial Similarities, Fundamental Differences." In Horst Siebert, ed., *Globalisation and Labour.* Tübingen, Germany: J. C. B. Mohr for Kiel Institute of World Economics.

Barro, Robert J. 2000. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth* 5(1): 5–32.

Ben-David, Dan. 1993. "Equalizing Exchange: Trade Liberalization and Income Convergence." *Quarterly Journal of Economics* 108(3): 653–79.

Boltho, Andrea, and Gianni Toniolo. 1999. "The Assessment: The Twentieth Century—Achievements, Failures, Lessons." *Oxford Review of Economic Policy* 15(4): 1–17.

Bordo, Michael D., Barry Eichengreen, and Douglas A. Irwin. 1999. "Is Globalization Today Really Different Than Globalization a Hundred Years Ago?" *Brookings Trade Forum 1999,* pp. 1–50.

Bordo, Michael D., Barry Eichengreen, and Jongwoo Kim. 1998. "Was There Really an Earlier Period of International Financial Integration Comparable to Today?" In *The Implications of Globalization of Financial Markets*. Seoul: Bank of Korea.

Borjas, George. 1999. *Heaven's Door: Immigration Policy and the American Economy*. Princeton, N.J.: Princeton University Press.

Bourguignon, François, and Christian Morrisson. 1999. "The Size Distribution of Income Among World Citizens: 1820–1990." World Bank, Paris. Processed.

Burniaux, Jean-Marc, Thai-Thanh Dang, Douglas Fore, Michael Förster, Marco Mira d'Ercole, and Howard Oxley. 1998. "Income Distribution and Poverty in Selected OECD Countries." OECD Economics Department Working Paper 189. Organisation for Economic Co-operation and Development, Paris.

Chiswick, Barry, and Timothy J. Hatton. 2001. "International Migration and the Integration of Labor Markets." Paper presented at National Bureau of Economic Research conference on Globalization in Historical Perspective, Santa Barbara, Calif., 3–6 May.

Clemens, Michael A., and Jeffrey G. Williamson. 2001. *Where Did British Foreign Capital Go? Fundamentals, Failures and the Lucas Paradox, 1870–1913*. NBER Working Paper 8028. Cambridge, Mass.: National Bureau of Economic Research.

Coe, David T., and Elhanan Helpman. 1995. "International R&D Spillovers." *European Economic Review* 39(5): 859–87.

Collins, William J. 1996. "Regional Labor Markets in British India." Harvard University, Department of Economics, Cambridge, Mass.

Coppel, Jonathan, and Martine Durand. 1999. "Trends in Market Openness." OECD Economics Department Working Paper 221. Organisation for Economic Co-operation and Development, Paris.

Currie, Janet, and Ann E. Harrison. 1997. "Sharing the Costs: The Impact of Trade Reform on Capital and Labor in Morocco." *Journal of Labour Economics* 15(3): S44–72.

Davis, Donald R. 1996. *Trade Liberalization and Income Distribution*. NBER Working Paper 5693. Cambridge, Mass.: National Bureau of Economic Research.

Deininger, Klaus, and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *World Bank Economic Review* 10(3): 565–91.

Dowrick, Steve, and Muhammad Akmal. 2001. "Contradictory Trends in Global Income Inequality: A Tale of Two Biases." Australian National University, School of Economics, Canberra.

Dunning, John H. 1993. *Multinational Enterprises and the Global Economy*. New York: Addison-Wesley.

Eichengreen, Barry. 1993. "Labor Markets and European Monetary Unification." In Paul R. Masson and Mark P. Taylor, eds., *Policy Issues in the Operation of Currency Unions*. Cambridge: Cambridge University Press.

Feenstra, Robert C., and Gordon H. Hanson. 1996. "Foreign Investment, Outsourcing, and Relative Wages." In Robert C. Feenstra, Gene M. Grossman, and Douglas A. Irwin, eds., *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*. Cambridge, Mass.: MIT Press.

Feis, Herbert. 1930. *Europe, the World's Banker, 1870–1914*. New Haven, Conn.: Yale University Press.

Feldstein, Martin, and Charles Horioka. 1980. "Domestic Saving and International Capital Flows." *Economic Journal* 90(358): 314–29.

Findlay, Ronald, and Kevin H. O'Rourke. 2001. "Commodity Market Integration, 1500–2000." Paper presented at National Bureau of Economic Research conference on Globalization in Historical Perspective, Santa Barbara, Calif., 3–6 May.

Freeman, Richard B., and Remco H. Oostendorp. 2000. *Wages around the World: Pay across Occupations and Countries*. NBER Working Paper 8058. Cambridge, Mass.: National Bureau of Economic Research.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy.* Cambridge, Mass.: MIT Press.

Hanson, Gordon H., and Ann E. Harrison. 1999. "Trade and Wage Inequality in Mexico." *Industrial and Labor Relations Review* 52(2): 271–88.

Harley, C. Knick. 1980. "Transportation, the World Wheat Trade, and the Kuznets Cycle, 1850–1913." *Explorations in Economic History* 17(3): 218–50.

———. (1988). "Ocean Freight Rates and Productivity, 1740–1913: The Primacy of Mechanical Invention Reaffirmed." *Journal of Economic History* 48: 851–76.

Hatton, Timothy J., and Jeffrey G. Williamson. 1998. *The Age of Mass Migration: Causes and Economic Impact.* Oxford: Oxford University Press.

———. 2001. *Demographic and Economic Pressure on Emigration out of Africa.* NBER Working Paper 8124. Cambridge, Mass.: National Bureau of Economic Research.

Higgins, Matthew, and Jeffrey G. Williamson. 1999. *Explaining Inequality the World Round: Cohort Size, Kuznets Curves, and Openness.* NBER Working Paper 7224. Cambridge, Mass.: National Bureau of Economic Research.

Hummels, David. 1999. "Have International Transportation Costs Declined?" Purdue University, Department of Economics, West Lafayette, Ind.

International Monetary Fund. Various years. *International Financial Statistics.* Washington, D.C.

Isserlis, L. 1938. "Tramp Shipping Cargoes and Freights." *Journal of the Royal Statistical Society* 101(Pt. 1): 304–417.

Jones, Geoffrey. 1996. *The Evolution of International Business: An Introduction.* London: Routledge.

Jones, Matthew T., and Maurice Obstfeld. 1997. *Saving, Investment, and Gold: A Reassessment of Historical Current Account Data.* NBER Working Paper 6103. Cambridge, Mass.: National Bureau of Economic Research.

Kanbur, Ravi. 1999. "Income Distribution Implications of Globalization and Liberalization in Africa." Cornell University, Department of Economics, Ithaca, N.Y.

Keller, Wolfgang. 2001. *The Geography and Channels of Diffusion at the World's Technology Frontier.* NBER Working Paper 8150. Cambridge, Mass.: National Bureau of Economic Research.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45(1): 1–28.

Lindert, Peter H., and Jeffrey G. Williamson. 2001. "Does Globalization Make the World More Unequal?" Paper presented at National Bureau of Economic Research conference on Globalization in Historical Perspective, Santa Barbara, Calif., 3–6 May.

Lucas, Robert E. 2000. "Some Macroeconomics for the 21st Century." *Journal of Economic Perspectives* 14(1): 159–68.

Melchior, Arne, Kjetil Telle, and Henrik Wiig. 2000. "Globalisation and Inequality: World Income Distribution and Living Standards, 1960–1998." Studies on Foreign Policy Issues, Report 6B: 2000. Royal Norwegian Ministry of Foreign Affairs, Oslo.

Milanovic, Branko. 1999. "True World Income Distribution, 1988 and 1993: First Calculations, Based on Household Surveys Alone." Policy Research Working Paper 2244. World Bank, Development Research Group, Washington, D.C.

Obstfeld, Maurice, and Giovanni Peri. 1998. "Regional Non-Adjustment and Fiscal Policy." *Economic Policy* 26: 205–59.

Obstfeld, Maurice, and Alan M. Taylor. 1998. "The Great Depression as a Watershed: International Capital Mobility in the Long Run." In Michael D. Bordo, Claudia D. Goldin, and Eugene N. White, eds., *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century.* Chicago: University of Chicago Press.

———. 2001. "Globalization and Capital Markets." Paper presented at National Bureau of Economic Research conference on Globalization in Historical Perspective, Santa Barbara, Calif., 3–6 May.

O'Rourke, Kevin H. 2002. "Europe and the Causes of Globalization, 1790 to 2000." Paper presented at the conference From Europeanization of the Globe to the Globalization of Europe, Geneva, 10–12 January.

O'Rourke, Kevin H., and Jeffrey G. Williamson. 1997. "Around the European Periphery, 1870–1913: Globalization, Schooling and Growth." *European Review of Economic History* 1(2): 153–91.

———. 1999. *Globalization and History: The Evolution of a Nineteenth Century Atlantic Economy*. Cambridge, Mass.: MIT Press.

———. 2001. *After Columbus: Explaining the Global Trade Boom, 1500–1800*. NBER Working Paper 8186. Cambridge, Mass.: National Bureau of Economic Research.

O'Rourke, Kevin H., Alan M. Taylor, and Jeffrey G. Williamson. 1996. "Factor Price Convergence in the Late Nineteenth Century." *International Economic Review* 37(3): 499–530.

Pritchett, Lant. 1997. "Divergence, Big Time." *Journal of Economic Perspectives* 11(3): 3–18.

Robbins, Donald J. 1996. "Evidence on Trade and Wages in the Developing World." OECD Technical Paper 119. Organisation for Economic Co-operation and Development, Paris.

Rodríguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2000*. Cambridge, Mass.: MIT Press.

Rodrik, Dani. 1999. *The New Global Economy and Developing Countries: Making Openness Work*. Washington, D.C.: Overseas Development Council.

———. 2000. "How Far Will International Economic Integration Go?" *Journal of Economic Perspectives* 14(1): 177–86.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–95.

Schultz, T. Paul. 1998. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." *Journal of Population Economics* 11(3): 307–44.

Slaughter, Matthew J. 2001. "Trade Liberalization and Per Capita Income Convergence: A Difference-in-Differences Analysis." *Journal of International Economics* 55(1): 203–28.

Spilimbergo, Antonio, Juan Luis Londoño, and Miguel Székely. 1999. "Income Distribution, Factor Endowments, and Trade Openness." *Journal of Development Economics* 59(1): 77–101.

Stokey, Nancy L. 1991. "Human Capital, Product Quality, and Growth." *Quarterly Journal of Economics* 106(2): 587–616.

Taylor, Alan M. 1996. *International Capital Mobility in History: The Saving-Investment Relationship*. NBER Working Paper 5743. Cambridge, Mass.: National Bureau of Economic Research.

———. 2000. "Latin America and Foreign Capital in the Twentieth Century: Economics, Politics, and Institutional Change." In Stephen H. Haber, ed., *Political Institutions and Economic Growth in Latin America: Essays in Policy, History, and Political Economy*. Stanford, Calif.: Hoover Institution Press.

Taylor, Alan M., and Jeffrey G. Williamson. 1994. "Capital Flows to the New World as an Intergenerational Transfer." *Journal of Political Economy* 102(2): 348–71.

———. 1997. "Convergence in the Age of Mass Migration." *European Review of Economic History* 1(1): 27–63.

UNCTAD (United Nations Conference on Trade and Development). 2000. *World Investment Report 2000*. New York and Geneva: United Nations.

Urquhart M.C. and K.A.H. Buckley, eds. 1965. *Historical Statistics of Canada.* Cambridge, UK: Cambridge University Press.

U.S. Department of Commerce. 1975. *Historical Statistics of the United States: Colonial Times to 1970.* Washington, D.C.: U.S. Bureau of the Census.

Williamson, Jeffrey G. 1995. "The Evolution of Global Labor Markets Since 1830: Background Evidence and Hypotheses." *Explorations in Economic History* 32(2): 141–96.

———. 1997. "Globalization and Inequality, Past and Present." *World Bank Research Observer* 12(2): 117–35.

———. 1998. "Real Wages and Relative Factor Prices in the Third World before 1940: What Do They Tell Us about the Sources of Growth?" HIER Discussion Paper 1855. Harvard Institute for Economic Research, Cambridge, Mass.

———. 2000. *Land, Labor and Globalization in the Pre-Industrial Third World.* NBER Working Paper 7784. Cambridge, Mass.: National Bureau of Economic Research.

Wood, Adrian. 1994. *North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World.* Oxford: Clarendon Press.

———. 1997. "Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom." *World Bank Economic Review* 11(1): 33–57.

World Bank. 2000. *Global Development Finance 2000.* Washington, D.C.

———. Various years. *World Development Indicators.* Washington, D.C.

Zlotnik, Hania. 1999. "Trends of International Migration since 1965: What Existing Data Reveal." *International Migration* 37(1): 21–61.

This Page Intentionally Left Blank

Fear of Globalization:
The Human Capital Nexus

Daniel Cohen

This article builds on the Mincerian approach to human capital to document the returns to education over time and across countries. Analysis of cross-sectional macroeconomic data shows that the social returns to education do not appear to have risen in rich countries, staying around 10 percent for the past three decades. But the returns do appear to have risen in poor countries. This finding contradicts two bits of conventional wisdom: that technological progress has been biased in favor of skilled workers in rich countries and that poor countries should experience a decline in their returns to education as globalization raises the demand for low-skilled workers in the South. The article shows that, in fact, openness raises the returns to education in poor countries. It also shows that the countries that opened their economies in the past two decades experienced a significant increase in inequality, while those whose economies were already open did not.

All over the world globalization has raised fears that the market could rend the social fabric of societies. Antiglobalization protestors proclaim that "the world is not for sale." French Prime Minister Lionel Jospin tried to draw a line in the sand, saying on television that he would go for a "market economy but not for a market society." Many more quotations, from all quarters, would point to the same idea: globalization, these days, is not being warmly welcomed. And quotations from centuries past would show that fears about globalization have long been prevalent. Back in the time of the Roman Empire Pliny the Elder was already complaining about "India, China and Arabia robbing our Empire one hundred million sesterces every year."

Nor should it be surprising that complaints about globalization are heard from both the North and the South, for globalization is a shock that alters the distribution of income among agents and sectors. More surprising perhaps is that the same fears about rising inequality are often expressed in both the North and the South. The simplest model of North-South trade, such as that encompassed in, say, the Stolper-Samuelson model, would argue that globalization is potentially a vector of rising inequality in the North but of declining inequality in the South, because it drives down the demand for unskilled labor in the North and the demand for skilled labor in the South. This argument has received much attention, especially in the North. Ross Perot's characterization of the effect of the North American Free Trade Agreement (NAFTA) as a "giant sucking sound," from U.S. markets being gobbled up by Mexican production, generated a large literature—though one that more often than not has concluded for the "small hiccup" hypothesis. So the search is on for other culprits in the rising inequality in the North and South during the 1980s.

Two key candidates have emerged: skill-biased technological progress and labor market deregulation. A wealth of literature in favor of the first has developed since the pioneering paper by Katz and Murphy (1992). Similarly, a series of papers after Di Nardo, Fortin, and Lemieux (1996) has argued that labor market deregulation, such as that triggered by the decline in unions, the lowering of minimum wages, and the like, is the key factor underlying the trend toward greater inequality. To these two traditional explanations can be added a third: the rise of new organizational patterns. More than technologies alone, the way an economy organizes its factory system has a critical bearing on the way it rewards labor (see, for example, Askenazy 1999).

Do these views on globalization originating in the North have a bearing on the debate in the South? Interestingly, many papers have argued that the rise in inequality in many countries in the South can be understood in terms of the debates in the North. Robinson (2000), for example, asserted that the rising inequality in Latin America should be interpreted along the same lines as in Di Nardo, Fortin, and Lemieux. At any rate, as we shall see, the 1990s were not years of declining inequality in the South.

In this article I first compare the private returns to human capital in poor and rich countries and look at how they have evolved. The returns appear to be marginally higher in poor countries, and there are some indications that private returns to education might have declined over the past decade. Using cross-sectional macroeconomic data, I then replicate the empirical results at the aggregate level. The social returns to education do not appear to have risen in rich countries. Instead, if anything, they have risen in poor countries. This finding contradicts two bits of conventional wisdom: that technological progress has been biased in favor of skilled workers in rich countries and that poor countries should experience a decline in their returns to education as globalization raises the demand for low-skilled workers in the South.

To analyze more specifically how globalization affects the returns to education, I divide the sample of countries into two groups: open and closed. The open group experiences social returns to education that are fairly similar to the private returns,

while the closed group appears to annihilate such returns. This result lends support to the view that in closed economies education is a useful private asset to the extent that it allows a person to reach a certain position in society, but a useless social asset to the extent that the knowledge a person acquires is not used productively. In contrast, open economies, perhaps because of the pressure of markets, make good use of their members' knowledge. From this perspective, opening an economy is more than merely a change in the relative supply of and demand for existing skills—it is a change in regime that may dramatically improve the allocation of resources.

Rewards to Human Capital

The concept of human capital is a fundamental nexus in many of the debates about the effects of globalization. Does globalization raise the returns to human capital in the North and lower them in the South? Does globalization help transform growth in human capital into growth in output?

Schooling and Experience

The standard analytical approach to human capital is from the work by Mincer (1974), which analyzes human capital through two inputs: education and experience. The inputs are, in Mincer's work, two sides of the same coin—an investment purchased by an agent, before working and while working, to enhance her productivity. After a given point in time the worker stops accumulating experience, and the decline in her stock of accumulated experience progressively reduces her earnings. In its simplest form the functional equation can be written as a log quadratic functional expression:

$$\log H_{it} = a + bYS_{it} + cEXP_{it} - dEXP_{it}^2 + \varepsilon_{it},$$

where $\log H_{it}$ is the log of the human capital of a worker i at a given time t, YS_{it} is the number of years of schooling, and EXP_{it} is the experience the worker has accumulated since starting to work, usually proxied by the age of the worker minus years of schooling minus 6.

For returns to schooling, the benchmark can be taken to be the United States, where they are 9.3 percent (table 1). The returns to the first year of experience in the United States are about a third that, at 3.2 percent, and the marginal returns to experience become nil after 33 years.

Results averaged over country groups—Africa, Asia, Latin America, and the OECD—show some disparities, although they are rather small, especially when compared with differences within groups. Cross-country data would show that the greatest dispersion occurs among OECD countries, some of which achieve extremely low returns to education, such as Austria (3.9 percent) and Sweden (2.6 percent, although this result is based on data from 1981, a time that the structure of wages was very compressed). Returns in Asia, Latin America, and even Africa

Table 1. Private Returns to Schooling and First Year of Experience in Major Country Groups and the United States, Various Years, 1980s and 1990s

Country group or country	Schooling	Experience	Max EXP (years)
Africa	0.131	0.052	31.1
Asia	0.095	0.046	42.1
Latin America	0.109	0.055	39.7
OECD	0.069	0.043	31.1
United States	0.093	0.032	33.0

Note: Max EXP is the upper limit to the accumulation of experience.
Source: Bils and Klenow 2000; and author's calculations.

(except for Côte d'Ivoire, where the returns to education are abnormally high at 20 percent) are all within the range of 8.5–15.0 percent, with average returns very close to those in the United States. The returns to the first year of experience are almost always higher than those in the United States, both in absolute terms and relative to the returns to education. The ratio between returns to experience and returns to education stands closer to 1 to 2 than to the U.S. ratio of 1 to 3.

Changes in the Patterns of Reward

In a new set of studies at the World Bank François Bourguignon and associates attempt to explain how patterns of inequality in developing economies could be analyzed from a Mincerian approach to income. These studies analyze schooling, experience, and labor force participation rates to simulate how the distribution of income has shifted over the years. Here I simply stick to the Mincerian equations that they reach, putting them in the format of the standard log-linear model presented above to ease comparison. The World Bank studies include Argentina, Brazil, Indonesia, Malaysia, Mexico, and Taiwan (China). For comparison, I add France and the United States. I focus on urban male workers, since this is the group for which results are most comparable across economies.

Argentina experienced a rise in inequality in recent decades, with its Gini coefficient exploding from 0.350 in 1974 to 0.448 in 1999. This rise in inequality was due more to the rise in unemployment and residual inequality than to a change in the returns to education or experience, which appeared to remain remarkably stable (table 2). In Brazil inequality fell marginally as the returns to schooling declined but the education level increased. In Mexico the pattern was similar to that in Brazil: the returns to education fell, but in contrast to Brazil, inequality in earnings rose by 0.08 and inequality in income by 0.06. The increase in inequality resulted from a (new) convexity of the earnings function that has benefited high-income more than low-income groups.

In Indonesia inequality rose moderately (with the Gini coefficient up from 0.384 to 0.402 between 1980 and 1996), but as a result of powerful forces working in opposite directions. In particular, a rise in the average level of schooling, usually expected to be a good thing as far as inequality is concerned, had an unequalizing

Table 2. Private Returns to Schooling and First Year of Experience for Urban Male Workers, Selected Economies and Years

Economy[a]	Schooling		Experience		Years after which income declines	
	Beginning of period	End of period	Beginning of period	End of period	Beginning of period	End of period
Asia						
Indonesia[b]						
(1980, 1996)	0.084	0.049	0.072	0.0854	36.1	35.6
Malaysia[c]						
(1984, 1997)	0.0624	0.0623	—	—	—	—
Taiwan (China)						
(1979, 1994)	0.0325	0.0582	0.0694	0.0641	24.8	26.7
Latin America						
Argentina						
(1974, 1999)	0.062	0.060	0.040	0.041	50	50
Brazil[b]						
(1976, 1996)	0.134	0.101	0.075	0.085	35.7	38
Mexico[d]						
(1984, 1996)	0.099	0.078	0.071	0.064	39	32
France						
(1982, 1998)	0.073	0.079	0.042	0.043	30.5	36.1
United States						
(1984, 1996)	0.107	0.112	0.0349	0.324	35.7	28.5

— Not available.
a. Years in parentheses refer to beginning and end of period for each economy.
b. Returns to schooling refer to first five years of education.
c. Returns to secondary education are not comparable.
d. Returns to schooling refer to first 5 years of education.
Source: World Bank and Inter-American Development Bank data.

effect. The main reason is that the schooling variable is convex, so that groups that were initially better endowed typically gained more than others. In addition, residual inequality has been rising. In Malaysia inequality fell in 1984–89 (with the Gini coefficient declining from 0.486 to 0.461) and then rose in 1989–97 (with the Gini coefficient increasing to 0.499). A decline in the returns to education was among the key equalizing factors in the first period, yet had only a marginal effect in the second. Finally, in Taiwan (China) the returns to education rose from a very low value to an average level. Throughout the period inequality remained stable, with a Gini coefficient of around 0.300. The rise in the returns to education coincided with a dramatic growth in the supply of educated workers, which canceled out the unequalizing effect of the increase in these returns.

In sum, the World Bank studies portray several conflicting trends. The returns to education remained stable in Argentina and Malaysia, rose in Taiwan (China), and fell in Brazil, Indonesia, and Mexico. In the three countries where the average returns dropped, the change in inequality was the outcome of complex phenomena: the fall in the returns to secondary education (as opposed to general education, to which I refer elsewhere in the text) was accompanied by a growing convexity of the reward curve that worked in an opposite direction with respect to inequality. In

Taiwan (China) the effect on inequality from the rise in the returns to education was offset by the growth in the supply of educated workers. With Indonesia excluded, there is a mild pattern of convergence toward average returns to education of about 6 percent. The two countries that started at this level, Argentina and Malaysia, stayed there. Those above that level, Brazil and Mexico, dropped. And the economy below it, Taiwan (China), rose.

In France and the United States the returns to schooling and experience remained stable, at least early in the period. But in the United States there was then a decline in the returns to experience as captured in the quadratic terms: in the late 1990s careers seemed to end early. This may be a sign that labor force reorganization wiped out the benefits of earlier on-the-job training.

Social Returns to Human Capital

The studies discussed in the previous section are fascinating investigations of the rise and fall of private returns to human capital. But they do not shed light on the extent to which the changes in the returns to schooling and experience are a result of social factors—such as new labor market environments—or underlying macroeconomic and technological changes. Moreover, case studies do not provide a convenient means for comparing experience across countries. In this section I attempt to fill this gap by applying at the aggregate level the Mincerian approach that has been developed in the case studies.

Theoretical Benchmark

Let's start with a simple neoclassical production function following the approach of Mankiw, Romer, and Weil (1992). Production can be expressed as:

$$Q_t = A_t^{1-\alpha} K_t^{\alpha} H_t^{1-\alpha}$$

where K_t is aggregate physical capital, H_t is aggregate human capital (human capital per capita multiplied by the population), and A_t is technological progress. Physical capital is accumulated according to the usual law of motion:

$$\dot{K}_t = -dK_t + sQ_t$$

where the dot is time derivative, d is the depreciation rate of capital, and s is the savings rate. Assume that μ is the rate of growth of technological progress and that n is the rate of growth of aggregate human capital. In the steady state of this revised Solow model, one can write:

$$\left(d + n + \mu\right)K_t = sQ_t \, .$$

One can then rewrite the expression as:

$$\log Q_t = (1 - \alpha) \log A_t + \alpha \log \frac{s}{d + n + \mu} Q_t + (1 - \alpha) \log H_t$$

or equivalently:

(1) $$\log Q_t = \log A_t + \frac{\alpha}{1 - \alpha} \Big[\log s - \log (d + n + \mu) \Big] + \log H_t .$$

In the standard neoclassical case where $\alpha = 1/3$, one should then find that $\frac{\alpha}{1 - \alpha} = 0.5$. In this case the dynamics of capital accumulation should be fairly rapid, so that the steady-state assumption is not too extreme.

The critical question is how one should proxy human capital. Mankiw, Romer, and Weil (1992) have addressed this question indirectly by focusing on a presumed law of motion of human capital, under which human capital is accumulated in a manner that is perfectly collinear to the accumulation of physical capital. Specifically, they write:

$$\dot{H}_t = -dH_t + s_H Q_t$$

where d, the depreciation rate for human capital, is taken to be identical to the depreciation rate for physical capital, s_H is essentially equal to the secondary school enrollment ratio, and Q_t is total output. Mankiw, Romer, and Weil then indirectly measure human capital as the steady state of this law of motion. This formulation implies that the dynamics of per capita income do not depend on the composition of human and physical capital, an assumption that is rejected by the data (see Cohen 1996, which shows that human capital accumulation relies more on human capital than on output).

An alternative method is simply to proxy human capital by the number of years of schooling (as in Benhabib and Spiegel 1994). This method seems innocuous, but it has significant implications for the rate of growth. In this article I simply follow the Mincerian approach to human capital, which shows that a log-linear model should be favored in cases where agents optimally choose the number of years of study, an investment that pays constant returns over their lifetime. This Mincerian approach has gained preeminence in macroeconomic studies, after the work by Bils and Klenow (2000, but circulated as a working paper in 1998) and Heckman and Klenow (1997). Hall and Jones (1999), Krueger and Lindahl (2000), and Bloom and Canning (2000) have also adopted the Mincerian approach. Pritchett (forthcoming, but circulated in 1996) was among the early such formulations. In its simplest macroeconomic form the model can be written as:

(2) $$\log H_t = a + bYS_t + \varepsilon_t ,$$

where $\log H_t$ is the log of the human capital of a country at a given time t, and YS_t is the number of years of study. (I ignore the role of experience in this article.)

The Role of Education

In this section I test equations 1 and 2, drawing on a new data set developed with Marcelo Soto at the Organisation for Economic Co-operation and Development's (OECD) Development Centre. The cornerstone of our data set for high-income countries is the detailed information published by the OECD on educational attainment since the end of the 1980s, based on reports from member and nonmember countries. This information refers to the population ages 15–64, broken up into different age groups. The main advantage of the OECD's data is that they are based on a standard method across countries. Our effort has been aimed at extending the data set to missing periods and countries. (See Cohen and Soto 2001 for details and the appendix for a summary of the sources.) In what follows I distinguish high- and low-income countries (also referred to as rich and poor countries), based on the World Bank's classification.

In 2000 workers in high-income countries had an average of 12.1 years of schooling, while those in poor countries had an average of 5.7 years (table 3). Note the contrast between the average growth in schooling in poor countries and the absolute increase in 1960–2000. In relative terms there was a mild pattern of convergence, as the ratio of years of schooling in poor countries to those in rich countries increased from 1 to 4 to 1 to 2. But the absolute difference between rich and poor countries stayed essentially constant over the years: no catch-up is embodied in the accumulation of human capital.

I estimate equation 1 in levels and take log$H$ as in equation 2. Total factor productivity is proxied by the lagged urbanization rate (URBAN), regional dummy variables (one for each region), and time dummy variables. To my knowledge this specification, which simply matches that of Mankiw, Romer, and Weil (1992) and Mincer (1974), has not been tested directly. Hall and Jones (1999) and Bils and Klenow (2000) calibrate but do not directly test this regression. Krueger and Lindahl (2000) estimate only a growth version. Heckman and Klenow (1997) do not use investment.

The regressions presented throughout the article are obtained by generalized method of moments (GMM). The instruments include lagged degree of urbanization, lagged child mortality, lagged population growth, and measures of policy distortion such as the relative price of investment and the Easterly-Levine ethnicity index (which measures the degree of ethnic diversity within countries; see Easterly and Levine 1997). All lags are 10-year periods. Africa is excluded from the sample, for reasons to which I return and that have to do with the fact that the steady-state assumption for capital accumulation is not appropriate for that region. The log of

Table 3. Average Years of Schooling in the Labor Force, Selected Years, 1960–2000

Country group	1960	1970	1980	1990	2000
Rich	8.7	9.8	10.9	11.6	12.1
Poor	2.1	2.9	3.7	4.8	5.7

Source: OECD Development Centre data.

per capita income is regressed on lagged urbanization, the number of years of schooling, and a dummy variable, POOR, for all developing countries. Similar point estimates would be obtained by focusing only on poor countries or only on OECD countries; the intercept would differ, however, which is why I have incorporated the POOR dummy variable.

This regression is almost miraculous. First, as already reported in a different format by Mankiw, Romer, and Weil (1992), the coefficient of $\log \frac{s}{d+n+\mu}$ exactly fits its theoretical value, 0.5 (table 4). (In the regression I use the domestic investment rate rather than the savings rate.) Second, the returns to education, 8.4 percent, are in line with the average returns obtained from microeconomic data. Other things equal, however, poor countries experience a 45 percent gap in income relative to rich countries, which can be interpreted as a technological barrier.

Cohen and Soto (2001) present results based on a different modeling strategy. Rather than relying on the rich-poor dichotomy, the analysis splits the world into five areas and gets the following discrepancy in per capita income with respect to rich countries: –27 percent for Eastern Europe, –29 percent for the Middle East and North Africa, –52 percent for Latin America, and –66 percent for South Asia. Averaging per capita income across the poor regions, the analysis finds the same result: a 45 percent gap between poor and rich countries. Cohen and Soto (2001) also use a different instrumenting strategy, based on a beginning-of-the-century schooling variable (the paper explains why this is a useful instrument). That strategy yields results identical to those presented here (all are quite similar to the ordinary least squares results, with measurement problems more of an issue than endogeneity). I return to this critical question below.

Human Capital and Globalization

In this section I look at how the returns to human capital have evolved over the years and the extent to which the changes may be associated with globalization.

Table 4. Per Capita Income (in Log)

Variable	Results
URBAN	$1.02 \cdot 10^{-2}$
	(5.27)
POOR	–0.45
	(–3.35)
Log [INV/($d+n+\mu$)]	0.51
	(4.5)
YS	0.084
	(2.6)
R^2	0.84

Note: INV is the domestic investment rate. Time dummy variables are omitted. *t*-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Time Patterns in the Returns to Human Capital

To analyze how the returns to education have evolved over time, I first split the sample into poor countries and rich countries. I leave aside investment (that is, the effect of education on investment).

One critical feature that emerges is the decline in the returns to education in rich countries in the past two decades (table 5). This result refutes the view that skill-biased technological progress has raised the returns to education, at least at the aggregate (social) level. In contrast, in poor countries the returns to education have risen, to levels similar to those observed in OECD countries. This rejects the view that increasing globalization, if this is taken to be a characteristic of our times, has depressed the returns to education in poor countries.

I also investigate how returns to the other component of human capital, experience, have evolved over time (but do not report the results here). The returns to experience appear to have declined somewhat in the 1990s. This may be taken as an indication that reorganization of production, more than skill-biased technological progress, was the driving force behind the rise in inequality. Given the difficulty in analyzing experience, however, I now focus on education alone.

Income and Openness

To analyze more directly the relationship between globalization and human capital, I introduce two important policy variables. The first is the Sachs-Warner dummy variable (SW) (Sachs and Warner 1995). This variable assigns the value 1 to countries regarded as "open," with openness measured by many things, ranging from tariff levels to black market disequilibria. (I discuss below whether this variable is more an all-encompassing policy variable than, strictly speaking, a measure of trade openness.) The second variable, recently put forward by Forteza and Rama (2001), measures labor market rigidities (FR).

Starting with the SW variable, I rely here on the distinction between open and closed as defined in 1970. Sorting the sample countries by this variable results in a relatively balanced split: 29 countries that are open and 30 that are closed. To see how these two groups fit the model estimated above, I run the simple regression of income on education (leaving aside the role of experience for the moment), along with the other variables (urbanization and the POOR dummy variable). (I drop the

Table 5. Returns to Education, 1979 and 1997

Country group	1979	1997
Rich	0.131	0.089
	(6.4)	(3.9)
Poor	0.065	0.138
	(1.8)	(2.6)

Note: t-statistics are in parentheses. Other factors are not reported.
Source: Author's calculations based on OECD Development Centre data.

investment term here, so as to capture the full impact of human capital on income.) The regression is obtained using the same GMM estimator as before.

The regression yields strong results (table 6). While the (total) returns to education appear to be about 13 percent for the open economies, they fall to absolute insignificance for the closed economies. Thus globalization, as measured by SW, does seem to be correlated with returns to education. There is clearly a risk, however, that the SW split introduces a sample selection bias. I attempt to correct this bias with Heckman's two-step procedure, estimating a probit equation for the decision to be open. Education is strongly positively significant; the POOR dummy variable is significantly negative, though per capita income is not. When I control for the Mills ratio for each subgroup, the results are essentially identical; in fact, they are somewhat more extreme for the closed economies (where the returns to education appear to be significantly negative).

The SW variable, however, is a proxy for many other choices that have a bearing on the internal organization of an economy. Thus I add to the analysis the FR variable, again for each subgroup of countries. To do so, I interact education with the FR variable (interestingly, the level itself is not significant).

The results show that labor market rigidities do not help to explain the low returns to education in the closed countries (table 7). But such rigidities are highly significant in the open countries. So, for a country with an open trade regime, it pays to liberalize the labor market. But for an otherwise closed economy, freeing the labor market will have insignificant (and marginally negative) effects. Conversely, it pays to open to trade for a country with a flexible labor market, but not for one with a rigid labor market. This shows that the SW variable captures a specific dimension that is somewhat complementary to labor market rigidity rather than collinear to it.

Is It Really Globalization?

What does the Sachs-Warner variable actually capture—openness or something else? To address this question, I first regress openness (imports plus exports as a fraction of GDP) against the set of variables used previously (urbanization and the POOR dummy variable) and against the inverse of the square root of population (1/SRPOP), land area (LAREA), and distance to the sea (DIST) (table 8).

Table 6. Income and Education

Variable	Open countries	Closed countries
URBAN	$2.0 \cdot 10^{-3}$	$2.88 \cdot 10^{-2}$
	(2.1)	(11.4)
POOR	−0.67	−0.57
	(7.8)	(−2.5)
YS	0.127	−0.039
	(7.8)	(−1.24)
R^2	0.88	0.72

Note: t-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Table 7. Income and Education with Labor Market Rigidities

Variable	Open countries	Closed countries
URBAN	$0.44 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$
	(3.01)	(9.8)
POOR	−0.95	−0.70
	(−7.5)	(−2.5)
YS	0.11	−0.047
	(5.4)	(−1.0)
YS*FR	$-4.7 \cdot 10^{-4}$	$-5.8 \cdot 10^{-5}$
	(−2.6)	(−0.1)
R^2	0.87	0.64

Note: *t*-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Creating a variable, OPF, that is the fitted value of the above regression, I then analyze how the returns to education interact with this value along with the FR variable. The results show that openness, as predicted by the regression results in table 8, appears to play no role in the group of countries characterized as open by SW, but is highly significant in the closed countries (table 9). This suggests an interesting explanation. In the group of countries with open trade regimes, the volume of trade itself is not important, while in the group of countries with constrained trade, relaxing the binding constraint does matter for the returns to education. This can be taken as confirmation that globalization does raise the returns to education whenever trade barriers are relaxed.

The Rise in Inequality

Has openness, as captured by SW, been associated with a rise in inequality? During the 1980s and 1990s poor countries experienced a significant rise in inequality. Theil indexes such as those measured by James Galbraith (2001) increased from 5 percent to 8 percent for poor countries during that period. A rise in the returns to

Table 8. Openness

Variable	Results
URBAN	$1.78 \cdot 10^{-3}$
	(2.8)
POOR	−0.22
	(−4.3)
1/SRPOP	8.5
	(4.9)
LAREA	$-1.9 \cdot 10^{-2}$
	(−1.99)
DIST	$-1.7 \cdot 10^{-2}$
	(−2.5)
R^2	0.48

Note: *t*-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Table 9. Per Capita Income

Variable	Open countries	Closed countries
URBAN	$0.47 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$
	(2.7)	(9.2)
POOR	−0.88	$-0.34 \cdot 10^{-2}$
	(−5.6)	(−1.08)
YS	0.093	−0.07
	(4.2)	(−1.46)
YS*FR	$-7.8 \cdot 10^{-4}$	$-4 \cdot (10^{-4}$
	(−2.95)	(−0.64)
YS*OPF	0.056	0.17
	(1.14)	(2.6)
R^2	0.87	0.69

Note: Africa is excluded. Time dummy variables are omitted. *t*-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

education is likely to generate an increase in inequality if there is no corresponding increase in years of education for the least educated group. Other things equal, a rise in the returns to education means that the existing distribution of educational achievements is multiplied by a larger number. More specifically, the income of individual i in country s can be expressed as:

$$\log y_{is} = a + bYS_{is} + \varepsilon_{is}$$

where *YS* represents, as before, the number of years of schooling of individual i and the residual includes all forms of correction, whether resulting from redistribution or other factors. One can draw from this equation that

$$\mathrm{var}_s\left(\log y_{is}\right) = b^2 \, \mathrm{var}_s\left(YS_{is}\right) + \mathrm{var}\left(\varepsilon_{is}\right) + 2b \, \mathrm{cov}\left(YS_{is}, \varepsilon_{is}\right).$$

After some manipulation, the partial derivative of the variance in income with respect to the returns to education can be derived as

$$\partial \, \mathrm{var}_s\left(\log y_{is}\right) / \partial b = 2 \, \mathrm{cov}\left(YS_{is}, \log y_{is}\right).$$

This is likely to be positive, unless the private returns to education are totally at odds with the social returns.

To investigate whether openness has been associated with a rise in inequality, I start with a simple cross-country analysis of inequality in the 1990s. I analyze the role of education, income (LRGDP), income squared (LRGDP²), the POOR dummy variable, the SW dummy variable, and a new dummy variable, MOVER, which takes 1 if a country was described as closed in the 1970s and open in the 1990s. The data are from the database of James Galbraith (2001). The index is a Theil inequality measure drawn from national surveys.

The results are telling (table 10, regression 1). Neither urbanization nor being poor or rich seems to matter much, nor does the SW variable. However, the results yield a Kuznets curve that describes an inverted U-shaped curve with income. Years

of schooling have a negative effect on inequality, although this effect is not significant at the 5 percent threshold. An important finding is that the countries that shifted from being closed to being open experienced a significant increase in inequality.

Ideally, one would want to correlate the rise in inequality with a measure of the covariance term that appears in the equation. This term is itself likely to be correlated with the variance in schooling. Lacking such data, I can take as an assumption that a country experiences greater inequality at low levels of education than at high levels. I then simply interact the MOVER variable with the level of education (table 10, regression 2). The results show that all the action comes from this interaction term. The higher the education level is in a country, the less important is the effect on inequality of switching from being closed to being open.

To investigate this relationship further, I estimate the same equation in first differences for the poor countries only (table 10, regression 3). The same result emerges: countries that liberalized trade experienced an increase in inequality, but this increase was smaller if the country was initially highly educated.

The fact that countries that were already open did not experience rising inequality in the past two decades, while those that changed their trade regime did, can be interpreted from many different angles. Following Banerjee and Duflo (2000), one could argue that a change in either direction is associated with a rise in inequality. But the fact that the outcome is mitigated by the level of education points to a specific channel: countries with a low level of education have been more prone to experience an increase in inequality resulting from rising returns to education. One potential explanation is that higher education is correlated with lower variance in educational attainment. Another is that it is a better predictor of an increase in educational attainment. It is not possible to go much beyond these conjectures at this stage.

Table 10. Inequality

Variable	Regression 1	Regression 2	Regression 3 (variation, poor countries)
URBAN	$-3.8 \cdot 10^{-5}$	$-1.3 \cdot 10^{-4}$	$-4.7 \cdot 10^{-5}$
	(−0.07)	(−0.28)	(−0.08)
POOR	−0.05	$-1.1 \cdot 10^{-2}$	
	(−1.50)	(−0.40)	
LRGDP	0.065	0.045	
	(2.93)	(2.32)	
$LRGDP^2$	$-0.5 \cdot 10^{-2}$	$-0.48 \cdot 10^{-2}$	
	(−2.27)	(−2.2)	
SW	$0.94 \cdot 10^{-2}$	$-0.3 \cdot 10^{-3}$	−0.02
	(0.40)	(−0.17)	(−1.03)
MOVER	0.068	0.25	0.19
	(3.2)		(4.1)
YS	$-0.7 \cdot 10^{-2}$	$0.35 \cdot 10^{-2}$	
		(0.70)	
MOVER*YS		$-2.9 \cdot 10^{-2}$	$-2.0 \cdot 10^{-2}$
		(−4.35)	(−2.70)
R^2	0.45	0.62	0.49

Note: t-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

The Convergence Debate Revisited

Now let's see how much bearing the previous discussion has on the impact of globalization on the distribution of income across countries.

Growth and Education

As a first test of the robustness of the results obtained above, I run the previous equations in first differences. I focus here on the role of education to make the results comparable with those of previous studies. To directly address the potential endogeneity of the right-hand-side variables, I lag the growth of human capital (the increase in the number of years of schooling, or GS) by one decade and estimate the growth equations by ordinary least squares (regression 1, table 11). I find a fit perfectly in line with the returns to education found before, with average returns to education of about 10 percent.

Why have I obtained this regression while other researchers have apparently denied its existence? (See de la Fuente and Doménech 2000 and Cohen and Soto 2001.) First, I chose a log-linear specification, along the lines of the Mincerian approach, while many formulations have used a log-log specification. With my specification, a log-log specification would not be statistically significant. Indeed, many poor countries start with little education. Growth rates of education may therefore appear to be very high even though the actual numbers remain quite small.

Second, there is a difference in the quality of data. The importance of this is emphasized by de la Fuente and Doménech (2000), who show that the results are quite sensitive to the quality of the data. First differences of growth rates are quite badly measured. The data set I use builds on improvements made possible by the work of many researchers and by a specific effort at the OECD and the OECD Development Centre. With Barro-Lee (1993) estimates such as those available in 1996, for example, the results would go nowhere.

Table 11. Growth of Per Capita Income

Variable	Regression 1	Regression 2	Regression 3 (variation, poor countries)
URBAN	$-2.63 \cdot 10^{-4}$	$-2.07 \cdot 10^{-4}$	$-2.0 \cdot 10^{-4}$
	(-2.99)	(-2.39)	(-2.41)
POOR	$-1.09 \cdot 10^{-2}$	$1.6 \cdot 10^{-3}$	$-5 \cdot 10^{-3}$
	(-2.74)	(0.3)	(-1.05)
GS(−1)	0.104	$-2 \cdot 10^{-2}$	0.079
	(2.64)	(-0.40)	(1.37)
GS(−1)*SW		0.146	0.08
		(3.98)	(2.11)
GS(−1)*FR			$-1.7 \cdot 10^{-3}$
			(-3.0)
R^2	0.20	0.26	0.32

Note: Africa is excluded. Time dummy variables are omitted. The number of observations is 161. *t*-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Table 12. Income Growth

Variable	Results
URBAN	$-1.88 \cdot 10^{-4}$
	(-2.2)
POOR	$-3 \cdot 10^{-3}$
	(-0.6)
GS	$-4 \cdot 10^{-2}$
	(-0.8)
GS*SW	$19 \cdot 10^{-2}$
	(4.42)
GS*RIG*SW	$-15 \cdot 10^{-2}$
	(-2.6)
R^2	0.29

Note: RIG = 1 if the FR (labor market rigidity) index for a country is above the mean for its reference group, and 0 otherwise. *t*-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Finally, perhaps the most substantial difference originates from my inclusion of the POOR dummy variable, which follows my favored specification in the estimates of the previous section. I now try to extend our understanding of this variable.

I begin by adding as an explanatory variable the interacted term SW with the growth rate of education (interestingly, the variable alone is not significant). The results are striking: the point estimate of the growth rate of education alone falls to insignificance, and all the variation is captured by the interaction term that goes with the SW variable. The POOR dummy variable also falls into insignificance. These results suggest two facts. First, the ability to channel growth in education into growth in output seems to be highly correlated with a set of policy measures proxied by the trade liberalization dummy variable. And second, this phenomenon seems to explain the difficulties of poor countries.

Now I add the other measure of policy distortion, the labor market rigidity variable (FR). The results are in line with those obtained in the previous section. When I run the regressions separately for open and closed countries (not reported here), I find that growth in the open subgroup is highly correlated with the growth of education and with its interaction with the labor market rigidity index. In contrast, for the closed economies the results are essentially negative: the returns to schooling are not significant, nor is the labor market term. Interestingly, the mirror exercise of splitting the countries into two groups, one with low labor market rigidities and one with high rigidities, and controlling for the residual explanatory power of the SW variable yields the same type of results. Merging these results, I run a single regression in

Table 13. Returns to Growth in Education

Nature of labor market	Open countries	Closed countries
Flexible	0.15	−0.04
Rigid	0	−0.005

Source: Author's calculations based on OECD Development Centre data.

which both policy variables are interacted. More specifically, I discretize the FR variable, using 1 when a country is above the mean for its reference group (rich or poor).

The results capture the features obtained above: the SW variable is highly significant in the group with flexible labor markets, while it is not in the group with rigid labor markets (table 12). Moreover, the POOR dummy variable is insignificant. These results, summarized in table 13, confirm my previous results and point to strong complementarities between internal and external policy choices.

Total Factor Productivity

Now let's focus on the rate of growth of technological progress. From the production function (equation 1), I can simply define the Solow residual as

$$Residual = \log Q / L - 0.5 * LINV - 0.10 * YS$$

where *LINV* is the log of investment.

The results show that while total factor productivity has been on the rise in the OECD countries, it has stagnated in the poor countries, even when Africa is excluded (table 14). Only the Asian countries appear to have registered an increase—about 25 percent over the past two decades. To analyze the dynamics behind these patterns, I regress the rate of growth of the residual on its initial value and on the factors that I analyzed above.

The results reveal a pattern of convergence similar to that found in many previous studies of convergence, with a "β-convergence" of 2.1 percent (table 15). Two specific features should be acknowledged, however. First, I control for the POOR dummy variable, which accounts for all developing economies. Second, I obtain this conditional convergence while controlling for essentially nothing else, neither school enrollment nor any policy variables. When I run the same regression with the growth rate of GDP as a dependent variable and initial income on the right-hand side, the regression goes nowhere (not reported here).

Many studies have attempted to demonstrate that key policy mismanagement problems are behind the failure of poor countries to catch up to the rich. Drawing on a key study by Easterly and Levine (1997), but again leaving aside Africa, I incorporate two other determinants of policy mismanagement in addition to the SW

Table 14. Total Factor Productivity, 1979 and 1997

Variable	1979	1997
OECD	8.47	8.96
POOR without Africa	8.02	7.96
AFRICA	7.93	7.40
ASIA	7.66	7.91
LATINCA	8.21	8.00

Source: Author's calculations based on OECD Development Centre data.

dummy variable for trade liberalization and the FR variable for labor market rigidity: ethnic diversity (EL, from Easterly and Levine 1997) and distortion of investment measured by the price of investment relative to that in the United States (PISH, from Summers and Heston 1991).

These variables are reported in the regression but do not add much by themselves. To further explore the role of the SW variable, I run the same regression for the subsample of countries that liberalized and for the subsample that did not.

OPEN ECONOMIES. Two striking findings emerge for the open economies (table 16). One is that the speed of convergence is now about twice as fast as the average. The other is that the steady state remains virtually unchanged, since the point estimate of the POOR dummy variable is about twice as large. At this stage one would then argue that economic integration appears to speed the convergence of productivity, which can be good for the growth of poorly endowed countries, yet seems to have little effect on the final outcome. But as we shall see, things are not quite so simple. To look further into what might cause the discrepancy in the POOR dummy variable, I interact that variable with the policy variables. One variable turns out to be highly significant: the price distortion variable PISH.

The fit improves significantly, with about 10 percent of variance additionally explained. The POOR dummy variable becomes insignificant (the point estimate actually becomes positive), with all the action carried out by the interaction term. Thus I can tentatively conclude that the ability to converge rests on the ability to create the proper incentives to invest, once the other policy measures encompassed in SW are taken care of.

Table 15. Growth of Total Factor Productivity

Variable	Regression 1	Regression 2
URBAN	$2 \cdot 10^{-3}$	0
	(1.4)	(0.31)
POOR	$-3.6 \cdot 10^{-2}$	$-3.5 \cdot 10^{-2}$
	(−4.8)	(−3.5)
LATINCA	$-0.9 \cdot 10^{-2}$	$-0.3 \cdot 10^{-2}$
	(−1.51)	(−0.5)
Residual	$-2.1 \cdot 10^{-2}$	$-2.1 \cdot 10^{-2}$
	(−3.01)	(−2.6)
SW		$0.6 \cdot 10^{-2}$
		(0.8)
FR		$-0.0 \cdot 10^{-2}$
		(−0.04)
EL		$-0.4 \cdot 10^{-2}$
		(−0.41)
PISH		$0.4 \cdot 10^{-2}$
		(0.5)
R^2	0.40	0.40

Note: Africa is excluded. Time dummy variables are omitted. The number of observations is 161. *t*-statistics are in parentheses. Residual is beginning of period total factor productivity.
Source: Author's calculations based on OECD Development Centre data.

CLOSED ECONOMIES. Now let's focus on the pattern of productivity growth in the closed economies. My previous analysis suggested that schooling was somewhat wasted in these countries and did not show up as a determinant of growth. So I try another definition of the residual:

$$\text{Residual } B = \log Q / L = 0.5 * LINV$$

which does not take account of schooling. The results show that although convergence is slower in closed than in open economies, it is still fairly rapid, at about 3.2 percent (table 17). For the closed economies, the POOR dummy variable and the interaction term with the price of investment are hardly significant.

The lesson seems to be that distortion of investment is less important in closed economies. The results also suggest that the claim that open economies converge more rapidly is somewhat exaggerated: closed economies also converge rapidly to a lower steady state. It is when the two groups are merged that the picture becomes blurred.

I also report the regression obtained by conditioning on productivity as defined for the open economies, that is, net of school attainment. The results show that productivity so defined has no predictive power. Symmetrically, when I run a similar regression for the open economies (not reported here), the opposite happens (residual B has no predictive power, $t = 0.2$), providing additional confirmation that productivity is properly defined for each subgroup.

Thus this exercise confirms that proper use of human capital is a critical factor in the analysis of the effect of globalization. I do not find, however, that human capital is enough to break the technological barrier between rich and poor countries. It also takes the ability to increase the incentive to invest (see Rodrik and Rodríguez 2000 for a similar conclusion).

Table 16. Growth of Total Factor Productivity in Open Economies

Variable	Regression 1	Regression 2
POOR	$-5.4 \cdot 10^{-2}$	2.24
	(5.3)	(1.22)
LATINCA	$0.5 \cdot 10^{-3}$	0.0
	(0.46)	(0.0)
EL	$0.8 \cdot 10^{-2}$	$-1.3 \cdot 10^{-2}$
	(0.8)	(1.22)
FR	$1.1 \cdot 10^{-4}$	$1.4 \cdot 10^{-5}$
	(0.98)	(0.13)
PISH	$0.55 \cdot 10^{2}$	$2.98 \cdot 10^{-2}$
	(0.33)	(1.90)
PISH*POOR		-0.11
		(-4.6)
Residual	$-4.5 \cdot 10^{-2}$	$-3.8 \cdot 10^{-2}$
	(-4.7)	(-4.4)
R^2	0.46	0.57

Note: t-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Table 17. Growth of Total Factor Productivity in Closed Economies

Variable	Regression 1	Regression 2	Regression 3
URBAN	$0.9 \cdot 10^{-3}$	$0.9 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$
	(3.8)	(3.9)	(3.7)
POOR	$-2.2 \cdot 10^{-2}$	$-1.3 \cdot 10^{-2}$	$-2.7 \cdot 10^{-2}$
	(−1.6)	(0.88)	(−1.4)
Residual B	$-3.2 \cdot 10^{-2}$	$-3.2 \cdot 10^{-2}$	$-4.1 \cdot 10^{-2}$
	(−3.92)	(−3.92)	(−1.91)
POOR*PISH		$-0.9 \cdot 10^{-2}$	
		(−1.15)	
Residual			$0.9 \cdot 10^{-2}$
			(0.42)
R^2	0.53	0.54	0.55

Note: t-statistics are in parentheses. Residual is beginning of total factor productivity.
Source: Author's calculations based on OECD Development Centre data.

The Case of Africa

Now let's look briefly at the case of Africa, which had to be dropped from the sample primarily because of its much lower capital-output ratios. To take a first glance at the data, I write the following simple decomposition:

$$Q/L = \frac{Q}{K} \cdot \frac{K}{L}.$$

Using data for a number of countries from the Summers-Heston database, I find that African countries appear to be very different from the others. The capital-output ratios for African countries are a third to a half that of the United States, while Latin American ratios are about the same as the U.S. ratio and Asian ratios are a bit lower (table 18).

The approximation according to which capital could be thought of as in its steady state could not be taken without reservation for Africa, unless capital is significantly less productive in Africa than elsewhere (as Easterly and Levine [1997] suggest). I nevertheless fit my model to the case of Africa, without taking account of investment. I first regress income on education and then interact education and openness as for closed economies.

Africa appears to exhibit large returns to education, and these returns appear to be driven essentially by exposure to trade (table 19). From this perspective, African countries do not appear to differ too much from the others.

Conclusion

Building on macroeconometric evidence, I have found several results that go against many of the conventional ideas about globalization and technological progress. To start with, I have found no evidence of rising returns to education in rich countries. If anything, it appears instead that the returns to experience have declined, suggesting a change in the organization of production rather than skill-biased technological progress. Among poor countries I have found a stark contrast between the

Table 18. Capital and Output

Economy	Q/L	Q/K	K/L	K/Q
Africa				
Kenya	1.9	2.3	0.8	0.4
Mauritius	10.2	2.3	4.4	0.4
Nigeria	2.1	2.8	0.7	0.35
Sierra Leone	2.5	1.0	0.2	0.09
Zambia	2.1	1.5	1.35	0.66
Zimbabwe	2.4	0.7	3.4	1.4
Asia				
Hong Kong (China)	22.8	1.6	14.0	0.6
India	3.2	1.6	2.0	0.6
Malaysia	1.2	2.9	0.4	0.35
Philippines	4.8	1.33	3.6	0.75
Thailand	6.7	1.15	5.8	0.87
Latin America				
Bolivia	5.3	1.0	5.3	1.0
Chile	11.8	1.0	11.3	1.0
Colombia	10.1	0.8	12.7	1.3
Mexico	17.0	1.2	13.7	0.8
Peru	6.8	0.8	6.5	1.2
Venezuela, República Bolivariana de	17.4	0.95	18.3	1.05
OECD				
France	30.3	0.8	37.5	1.2
Germany	30.1	0.1	41.2	1.4
Japan	22.6	0.5	41.2	1.8
United States	36.7	1.7	36.0	1.0

Source: Summers and Heston 1991.

returns to education in open economies and those in closed economies. In countries where the constraints to trade are binding, trade appears to raise the returns to education, while those returns appear to be uniformly high in open countries. I take this as confirmation of the counterintuitive view that in poor countries international trade raises the returns to education rather than reduces them. This finding contradicts the naive version of the North-South trade model in which the returns to skills are depressed in poor trading economies. In the countries that changed their trade regimes, I have found that inequality rose, more so in the countries that initially had low levels of education.

Table 19. Income Growth in Africa

Variable	Regression 1	Regression 2
URBAN	$2.1 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$
	(2.07)	(1.9)
EDUC	0.16	−0.11
	(5.6)	(−1.8)
Y*OPF		0.78
		(5.8)
R^2	0.70	0.56

Note: t-statistics are in parentheses.
Source: Author's calculations based on OECD Development Centre data.

Although openness appears to be critical to the ability of a country to build on its human capital base, it is not the only factor. Domestic labor markets also appear to play an important role. The two factors appear to be intertwined: it takes both internal and external openness to build on human capital accumulation. When the analysis is plugged into the pattern of inequality, countries whose economies are already open do not appear to be more unequal, yet countries that are opening their economies appear to experience rising inequality—all the more so if their initial endowment in education is low.

Being open or closed does not in itself appear to help bridge the gap between rich and poor countries, however. The returns to education appear to converge between rich countries and open countries, but this does not narrow the technological gap. The extent to which this gap is bridged seems to be highly correlated with the distortion of investment. More work is clearly needed to analyze the extent to which such distortion is due to poor internal credit markets. Altogether, the findings suggest that globalization is still full of risks, promises, and mysteries.

Data Appendix

In my work with Marcelo Soto, we have built a database drawing from three main sources: the OECD database on education, national censuses or surveys published in UNESCO's *Statistical Yearbook,* and censuses obtained directly from the Web sites of national statistical agencies.

To fill the gaps in the data available in the OECD database, we first split the population into five-year age groups (15–19, 20–24, and so on) for each of the years 1960, 1970, 1980, 1990, and 2000 from the United Nations population statistics database; we also include 2010 estimates from a forecast of the U.S. Bureau of the Census. We then estimate school attainment in each age group using an OECD, national, or UNESCO census wherever such a census exists. When no such census is available for the period being considered, but a later one is, we extrapolate backward all relevant information from the latest census by making the assumption that the school attainment of the population age T in one census is the same as that of the population age $T - 10$ in the census performed 10 years earlier.

For the data still missing from such backward computations, we extrapolate, whenever possible, the data available from an earlier census. Consider the case of a country for which no direct information exists for the age group 60–64 in 1980. If possible, we first try to extract the information from the 1990 census by considering the age group 70–74 in 1990. If this information is unavailable, we then try to extract the information from the 1970 census by considering the age group 50–54. When no relevant census exists, we rely on school enrollment data to fill in the missing information. To take the same example, if the entrance age for primary school is six, the age group 60–64 in 1980 was at the age to start primary education between 1922 and 1926. By calculating the ratio of new entrants in first grade to the population of six-year-olds—the net intake rate—in, say, 1924, one can estimate the share of the population ages 60–64 in 1980 that attended primary school. The

same procedure provides an estimate of the fraction of each age group that went through each education level for which no census information is available.

Several sources are used to determine the net intake rate. The main source is Mitchell (1993), who has published long series on primary, secondary, and tertiary enrollment for most countries, starting in the second half of the 19th century. This information is combined with data from UNESCO's *Statistical Yearbook,* which publishes systematic data on enrollment at different levels of education since 1950. The two sources generally coincide, but when important differences arise UNESCO data are used. Population tables by age are taken from Mitchell, the United Nations *Demographic Yearbook,* the U.S. Bureau of the Census, and national agencies.

Other authors (see Nehru, Swanson, and Dubey 1995) have already used Mitchell's series to build education indexes but have been criticized because they do not use census data. As a result, some of their country indexes bear little relationship to data taken directly from censuses. Moreover, de la Fuente and Doménech (2000) have noted some implausible results in Nehru, Swanson, and Dubey's (1995) database. In 1960, for example, Ireland's population is given 14 years of average schooling. Since most studies (including Nehru, Swanson, and Dubey's) assign less than 14 years to most countries with high levels of education in 1990, this figure must be an error. One important difference between Nehru, Swanson, and Dubey's approach and ours is that we use Mitchell's data only to fill missing cells in existing data rather than for the entire database. The only region for which we rely primarily on Mitchell's data is Africa, one reason that the region is dropped from the econometric analysis.

Two assumptions underlie the use of censuses to infer educational attainment before and after a census. First, it is assumed that the mortality rate is distributed homogeneously within each age group, independent of the education level of those in that age group. Although it can be argued that mortality rates are lower among more educated people than among less educated ones, the error introduced by the assumption of "death homogeneity" must be of the second order.

Second, and more troublesome, the methodology assumes that immigrants have the same education level as the corresponding age group in the host country. If this is not the case, and assuming that the host country's population is on average more educated than immigrants, the education level for the years prior to immigration will be understated if the immigration takes place before the census. An additional bias is introduced when net intake rates rather than census data are used to compute the education level. Since the historically observed intake rates are used to compute current education levels for some age groups, immigration by people with relatively low education will induce an overstatement of the education level for those age groups.

Similar arguments may be applied to countries with important flows of emigrants, though in these cases the distortions are arguably lower. The reason is that emigrants plausibly have an education level close to that of their compatriots. In any case the lack of information on the education level of migrants makes it impossible to take the effects of migration into account.

Bibliography and References

Alatas, Vivi, and François Bourguignon. 1999. "The Evolution of the Distribution of Income During Indonesian Fast Growth: 1980–1996." Research project on Microeconomics of Income Distribution Dynamics. World Bank and Inter-American Development Bank, Washington, D.C.

Askenazy, Philippe. 1999. "Neo-Stakhanovism." Centre d'Études Prospectives d'Économie Mathématique Appliquées à la Planification (Cepremap), Paris.

Banerjee, Abhijit V., and Esther Duflo. 2000. *Inequality and Growth: What Can the Data Say?* NBER Working Paper 7793. Cambridge, Mass.: National Bureau of Economic Research.

Barro, Robert, and Jong-Wha Lee. 1993. "International Comparisons of Educational Attainment." *Journal of Monetary Economics* 32(3).

Benhabib, Jess, and Mark M. Spiegel. 1994. "The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data." *Journal of Monetary Economics* 34: 143–73.

Bils, Mark, and Peter Klenow. 2000. "Does Schooling Cause Growth?" *American Economic Review* 90(5): 1160–83.

Bloom, David, and David Canning. 2000. "Health, Human Capital and Economic Growth." Commission on Macroeconomics and Health, Geneva.

Bourguignon, François, Marc Fournier, and Marc Gurgand. 1999. "Distribution, Development and Education in Taiwan, 1979–1994." Research project on Microeconomics of Income Distribution Dynamics. World Bank and Inter-American Development Bank, Washington, D.C.

Coe, David, Elhanan Helpman, and Alexander Hoffmaister. 1997. "North-South R&D Spillovers." *Economic Journal* 107: 134–49.

Cohen, Daniel. 1996. "Tests of the Convergence Hypothesis: Some Further Results." *Journal of Economic Growth* 1(3): 351–61.

Cohen, Daniel, and Marcelo Soto. 2001. *Human Capital and Growth: Good Data, Good Results.* Paris: OECD Development Centre.

de la Fuente, Angel, and Rafael Doménech. 2000. "Human Capital in Growth Regressions: How Much Difference Does Quality Data Make?" CEPR Discussion Paper 2466. Centre for Economic Policy Research, London.

Di Nardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages." *Econometrica* 64(5): 1001–44.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112(4): 1203–50.

Ferreira, Francisco H. G., and Ricardo Paes de Barros. 1999. "The Slippery Slope: Explaining the Increase in Extreme Poverty in Urban Brazil, 1976–1996." World Bank, Poverty Reduction and Economic Management Network, Poverty Division, Washington, D.C.

Forteza, Alvaro, and Martín Rama. 2001. "Labor Market Rigidity and the Success of Economic Reforms Across More Than 100 Countries." Policy Research Working Paper 2521. World Bank, Development Research Group, Washington, D.C.

Galbraith, James. 2001. University of Texas Inequality Project. [utip.gov.utexas.edu].

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics* 114(1): 84–116.

Heckman, James J., and Peter J. Klenow. 1997. "Human Capital Policy." University of Chicago, Department of Economics.

Katz, Lawrence, and Kevin Murphy. 1992. "Changes in Relative Wages, 1963–87: Supply and Demand Factors." *Quarterly Journal of Economics* 107(1): 35–78.

Krueger, Alan B., and Mikael Lindahl. 2000. *Education for Growth: Why and for Whom?* NBER Working Paper 7591. Cambridge, Mass.: National Bureau of Economic Research.

Legovini, Arianna, César Bouillon, and Nora Lustig. 1999. "Can Education Explain Income Inequality Changes in Mexico?" Research project on Microeconomics of Income Distribution Dynamics. World Bank and Inter-American Development Bank, Washington, D.C.

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107(May): 402–37.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings.* New York: Columbia University Press.

Mitchell, B. R. 1993. *International Historical Statistics: Africa, Asia and Oceania, 1750–1988.* New York: M. Stockton Press.

———. 1998a. *International Historical Statistics: Europe, 1750–1993.* New York: M. Stockton Press.

———. 1998b. *International Historical Statistics: The Americas, 1750–1993.* New York: M. Stockton Press.

Nehru, Vikram, Eric Swanson, and Ashutosh Dubey. 1995. "A New Database on Human Capital Stocks in Developing and Industrial Countries: Sources, Methodology and Results." *Journal of Development Economics* 46(2): 379–401.

OECD (Organisation for Economic Co-operation and Development). 2000. *Investing in Education: Analysis of the 1999 World Education Indicators.* Paris.

Pritchett, Lant. Forthcoming. "Where Has All the Education Gone?" *World Bank Economic Review.*

Robinson, James A. 2000. "Where Does Inequality Come From? Ideas and Implications for Latin America." Paper presented at OECD Development Centre conference on the Effects of Globalization, Paris, 30 November–1 December.

Rodrik, Dani, and Francisco Rodríguez. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." Harvard University, John F. Kennedy School of Government, Cambridge, Mass.

Sachs, Jeffrey, and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1, pp. 1–95.

Summers, Robert, and Alan Heston. 1991. "The Penn Mark IV Data Table." *Quarterly Journal of Economics* 106(2): 327–68.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). Various years. *Statistical Yearbook.* Paris.

United Nations. Various years. *Demographic Yearbook.* New York.

Comment on "Globalization and Inequality: Historical Trends," by Kevin H. O'Rourke, and "Fear of Globalization: The Human Capital Nexus," by Daniel Cohen

Juan-Luis Londoño

The articles by Kevin O'Rourke and Daniel Cohen are fascinating and full of valuable ideas that will enrich the discussions about globalization and inequality in Latin America. The two articles are also quite complementary. While Cohen looks at the effects of trade on the rewards to human capital, O'Rourke considers the effects of the integration of goods and factor markets on the rewards to all factors of production. While Cohen estimates small samples using cross-sectional data for the past two decades, O'Rourke takes a historical perspective on two centuries through simulations of general equilibrium models. But both focus on capturing the distributive effects of globalization through the evolution of returns to factors of production.

General Observations

Cohen brings simple messages: Education leads to growth in the countries that have maintained open trade systems and flexible labor markets. In the countries that have not met those conditions, education has had little effect. Greater trade openness is associated with greater inequality only in the countries in which capital has been scarce, education limited, and the incentive to invest small.

O'Rourke bears a more complex and challenging message: The globalization of the past two decades is not so new as is commonly believed, nor is it more intense than the globalization that took place in the 19th century. The relationship between globalization and inequality is not simple or straightforward, since globalization has many dimensions, each of which has many possible links with inequality. Rather than average relationships, what needs to be sought is the association between the inequality in a country, its resource endowments, and the movements of goods, workers, or capital. O'Rourke finds an important difference between the 19th and 20th centuries: while in the 19th century the changes in distribution within countries were

Juan-Luis Londoño is professor of economics at the Universidad de los Andes and director of the journal *Revista Dinero*.

greater, in the 20th century there was more convergence among countries. And in this convergence the integration of factor markets—through migration in the 19th century and capital movements in the 20th—has had as significant an impact as trade.

Both articles leave some loose ends—some hypotheses that are not fully checked out. Cohen, in suggesting that the returns to human capital in the developed countries have fallen with globalization—a hypothesis that runs counter to much of the North American literature of the 1990s—rules out the possibility of a technological change involving more intensive use of human capital. Cohen leans instead toward a huge reorganization of production, whose characteristics he has not properly documented.

O'Rourke, citing a recent study by Lucas (2000), goes so far as to suggest that in the 1990s humanity passed through a new "Kuznets-Lucas" turning point at which globalization would happen only if accompanied by greater equality and convergence among countries. I do not believe that there is as yet any empirically justified basis for this optimism.

Implications for Latin America

The two articles provide a general framework for discussing the possible links between globalization and inequality in the postwar period for the particular case of Latin America. To do so, I make intensive use of a study written with two coauthors (Londoño, Székely, and Spilimbergo forthcoming).

What Explains Inequality and Trade in General?

In a comparison of the world's regions, no direct connection is apparent between their degree of openness to trade and their measure of income inequality (Gini coefficient). Countries in Asia have less inequality than countries in other regions, although those in East Asia are more open and those in South Asia less so. Europe has the same relative intensity of trade as Africa, but its Gini coefficient is only about half that of Africa (figure 1).

O'Rourke and Cohen invite us to examine this apparent national and regional diversity as it relates to the development of resources. As Cohen wonders in his article, will we find that Latin America is undercapitalized and that it has accumulated human and physical capital more slowly than expected?

The relationship between trade, globalization, and income inequality can be examined empirically using a database that Miguel Székely, Antonio Spilimbergo, and I have built, with more than 3,000 observations for 1960–95. Using this database, we have established firm empirical relationships between inequality and globalization through the resource endowments of economies.

The intensity of trade, in the midst of a growth trend, is inversely related to the size of a country, in terms of both area and population, and to its distance from developed country markets (table 1). These economic geography variables explain a

Figure 1. Inequality and Trade

*Intensity of trade
(exports + imports
as percentage of GDP)*
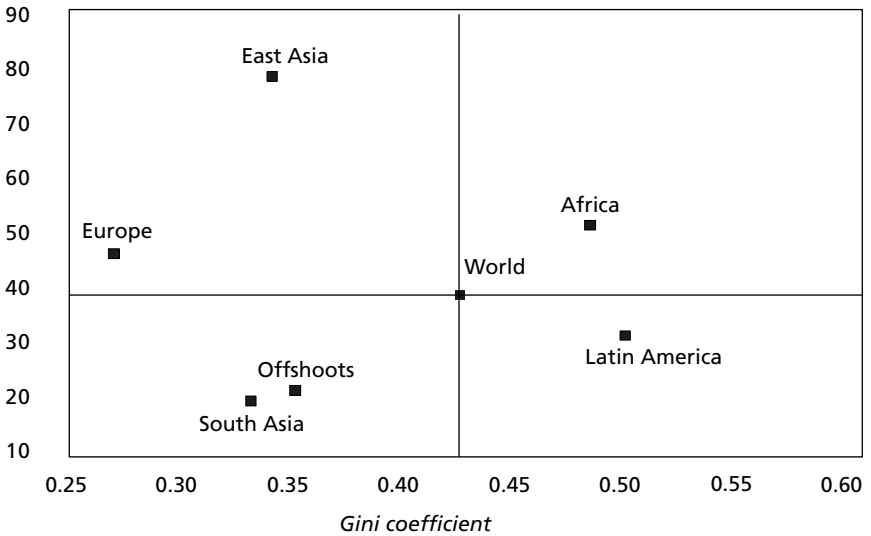
Note: Data refer to the mid-1990s. *Offshoots* refers to land-rich countries such as Australia, Canada, and the United States.
Source: Londoño, Székely, and Spilimbergo forthcoming.

third of the variance in trade. The land, human capital, capital, and labor endowments of each country relative to the world average explain another third: greater intensity of any factor of production tends to be associated with greater trade, as in any Ricardian model.

Income inequality, beyond being correlated with the intensity of trade (as is usual in the literature) or the level of income (as in the Kuznetsian variants), appears to have a specific regional component closely linked with the relative intensity of factors (table 2). The countries with greater abundance of primary resources, such as land or unskilled labor, tend to have greater inequality, while those better endowed with accumulable resources, such as human and physical capital, tend to have less.

In Latin America income inequality is pronounced and has remained relatively stable rather than declining (figure 2). How can this be, given the region's relative resource endowment?

Latin America has a greater abundance of the resources associated with higher inequality, and less of the accumulable factors associated with lower inequality. Compared with the world average, it has more natural resources but much less human capital—a situation that tends to be associated with greater inequality (figure 3). But its relative resource differences are smaller than those of other regions; indeed, its resource portfolio appears to be well in line with the world average. This

Table 1. Trade, Geography, and Resource Portfolio, 1960–95
(dependent variable: trade intensity)

Variable	Economic geography		Relative factor endowments	
	(1)	*(2)*	*(3)*	*(4)*
Constant	155	–1,163	–1,765	–1,872
	(29.4)	(13.5)	(10.3)	(10.8)
Area	–6.0	–5.4	–7.3	–6.8
	(23.3)	(21.9)	(15.0)	(14.9)
Size	–2.9	–4.1	–5.1	–6.8
	(10.0)	(14.3)	(8.9)	(11.4)
Distance	–12.2	–13.9	–16.4	–15.1
	(12.6)	(15.2)	(12.6)	(10.5)
Time		0.7	1.0	1.1
		(15.3)	(11.5)	(12.0)
Land				2.5
				(5.1)
Human capital				19.0
				(6.2)
Capital				5.5
				(3.8)
Unskilled labor				1.6
				(1.6)
Factor disequilibrium			3.4	
			(8.3)	
Adjusted R^2	0.35	0.37	0.71	0.72
F-test	460	440	243	188
N	2,930	2,930	436	436
Method	Huber correction	Huber correction	Huber correction	Huber correction

Note: t-statistics are in parentheses. Area is in square kilometers, size in millions of people, and distance (from developed country markets) in kilometers (all in natural logarithms). Panel data analysis was based on Huber correction method, using Stata.
Source: Londoño, Székely, and Spilimbergo forthcoming.

relative equilibrium in resource endowments tends to be associated with a lower intensity of trade.

The Case of Latin America

Contrary to what O'Rourke believes, Latin America has not been characterized by a structural trend toward lower income inequality. The Gini coefficient dropped in the 1970s, increased in the 1980s, and stabilized in the 1990s—all while continuing to fluctuate around a level close to 0.550. The intensity of trade, although it has increased over the past 30 years, has done so more slowly than was expected.

What shift in the relative factor endowment can explain the persistence of the pronounced inequality and the low intensity of trade in Latin America? Until the 1970s Latin America had a growing relative intensity of land and unskilled labor together with a declining relative intensity of capital per worker (figure 4). But since then the picture has changed radically. The integration of Asia and China into world trade has

Table 2. Trade and Inequality, 1960–95
(dependent variable: Gini coefficient)

|	Trade	Income	Regional effects			
Variable	(1a)	(2a)	(3a)	(3b)	(3c)	(4)
Constant	33.9	−49.9	−26.5	−12.6	−61.5	−55.7
	(57.6)	(1.6)	(0.8)	(0.6)	(0.8)	(2.1)
Trade intensity	0.02		0.02	0.01	0.04	−0.01
	(2.4)		(2.5)	(1.0)	(1.8)	(0.2)
Log (income)		23.5	17.6	5.8	29.5	22.3
		(3.0)	(2.2)	(1.2)	(1.5)	(3.3)
Log (income)^2		−1.57	−1.21	−0.4	−2.1	−1.3
		(3.3)	(2.5)	(1.2)	(1.7)	(3.2)
Asia						4.8
						(3.4)
Eastern Europe						−6.4
						(6.1)
Middle East						10.6
						(6.8)
Latin America and the Caribbean						18.6
						(18.0)
Land-rich						3.8
						(4.3)
East Asia						5.6
						(4.6)
Africa						15.9
						(9.3)
Method	Huber	Huber	Huber	Within effects	Between effects	Huber
R^2	0.002	0.09	0.09	0.006	0.17	0.64
N	565	565	565	565	565	565

Note: t-statistics are in parentheses.
Source: Londoño, Székely, and Spilimbergo forthcoming.

reduced Latin America's relative abundance of unskilled labor. The integration of the Russian Federation has tended to lessen its relative abundance of land. This relative reduction of primary factors should have diminished inequality in the region.

But Latin America has lagged increasingly behind in the relative accumulation of factors associated with lower inequality. The intensity of physical capital increased more rapidly in developed countries. And the intensity of human capital grew more quickly in other developing countries, especially those in Asia. As a result, Latin America lost its relative abundance of human and physical capital.

This development in Latin America's resource portfolio could reasonably explain the persistence of inequality and the low intensity of trade. A simulation of the changing influences on inequality shows that a rise in income and the reduction in the intensity of primary factors lowered inequality in Latin America, but this effect was more than offset by the relative scarcity of accumulable factors (figure 5). As a result, inequality could not have declined.
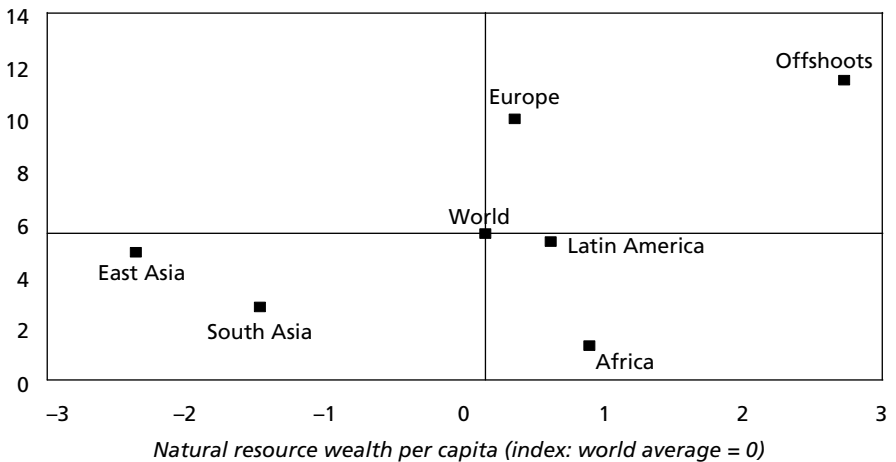
Figure 2. Income Inequality in Latin America, 1970–95

Gini coefficient
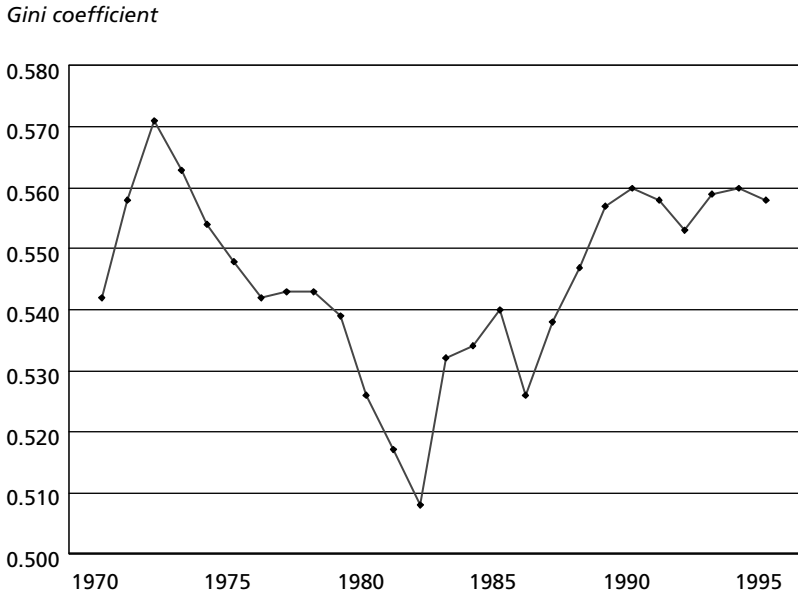
Source: Londoño. Székely, and Spilimbergo forthcoming.

Figure 3. Human Capital and Natural Resources, 1995

*Average years
of education*

Natural resource wealth per capita (index: world average = 0)
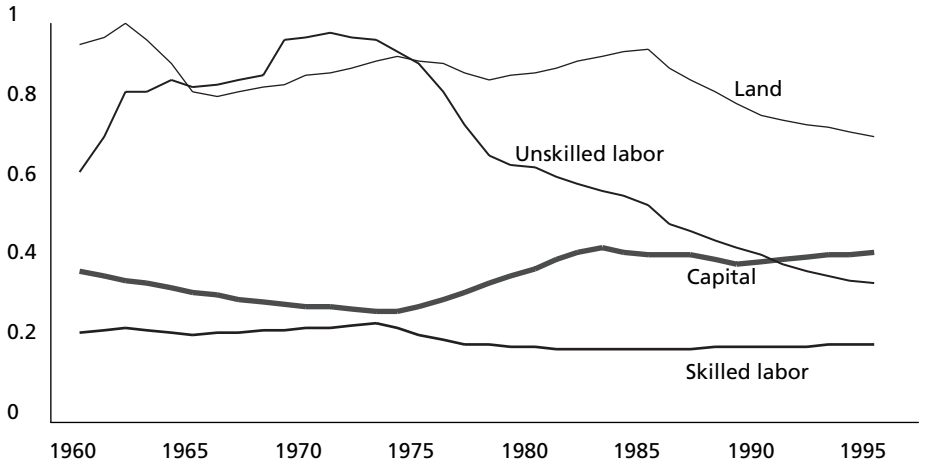
Source: Londoño, Székely, and Spilimbergo forthcoming.

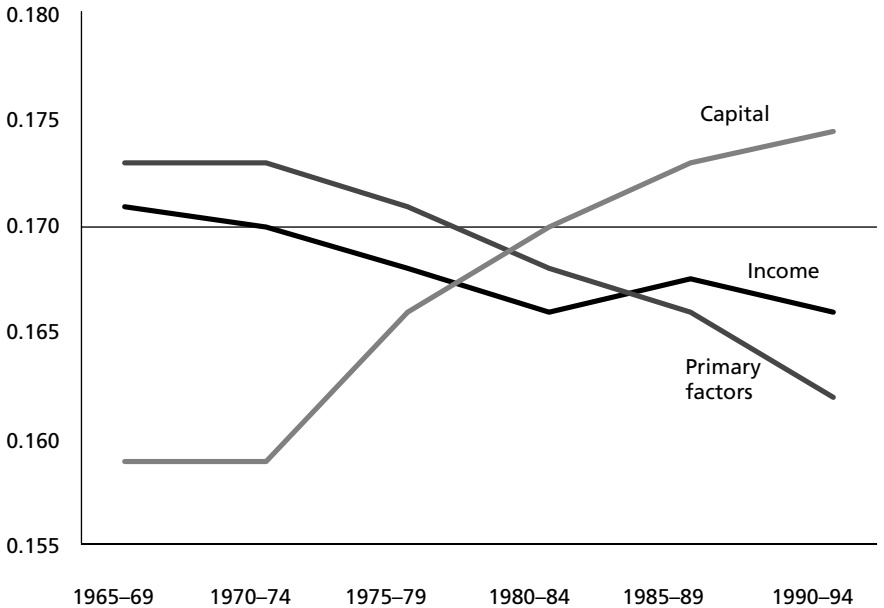**Figure 4. Relative Factor Endowment of Latin America, 1960–95**

Index: World average = 1.0

Source: Londoño, Székely and Spilimbergo forthcoming.

Figure 5. Trends Driving Inequality in Latin America, 1965–94

Contribution to Gini coefficient

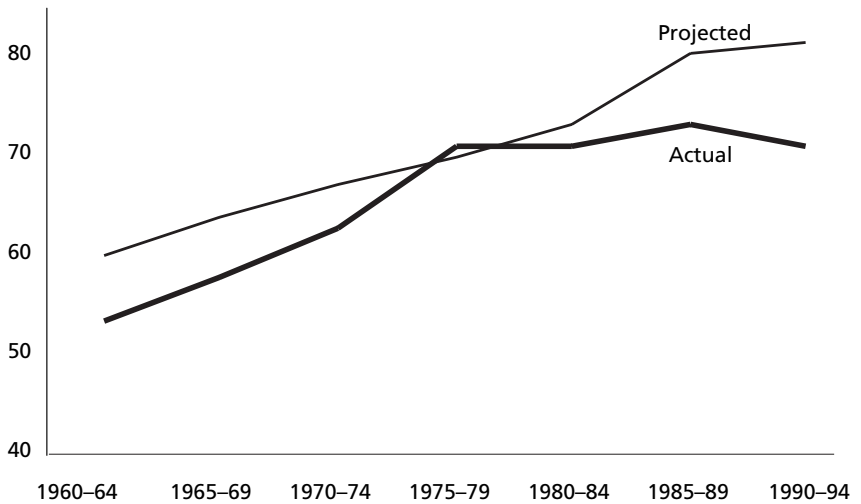*Source:* Londoño, Székely, and Spilimbergo forthcoming.

   With the incorporation of more developing countries into the world trade sys-
tem, the shift of the center of the world economy from Europe to the United States,
and the reduction in transport costs since the 1960s, Latin America's trade intensity
should have increased. In fact, the sum of exports and imports rose from 54 percent
of GDP in the 1960s to 90 percent in the 1990s (figure 6). But this growth in trade
was less than expected. Despite the reduction in obstacles to trade, the gap between
projected and actual trade for Latin America widened in the 1990s. Perhaps the
region's trade policies were less to blame than its resource portfolio. As Latin
America's relative abundance of primary resources steadily declined, the region
failed to build up a relative abundance of capital—a situation consistent with less
trade.

Conclusion

The insights O'Rourke and Cohen provide are most helpful for examining Latin
America's experience in the postwar period. The analysis of inequality needs to go
beyond the static two-factor trade models. The key lies in the factor markets—in the
general accumulation of capital.

Figure 6. Trade Intensity in Latin America, 1960–94

*Intensity of trade
(exports + imports as
percentage of GDP)*

Source: Londoño, Székely, and Spilimbergo forthcoming.

Why has inequality remained high in Latin America? Perhaps the answer has less to do with the greater or lesser opening to trade that the critics make so much of than with the low accumulation of physical and human capital in Latin America in the past 20 years compared with that in other regions. As Nick Stern suggests in his keynote address, the key for both trade and inequality appears to lie in the investment climate—for human capital as well as for machinery and equipment. And this theory is consistent with the explanations put forward by O'Rourke and Cohen in their stimulating articles.

References

Londoño, Juan-Luis, Miguel Székely, and Antonio Spilimbergo. Forthcoming. "Chi va piano non va lontano: comercio, recursos y desigualdad en America Latina." *Revista de la Cepal.*

Lucas, Robert E. 2000. "Some Macroeconomics for the 21st Century." *Journal of Economic Perspectives* 14(1): 159–68.

Comment on "Globalization and Inequality: Historical Trends," by Kevin H. O'Rourke, and "Fear of Globalization: The Human Capital Nexus," by Daniel Cohen

Patrick Asea

Kevin O'Rourke and Daniel Cohen both seek to identify the relationship between globalization and income inequality. Why income inequality? Why are they interested in the size distribution of income—the fraction of national income accruing to a particular share of the population—and not, for example, poverty?

Several reasons come to mind. For developing country policymakers inequality is a concern because it may lead to civil conflict. Inequality is also a concern if the talent in an economy is assumed to be normally distributed and the distribution of income is skewed, so that there appear to be losses resulting from suboptimal allocations between talent and outcomes. Correcting for such distributional inequalities might improve economic performance.

For members of civil society in industrial countries the notion of great disparities in the distribution of income may be morally repugnant. That is probably why we saw a lot of shouting in Seattle and in Washington, D.C. That is also a valid reason for social scientists like us to study the size distribution of income or income inequality.

For economists there is also a theoretical reason. Economists build conceptual models to understand how the world functions, and an important model—the Heckscher-Ohlin-Samuelson model—does not work the way these other models predict that it would.

The Heckscher-Ohlin-Samuelson model suggests that economic integration, or globalization, will drive down the wages of unskilled labor in high-wage countries but drive up unskilled wages in low-wage countries. In other words, income inequality should worsen in high-wage countries and improve in low-wage countries. That is a prediction of a powerful model that economists use as our conceptual framework for thinking about international trade issues.

There has been some evidence of the first implication of the Heckscher-Ohlin-Samuelson model—a widening dispersion in wage rates in industrial countries—in the United States, for example. People with only a high school education in the

Patrick Asea is director of the Economic and Social Policy Division, United Nations Economic Commission for Africa.

United States have lost both absolutely and relatively. There is some evidence that the Stolper-Samuelson effect accounts for the wide dispersion of wages.

There is no evidence of a reduction in the dispersion of incomes in developing countries, however, and this is a puzzle. Experience in Argentina, Chile, Colombia, and Costa Rica has shown that after trade liberalization—after the comparative static effect of tariff reduction—wage inequality increases.

The discrepancy between the data and the Heckscher-Ohlin-Samuelson model is another reason that economists might be interested in inequality rather than poverty. This discrepancy is highlighted by both O'Rourke and Cohen, and it bears more discussion because some observers believe that the interest economists have in increasing inequality may be misplaced.

These comments focus on the article by Cohen, who tackles the theoretical puzzle relating to the Stolper-Samuelson effect—the widening and shrinking of income dispersion. His hypothesis is that skill-biased technological progress might be one cause of the persistent inequality in developing countries, much as in the developed world. He thinks that labor market rigidities might be another.

Globalization and the Returns to Human Capital

The question Cohen poses is this: Does globalization raise the returns to human capital in the North and lower them in the South? And is there any evidence that globalization speeds the transformation of human capital into growth?

So Cohen's article is an empirical one. I would agree with everything he says about the robustness of these kinds of correlations between globalization and income inequality. The consensus is that it is very difficult to get robust and believable results. Having said that, however, I will tackle some of the issues having to do with Cohen's empirical analysis.

When Cohen looks at the returns to education, he draws a useful distinction between human capital gained through formal schooling and that gained through experience. As a baseline, he finds that the returns to schooling are higher than the returns to experience in Asia, Africa, Latin America, the United States, and other OECD countries. Furthermore, he ranks these groups or economies in descending order by the returns to education: Africa, Latin America, Asia, the United States, and the OECD. However, Cohen gives no reasons for this ranking or for its outcome.

The empirical question has to do with the quantity or quality of schooling, a question that can be addressed through the choice of variables. The average number of years of schooling is also important. Clearly, when one looks at the average, one ignores the second moment, and the variance may play a much more important role in developing countries.

Cohen's measure of experience may be less satisfactory than one would like. He uses as a proxy for experience the life horizon of a worker entering the labor force—life expectancy at five years of age minus average years of schooling. This measure clearly does not take into account the fact that in the North, although there has been a large increase in life expectancy, as Cohen correctly points out, there has been an

even greater fall in the average retirement age. As a result, people in developing countries may work longer relative to their life span than do those in developed countries. So the measure of experience may need a little more fine-tuning.

Cohen discusses the returns to education as a central plank in his understanding of the human capital nexus between globalization and inequality. However, he does not explain why the returns to education are so much larger than the returns to experience. This is an interesting point that deserves more attention.

Moreover, perhaps a different measure of experience should be used. While it could be argued that education profiles are similar even across countries, experience profiles may differ significantly within the population of a given country and across countries.

Cohen concludes that his empirical results refute the view that skill-biased technological progress has increased the returns to education over the past two decades, at least at the aggregate (social) level.

Theoretical Insights and Empirical Analysis

Having analyzed the returns to education, Cohen goes on to discuss whether poorer countries are continuing to converge with richer countries. What I find intriguing in his article is that although it is motivated by the fact that a theoretical model does not match the data, it lacks a sense that what might be useful is for theoretical insights to drive the empirical results. I will explain what I mean.

In the standard Heckscher-Ohlin-Samuelson two-good, two-factor model the rise in the relative price of the labor-intensive sector following economic liberalization causes an increase in the real wage rate. In the current models this does not occur in the poorer countries. So we need to develop at least a theoretical framework that will organize our thinking—one that will give us a clear comparative static result consistent with both divergence in industrial countries and greater divergence in developing countries.

Edward Leamer (1999) establishes such a model, one that reinforces Cohen's main point. It is a Heckscher-Ohlin-Samuelson model, but it includes effort and endogenizes it as an important variable in the production function. So effort appears in place of the usual total factor productivity. By making effort endogenous, the model confirms Cohen's results. The model can explain both the divergence in the North and the divergence in the South with a simple comparative static Stolper-Samuelson magnification effect.

The model can also show that labor market rigidities are another important factor in this divergence—and that capital market constraints are yet another. So theoretical insight drives an empirical analysis that will help establish consistency between theory and empirical analysis.

Reference

Leamer, Edward E. 1999. "Effort, Wages, and the International Division of Labor." *Journal of Political Economy* 107: 1127.

This Page Intentionally Left Blank

Growth and Inequality:
The Role of Foreign Trade and Investment

Richard N. Cooper

The article, a survey of theory and evidence, addresses the influence of foreign trade and investment on inequality, with a focus on developing countries. Since growth affects the level of poverty and the distribution of income, the trade-growth nexus is also addressed. The findings are inconclusive, perhaps leading to an agnostic view of the relationship of foreign trade and investment to world economic growth and its distribution. There are no compelling theoretical reasons to believe, in general, that trade promotes growth (as distinguished from an increase in real income), and the empirical work purporting to make a connection at the country level has been criticized on methodological grounds. The theoretical case that foreign investment should stimulate growth and diminish world and host country inequality of income is stronger, and some aggregate evidence credits foreign direct investment (FDI) with a significant growth-enhancing impact.

Despite the ambiguity of theory and evidence, it strains credulity to believe that trade liberalization did not play a significant role in the growth of the world economy in the last 50 years and hence in the great reduction in poverty. The article suggests the (rebuttable) presumption that a country wanting to develop should ensure that exports remain competitive and that domestic production not be insulated from foreign competition.

Global inequality should not be a focus of great interest or research. Undesirable consequences may well flow from greater inequality in particular circumstances, especially if the growth in inequality lacks legitimacy because of the way it was generated. Greater global inequality, on the usual measures, is a natural consequence of uneven growth—and uneven growth is better than none. The key question is whether people's lives are improving.

Richard N. Cooper is Maurits C. Boas Professor of International Economics at Harvard University.

This article addresses the influence of foreign trade and investment on inequality or, more generally, on the distribution of income, with a focus on developing countries. The influence on economic growth of economic openness to the rest of the world has been a topic of scholarly debate. Since growth affects the level of poverty and the distribution of income, the trade-growth nexus is also addressed.

Distribution of income has different meanings, apart from the different measurements used to describe it. Economic theory has been concerned mainly with the functional distribution of income, with the returns to identifiable factors of production and their respective shares in total income of a particular country, such as the share of labor income in national income. Popular and political discourse is more concerned with the size distribution of income, such as the fraction of national income accruing to the top or bottom 10 percent of residents, and with changes in inequality. In recent years, concern with the size distribution of income has extended to its global distribution, with observations by country grouped by per capita income, rather than by individuals.

The two concepts of distribution are related by the ownership of the factors of production, especially land in a predominantly agrarian economy and capital in a modern economy. If ownership of land and capital were evenly distributed across a population, even significant changes in the functional distribution of income would have little impact on the size distribution of income. Somewhat surprisingly, simulated empirical models suggest that the size distribution of income, while significantly influenced by a country's overall development strategy and its institutional structure, is little influenced by economic shocks or by modest changes in policy within a given strategy (Adelman and Robinson 1989).

The article starts with a simple parable of economic change, to fix ideas about the possible consequences for inequality of a single significant shock. Next, it selectively reviews why and how foreign trade might affect the (mainly functional) distribution of income. It also addresses the possible impact of foreign trade on economic growth. A similar review covers inward foreign investment. The article then summarizes some of the empirical work on the impact of foreign trade and investment on growth and income distribution. Finally, the article suggests a simple paradigm for characterizing world economic growth over the past half century and, within this paradigm, questions whether we should be concerned with the global size distribution of income.

A Parable of Change

Assume a coastline with many traditional fishing villages, each economically identical but autarkic. The only (nondomestic) economic activity is fishing. There are three kinds of fishermen (only males fish): ordinary, superior, and energetic. Superior fishermen have special, nontransferable fishing skills—for example, they have better than average instincts about where to find fish. Energetic fishermen work harder (longer hours) than other fishermen. And then there is an element of

luck affecting the daily and annual harvest by each fisherman, luck being distributed randomly.

Ordinary fishermen earn an ordinary (basic) annual wage in fish, adjusted up or down in any given year by good or bad luck. Superior fishermen earn a superior wage, also adjusted up or down by luck. Energetic fishermen earn the basic wage augmented by their additional effort, again adjusted up or down by luck.

The observed distribution of income among fishermen in any given year is thus determined by the skill premium of the superior fishermen, along with their numbers; the extra effort by the energetic, along with their numbers; and by an element of chance.

Introduction of Change and New Opportunities

This ecological/economic equilibrium is now disturbed by the arrival of foreign (to the villagers) ships engaged in some offshore activity that does not directly involve the villagers. But the ships need a local shore base, for resupply of food and fresh water, repairs, and crew rest and recreation. They choose one of the villages, perhaps because of its deeper channel, but to the villagers it is by chance; perhaps the foreign shipowners pay a distant government for the right to use the shore base and pay subsequent royalties as well. The distant government neither collects taxes from nor provides services to the villagers, beyond protecting them from marauders or invaders.

Now suppose that the regular coming and going of ships destroys the fishing activity, perhaps by driving the fish away, and thus deprives the villagers of their traditional livelihood. Their activity does not directly affect fishing in other villages.

The ships and their crews need unskilled labor and local services in the form of food preparation, repairs, eating establishments, and the like, skills not initially available in the village. The opportunity set facing the villagers has thus been drastically altered. Suppose further that the energetic fishermen recognize the possibilities and take steps to acquire the required skills, which command a premium wage. The wage is paid in money, which is used, at least initially, to buy fish from neighboring villages. With an elastic supply of fish, those villages are in competition, but they charge for delivery and use the money to import new goods, thus ending their autarky.

What is the new equilibrium, and how does it compare with the initial situation? Two cases can be distinguished, depending on whether the new demand for labor at the initial basic wage falls short of (case A) or exceeds (case B) the number of fishermen available.

In case A the basic wage will fall enough to employ all the fishermen (since they now need income). Superior fishermen will earn the new (lower) basic wage, since their fishing skills are not transferable to the new activities. Energetic fishermen will earn a new skill premium over the basic wage, which may leave them with either lower or higher income than they earned as fishermen.

In case B the basic wage will rise enough to ration the limited supply of labor. Real income will rise if the new basic wage is sufficient to cover the delivery cost of fish

from neighboring villages, which will be assumed. Superior fishermen will earn this new basic wage, and energetic fishermen will earn this wage plus a skill premium.

Thus, overall, in case A both ordinary and superior fishermen will be worse off than they were initially; energetic fishermen may be better off, if the skill premium is sufficiently high. The distribution of income, however defined, may be more or less unequal than initially, depending on the various wage premia before and after the change and depending on the number of fishermen in each category, but the element of chance (in this parable) will have been eliminated.

Neighboring villages will clearly be better off, on average, since their incomes will have risen by supplying the new demand for fish (presumably at the expense of other fish predators, micro and macro). If the additional fish are provided solely by the energetic fishermen in other villages, the distribution of income may have become less equal there.

In case B all ordinary workers in the directly affected village are materially better off. Superior fishermen will be worse off if the basic wage has not risen enough to cover their loss of skill premium plus the cost of fish delivery. Energetic fishermen will be better off if the new skill premium exceeds their previous extra fishing effort, making no allowance for their leisure or the cost of acquiring the new skills. The distribution of income may be more or less equal than initially, but the relative position of superior fishermen will have deteriorated. The position of neighboring villages will have improved even more than in case A.

Effects of Change on Income Distribution

The arrival of the ships was a major disturbance, destroying previous livelihoods in the village, but also creating new opportunities. In case A this drastic change leaves the directly affected village worse off, although perhaps with a more equal distribution of income. In case B it leaves the village materially better off (it can consume more fish), although perhaps with a less equal distribution of income. In both cases neighboring villages are on average better off, although perhaps with less equal distribution of income.

Case B, in my scale of values, is superior to case A, even if it involves greater inequality; and it may well be superior to the initial condition, even though a traditional way of life has been destroyed. Those who especially enjoyed fishing are big losers, although they can perhaps regain that enjoyment by migrating to one of the other villages. Those who especially disliked fishing are gainers. Such gains and losses in utility are not captured by the conventional economic measures of output and consumption.

This parable, especially case B, while necessarily oversimplified and perhaps overdramatic, captures the essence of economic change brought about by modern technology and globalization. That is, new and better economic opportunities are created for those willing to take advantage of them, but traditional methods of earning a living will generally be made less attractive and in extreme cases (as here) may become unviable. The aged and the inflexible are likely to be losers, but the young

and the more flexible face new opportunities, and on average the gainers will out-weigh the losers. The task of organized society is to ensure that they do.

The parable could be extended in several directions. One might be to allow immigration of labor from neighboring villages, thereby damping the wage increase in the directly affected village but also creating rental incomes there, thus introducing land ownership as a factor. Another might be to allow village women to produce a marketable product, such as woven baskets, used both in fishing and by the ships. The new level and distribution of income will then also depend on the price of baskets and on the distribution of basket-weaving effort and talent and its covariance with pre- and postimpact skill premia. But that would take us further afield than necessary.

The "growth" portrayed in case B may be accompanied by a less equal distribution of income, however measured. Is that necessarily undesirable, assuming that no worker is precluded by law or practice from acquiring the new skills? An element of chance may exist (being in the right place at the right time), but over time effort should prevail. As people acquire modern skills, the inequality of income distribution may rise, perhaps for a long time, before it falls, as Kuznets (1966) conjectured. For example, if income distribution is measured by the ratio of income earned by the top quintile in the distribution to that earned by the bottom quintile, where that quintile is engaged in traditional activities, the distribution will worsen until the income of the second decile (the top half of the bottom quintile) begins to rise more rapidly than average.

In every society some workers are more malleable than others. Malleable labor can adapt more easily to new opportunities. Nonmalleable workers will gain only insofar as demand for their labor rises. Local elites may have special talents that are well suited to the preimpact situation but not easily adapted to change. Thus Chinese scholar-officials of the 19th century generally resisted economic change, since it rendered their specialized knowledge largely obsolete. In our parable the superior fishermen can be expected to oppose the new activity, if they have anything to say about it, unless they are "bribed" by those favoring the change (the shipowners, the distant government, or even the local gainers) to accept it.

The main lesson from this parable is that a major disturbance can either reduce or increase average income and either reduce or increase inequality of income. And the two effects need not be closely related.

Influence of Foreign Trade on Growth, Inequality, and Poverty: Some Theoretical Considerations

In a country open to the world economy, the level and composition of its foreign trade, like GDP, distribution of income, growth of output, and a host of other variables, are determined by the underlying social and political structure, technological possibilities, factor endowments, and household and government preferences across available consumption possibilities. Thus trade is endogenous, like many other economic variables, and jointly determined by the structure and exogenous variables affecting the economy under consideration, where "exogeniety" is itself determined by the ambition of the structural characterization of the economy. Thus it is not pos-

sible in such a framework to discuss the influence of trade on the distribution of income; both are jointly determined by other factors.

Since one purpose of the Annual World Bank Conference on Development Economics is to discover what policy guidance the World Bank and other advisers should give developing countries, we will take an easier route of considering the incremental consequences of liberalizing a country's imports by eliminating import quotas or reducing tariffs—or both. This act of import liberalization is the exogenous change.

Consider a small increment to imports made possible by tariff reductions (for concreteness, think of imports of cut flowers into Europe and the United States). The exogenous change opens up a new range of choices. This development inevitably makes those who initially take advantage of it prospectively better off; otherwise they would not knowingly have taken advantage of it. (If the purchases were merely exploratory, those who initially took advantage of it will stop if purchasers do not like the results.)

The gain is initially in (usually nonmeasurable) utility or satisfaction. There is no necessary impact on the level of output, much less on the rate of growth.

What about the distribution of income? The new possibility for trade enlarges the menu of choice, often (but not always) by lowering price. (For example, cut flowers may be imported profitably only during seasons when domestic production is not available.) If the price of a good falls, that benefits consumers, but it hurts domestic producers of the good or of close substitutes for it. Thus domestic producers of close substitutes will experience a worsening of their terms of trade and hence of their real income. Inequality will increase if these producers are poorer than average; it will decline if they are richer than average. Poverty will rise if the domestic producers are initially just above the poverty line, however defined.

Over time, resources may be reallocated and the structure of domestic output altered as a result of this new trade. Domestic producers may exit production of the import-competing product in favor of now more lucrative productive pursuits. In doing so, they will certainly improve their position with respect to their condition after arrival of the imports (otherwise they would not make the change), and they might even improve their position with respect to the status quo ante, before the new imports arrived. This depends on how lucrative the new pursuits are, which in turn will depend among other things on the new export opportunities opened up—indeed made necessary—by the requirements for restoring macroeconomic equilibrium, e.g., through depreciation of the currency.

If the country is freshly importing a new product, one or more other countries must be exporting it, and that will affect relative prices and the distribution of income there as well, and usually output.

Economic Growth

With a reallocation of resources, the level of the country's output will rise when measured at world prices. When measured at pretrade domestic prices, it may actu-

ally fall, but need not. (This important distinction is usually neglected in empirical work on trade and growth.[1]) GDP at the new prices will of course rise as the real-location takes place. But once the reallocation occurs, this GDP "growth" will cease unless it is sustained by one or more of five factors:[2]

- The redistribution of real income raises the national savings rate, leading directly or indirectly (through the capital market) to a higher rate of investment.
- The relative price of investment goods is reduced, so that a given level of national savings finances greater real investment.
- Productive foreign investment flows into the country in greater amounts on a sustained basis.
- The redistribution of income or new competitive pressure leads people to attain higher levels of economically useful skills.
- The efficiency of labor and capital is continually improved as a result of the new imports, which may convey useful information from abroad as well as exerting greater competitive pressure on domestic producers (leading, in our example, domestic producers of cut flowers to improve their efficiency in production or to discover new products to sell).

The first four of these factors could, of course, have negative signs, leading to a reduction in subsequent growth, following the gains from the initial reallocation of resources. Full employment is assumed to prevail.

A new trading possibility, brought about by import liberalization or by changes in the prices of foreign goods, is closely analogous to an improvement in technology at home: both enlarge the menu of choice, raise the utility of consumers of the products in question, and worsen the terms of trade of producers of competing products. If *technical change* were substituted for *new imports* in the discussion above, the logic would be similar, except that a change in technology would necessarily lead to some change in the cost structure of domestic output.

That raises the question of why public debate draws a sharp distinction between imports and technical change. The answer, no doubt, is complex, arising in part because new imports are often more easily identified with changes in policy, while technical change is thought to be less political in origin (which may or may not be the case), and in part because technical change (by assumption here) is domestic in origin, whereas imports necessarily come from foreigners, who are easier to stigmatize and who have little or no voice in domestic politics.

Technical change can boost the growth rate through a stream of innovations, but a single innovation generally would follow the pattern sketched above. Similarly, the rate of growth should increase (at world prices) during a continuing process of trade liberalization and for a time thereafter, due to response lags, as resources are reallocated.

This discussion assumes that there are no serious market distortions.[3] Market distortions can either reinforce or weaken (and even reverse) the impact on output (measured at undistorted prices) of an exogenous change. With market distortions, incentives may guide the behavior of both firms and households—which are

assumed to respond to the actual incentives they face—in the wrong direction for maximizing output. This important point is taken up again in the discussion of foreign investment below.

The focus so far has been on small policy-induced increases in trade. Comparing free trade with autarky brings another consideration into play: the limited size of the domestic market for all products. If economies of scale or indivisibilities are significant in any sector, this limitation can be important for all but the largest countries. Overall growth under autarky will be limited by growth in the slowest sectors. Trade can help to break bottlenecks and permit a country to enjoy economies of scale whether they occur at home or abroad. This point is not an argument for free trade, but for some trade. The example of the Democratic People's Republic of Korea, once richer and more developed than the Republic of Korea, should warn everyone of the disadvantages of autarky, pursued there in the name of *jusen* (self-sufficiency), although of course many policies are involved, not just severe restrictions on imports.

To sum up the argument so far: there is, in theory, no systematic link between trade and sustained growth. Just as there is no single, simple connection between growth and trade (see Cooper 1996b and the references cited therein), there is no single or simple connection between trade and growth. The impact of new trade on growth may well be powerful in some countries, but it can as well be negligible or even negative in others. There is no reason to believe that the impact will be the same everywhere. Controlling a cross-country analysis of growth for real investment captures the first three of the five growth-sustaining factors listed above, but only by transferring to investment some (much?) of the impact of trade.[4]

This lack of systematic, theoretical connection between trade and growth is potentially important, because growth over time is probably the surest and most effective way to reduce poverty defined against an absolute standard, if not income inequality.

There is, however, an alternative view relating trade, in particular exports, to growth. This view stresses that growth may be constrained by inadequate demand or inadequate availability of foreign exchange. This is an old model, not intellectually fashionable these days, but not wrong for that reason.

Export growth can be the leading sector of a growing economy, stimulating investment. Exports can grow because world demand for them is growing smartly or because the country in question is able to increase steadily its share of the world market through a suitable combination of competitive price and quality.

Export production, of course, is constrained in the short run by installed capacity and labor force. But it need not be constrained in the medium run if the supply of relevant labor is elastic; the supply of investible funds is responsive to the public demand for them, either through national (public and private) savings or through funds from abroad; and any serious bottlenecks can be broken by imports of material inputs, machinery, or disembodied technology.[5]

Within this framework, an effective policy for growth would ensure that

- Exports are competitive, with strong implications for exchange rate policy.

- Supplies of relevant labor and capital are adequate, with implications for policies toward transportation, education, housing, and financial intermediation.
- Requisite imports are readily available, not subject to high tariffs or import restrictions.

As a rough generalization, these seem to be the policies pursued by such rapidly growing economies as the Republic of Korea, Taiwan (China), Singapore, and more recently Mexico and China—each with significant national idiosyncrasies. Some high-income countries have also relied heavily on export-led growth, and Japan continues to do so, as its accumulation over time of an extraordinary $350 billion in foreign exchange reserves—through market intervention to inhibit appreciation of the yen—testifies.

Distribution of Income

Trade can affect the distribution of income directly, without the mediation of its influence on economic growth. It does so by affecting relative commodity prices, which affect the real value of consumption and may also affect the relative and absolute rewards to factors of production. According to the celebrated Stolper-Samuelson theorem (1941), opening a (simple) economy to foreign trade will increase the real income of the owners of the factor of production used intensively in production of the export good and will reduce the real income of the owners of the factor of production used intensively in production of the good now subject to competition from imports, assuming both goods continue to be produced, regardless of the consumption pattern of either factor. (Whether this change increases or reduces inequality depends on initial ownership of the factors of production.)

This remarkable and elegant result, and the closely related factor price equalization theorem, both arising from the insights of the Swedish economists August Heckscher and his student Bertil Ohlin, have received far more attention from professional economists than they warrant. The Stolper-Samuelson theorem assumes competitive markets for goods and factors of production, two goods and two factors with one used intensively in the production of each good, and no specialization in production. It also assumes homogeneous factors of production—labor and capital, or (these days) unskilled labor and skilled labor—combined in known and stable production functions with constant returns to scale and unique factor intensities.

In the short and medium run skilled labor and capital are both specialized, not easily transferable to other uses, so they earn rents to their specialization, which may be high or low and which in a dynamic economy will generally change over time. In the long run all factors may be completely malleable—unskilled labor may be trained, amortized capital may be reinvested in different forms. But in dynamic economies the underlying production function changes over this same long run, often in unpredictable ways. Thus there is a serious mismatch in the Heckscher-Ohlin framework between the assumption of homogeneous factors of production and the assumption of an unchanged production function.

The 2x2 dimensionality of the Stolper-Samuelson result is also problematic. Attempts to generalize the theorem to *m* commodities and *n* factors of production have produced weaker results. For $m = n$ in a competitive economy with constant returns to scale, a rise in any one commodity price will unambiguously improve the real return to at least one factor of production and unambiguously worsen the return to at least one other factor of production, although it may be difficult in a complex economy to identify prospectively exactly which factors those are. The real returns to other factors will depend on their patterns of consumption and could go either up or down in response to a change in commodity price. The same proposition holds for $m < n$ (Jones 1977, 30–31). For $m > n$, the link between commodity and factor prices is further attenuated, and stronger assumptions are required to reach generalizations (Leamer 1995).

A case of special interest is $n = m + 1$, in which each commodity uses in its production a factor specialized to it, while sharing a common factor (unskilled labor, for example) with all other products. The percentage change in rewards to specialized factors, up or down, will be greater than the percentage change in prices of the commodities in whose production they are used. This case is of particular interest since, as noted above, many factors, both skilled labor and capital, are likely to be specialized in the medium run of 5–10 years or even longer.

Propositions deriving from the Heckscher-Ohlin framework assume that a sufficient number of identical products (or for some propositions, all products) are produced in the trading countries. In reality, countries typically specialize, partly because of geographic advantage (natural resources, climate), partly because endowments of common factors are too imbalanced to sustain the production of all goods. The consequence is that price movements in at least some imported goods can occur without affecting factor prices in the importing country, thus permitting all factors to garner a rise in real income from a decline in the price of those imported goods.[6]

Influence of Foreign Investment on Growth, Inequality, and Poverty: Some Theoretical Considerations

A similar agnosticism, or indeterminacy, applies to the influence of foreign investment on economic growth as applies in the case of foreign trade. That is at first surprising, since economic growth is strongly and systematically associated with the rate of investment; insofar as foreign investment augments national investment, it should contribute to growth in economic output. And so it probably does, in general. But not without qualification. And its influence on the distribution of income in the host or receiving country is problematic.

It is useful to distinguish among several different types or motives for foreign investment: loans from governments or international organizations (foreign aid), export credits, bank loans, portfolio investments in marketable securities by foreigners, and foreign direct investment. Foreign direct investment can be subdivided into resource investments, production slicing for reexport, and investment for local sale, the last usually involving a differentiated product of some kind.

All foreign investments except those directly associated with the importation of goods or services (amounting to deferred payments on imports) augment the spending power of the receiving country (unless offset by macroeconomic policy, which will be assumed not to occur in what follows), which in general will be divided between imports and domestic goods and services. This spending may or may not change relative prices; if it does, considerations such as those discussed in the previous section come into play.

Capital Inflows

In general, an inward flow of capital would be expected to lead to a rise in the prices of nontradable goods and services relative to imported goods and services. If the country is a price-taker on world markets, the price of nontradables will also rise relative to export products. This change will affect incomes (such as urban land rents) of factors that are used intensively either in nontradables or in tradables.

A specialized literature discusses possible influences of foreign investment on the terms of trade (export prices relative to import prices; Eaton 1989; Cardoso and Dornbusch 1989). An induced change in terms of trade affects real income for any given level of real output, which may or may not be affected. The terms of trade might improve if the capital inflow leads to a currency appreciation and the domestic prices of export goods do not fall correspondingly (the country faces a downward sloping foreign demand curve for its export products). The terms of trade might deteriorate if the foreign investment augments export supply into a world market with supply-sensitive prices. Or the terms of trade may (and for most countries, are likely to) remain unchanged. Unchanged terms of trade do not, however, imply unchanged relative prices, since, as noted above, the price of nontradables will generally rise relative to the prices of imports and exports. This effect generates the "Dutch disease" phenomenon: resources are drawn from tradables to nontradables, and exports fall as part of the macroeconomic reequilibration of the economy in response to (continuing) capital inflow, with corresponding changes to factor demand and prices. Distributional effects in turn flow from these changes.

Eaton (1989, pp. 1317–47) provides a masterful review of the theory of international capital movements inserted into models of international trade and growth. It is impossible to summarize the results concisely. Not surprisingly, the range of possible outcomes widens with the complexity of the model and with the presence of distortions from competitive market equilibrium. Capital inflows may, but need not, raise real national income. Even when they do, they may, but need not, raise the average national wage.

Mainline classical analysis of a net capital inflow suggests that the capital stock of the receiving country will be augmented. That in turn will depress the returns to capital (assumed to reflect, under competitive conditions, the marginal product of capital) within the country and raise the marginal product of labor and hence the real wage. The foreign investor will be paid the world interest rate or, if the physical capital is directly owned, the (now lower) domestic marginal product of capital,

less any taxes paid to the host government. This arrangement works to mutual advantage so long as the marginal product of capital (net of host country taxes) exceeds the world interest rate. The impact on the host country distribution of income depends on the distribution of ownership of domestic capital. If capital ownership is more concentrated than is labor income, the capital inflow should lead to a more equal distribution of income as well as to a higher national income.[7]

This happy picture disappears in the presence of import restrictions protecting a domestic capital-intensive industry in a small economy. An analysis foreshadowed by Johnson (1967) and developed by Brecher and Diaz-Alejandro (1977) shows that capital inflows under these circumstances (and under assumptions similar to those required for the Stolper-Samuelson theorem) will lower national income measured at world prices. This perverse result arises because in addition to the foreign capital, domestic resources (labor) are drawn into the protected industry from the rest of the economy. The (favored) return to capital and hence wages will remain unchanged, output in the capital-intensive industry will rise, output in the export industry will fall, payments will be made to foreign investors at the protected rate of return, and real national income will fall. This possibility is not merely of theoretical interest, since protectionism continues to be pursued by many countries, sometimes with the explicit objective of attracting foreign investment.

A second example of foreign capital inflows having potentially perverse results for national income and for the distribution of income concerns commodity aid, especially food aid extended on credit at concessional terms. Food aid, other things remaining unchanged, may depress the price of food in the receiving country below the level it would otherwise be. That benefits the urban poor, indeed all net consumers of food. But unless countered by policy it also depresses prices received by domestic food producers. That will lower rural land rents and agricultural wages, thus encouraging migration to the cities. The effects on the distribution of income obviously depend on the relative weights of farm-dependent and other, especially urban, populations; on the ownership of agricultural land; and on institutional factors influencing the relationship between urban and rural wages. But it is not difficult to construct plausible scenarios in which food aid makes the distribution of real income less equal, especially where food production absorbs a large fraction of the labor force.

Foreign assistance for infrastructure should raise national income. If it is devoted to the purchase of imported equipment, it will simply augment the domestic stock of capital, raising factor incomes all around except for capital in direct competition with the new investment. If it is devoted in part to local construction, it will raise demand for labor during the period of construction, both for unskilled workers and for workers with construction skills. That will be a transitory effect, but for large projects it may last for many years. When such aid flows continue over decades, they can create the basis for an indefinitely enlarged construction industry.

However, large public construction projects notoriously provide occasions for rake-offs by politicians and officials, as do large direct government or government enterprise purchases of imported equipment, usually on credit. Such effects also need to be included in reckoning the impact of capital inflows on the distribution of

income. Bribery and "commissions" are often large and are enjoyed by relatively few, often already privileged, individuals.

Foreign Direct Investment

Foreign direct investment (FDI) introduces a wider set of issues. Inflows of capital usually accompany FDI, but in some cases they may be its least important feature. FDI also may bring improved management, new production techniques, quality control, and access to foreign markets that would otherwise be difficult to develop, as well as exerting competitive pressures on local producers, in the markets for labor as well as for goods and services.

These days much new FDI in developing countries occurs in process manufacturing, with some part of a production sequence undertaken offshore, usually because of lower labor costs. The host country imports unfinished components and exports either assembled finished products or more refined components for further processing elsewhere.[8] This type of FDI typically adds little to the host country capital stock, apart from work in progress. It hires and often trains local labor, providing employment and typically raising local wages, at least for those working for the foreign firm. Since employment by foreign firms is rarely more than a small fraction of the labor force, the impact on the distribution of income will be limited unless the national labor market is tightly integrated, which it rarely is in developing countries.

Thus while this type of FDI might in theory raise wages across the board, thereby reducing income disparities, in practice it is more likely to raise wages for a small fraction of the labor force, thereby perhaps widening income disparities by creating a favored local group. More generally, the likely result will be to improve the absolute and relative positions of a group of workers who were already well above the lowest paid, but way below the highest-income residents. Thus it reduces the income shares of both the lowest decile and the highest.

Traditionally, most FDI has not been in process manufacturing but in exploitation of the natural wealth of the host country (minerals, or climate and land suitable for agriculture); in manufacturing of branded products for local sale and possibly regional export (such as soft drinks, cosmetics, and pharmaceuticals); and in maintenance and repair units for complicated and branded imported equipment (high-rise elevators, for example). For these cases the analysis is more complicated, because rents are typically earned in all three activities, and competition is imperfect—sometimes very imperfect—in the manufacturing of branded products for local sale and in maintenance and repair activities.

The economic rents can be shared in various ways, subject to taxation and explicit or implicit bargaining. How the rents are shared affects both the total gains to the host country from the FDI, and the domestic distribution of those gains. For instance, mineral extraction may be heavily taxed, with the revenues disposed of in many different ways. Or the foreign firm may pay exceptionally high wages and commissions in order to build worker loyalty and local political support. As noted above, however, if the FDI was stimulated by protection against imports, its contri-

bution to real national income may be negative, even while it is privately profitable because of the high domestic product prices.

Caves (1999), observing that firms are much larger in rich than in poor countries, considers the obstacles to growth of domestic firms and addresses the potential for spillovers from FDI that could benefit domestic firms by reducing obstacles to growth. He conjectures that the most helpful spillover may be simply demonstrating to domestic firms what is feasible and that this effect is likely to be greater in countries pursuing an outward-oriented trade strategy than in countries relying heavily on import substitution.

Empirical Evidence on Trade, Investment, and Inequality

Theoretical considerations address the impact of trade or investment on the distribution of income within rather than between countries. This section first addresses the impact of trade (mainly U.S.) on domestic income distribution and then looks at the evidence for distribution across countries, the influence of openness on growth, and the distributional impact of foreign investment.

Domestic Income Distribution

From the mid-1970s to the mid-1990s the distribution of income widened substantially in the United States. Some of this involved changes in family structure, toward more single-parent families, and thus is in part a measurement or definitional issue—and may be related to higher incomes and employment opportunities for women. But dispersion also increased considerably in the earnings of full-time male employees. The ratio of male earnings at the ninth decile to those at the first decile increased from 3.18 in 1979 to 4.35 in 1995 (Freeman 2000, p. 38) (This trend seems to have reversed in the late 1990s). The figures have been analyzed from many perspectives. Wage dispersion increased not only across educational levels, but also within skill categories. Indeed, people at the bottom of the scale, without high school education, experienced a decline in real income, an extraordinary development in an economy at high levels of employment and with 50 percent growth in real per capita output over the two decades.[9]

To what extent can this increased dispersion be attributed to foreign trade? The first point to note is that foreign trade developments experienced by the United States, especially the rapid growth of imports of manufactured goods from developing countries, were also generally experienced by Europe and Japan, albeit to lesser degree. With the notable exception of Germany, other rich countries also experienced some widening in dispersion of male earnings over the 1980s, though generally considerably less than that in the United States (Freeman 2000, p. 38). However, in Japan and many European countries, unlike in the United States, unemployment grew over this period, suggesting that pressures similar to those leading to wider wage dispersion in the United States may have led instead to increased unemployment in Europe and Japan, also concentrated among those with lower education or skills.

An extensive literature has developed on the reasons for increased wage disper-sion in the United States, in particular on the portion that foreign trade might explain.[10] The motivating thought is that increased imports of manufactures from developing countries—due partly to continuing trade liberalization in the importing countries but mainly to policy changes in the exporting countries leading to greater engagement with the world economy—in effect enlarged world "endowments" of unskilled labor. Imports of labor-intensive goods put downward pressure on the wages of unskilled labor in rich countries, leading to unemployment in countries where for institutional reasons relatively low wages could not be reduced further. The Stolper-Samuelson theorem was at work.

This is an attractive hypothesis, but it cannot stand close scrutiny. The Stolper-Samuelson theorem operates on factor prices through changes in commodity prices, and the changes in commodity prices required to explain reduced wages of unskilled workers cannot be robustly observed. Hypotheses about less straightforward chan-nels of causation linking trade (especially imports of manufactures) to pressure on unskilled wages have not fared much better in the empirical literature.

Cooper (1996a) examined U.S. imports, production, and employment in textiles, apparel, and leather industries—the tradable sectors that rely most heavily on unskilled labor—and concluded that during the 1980s imports could explain only about 10 percent of the relative decline in wages of unskilled workers, who are also widely employed in the nontradable retail sector. Other studies, very different in approach, produce results of similar magnitude. Most studies attribute the bulk of the increased wage dispersion to technical change that has increased the premium for greater education. Some of the technical change can be identified directly at the plant level (see, for example, Krueger 1993 and Jensen and Troske 2000), but tech-nical change is difficult to measure adequately and much of the attribution is infer-ential or anecdotal.

Blanchflower and Slaughter (2000, p. 78) conclude their review of the impact of foreign trade on wage dispersion by noting that

> The methodological issues surrounding the proper way to gauge trade's role have not been resolved. Nevertheless, what is important to emphasize is that the large majority of studies to date—regardless of their methodol-ogy—find only a small role for international trade in rising U.S. income inequality. Product prices, labor shifts, trade flows: All these data have been analyzed in different ways, and the recurring conclusion is that trade has not mattered much.

Immigration has also been an important feature of the U.S. economy in recent decades, on a much larger scale than had been experienced since before 1920. Numerically, much immigration has been of relatively uneducated, unskilled work-ers, mainly from Mexico and other countries in Latin America. There is also a sub-stantial literature on the impact of immigration on U.S. wage dispersion, again relying on a variety of methodological approaches. It has been summarized by

Camarota and Krikorian (2000), who in their own work find a negative impact of 7–10 percent on the wages of unskilled workers; other studies show a somewhat larger impact. But that is a different channel from foreign trade or investment.

In the Heckscher-Ohlin framework, policy-induced increases in labor-intensive exports would be expected to lower the demand for labor-intensive production in capital-rich importing countries, which would reduce demand for unskilled labor, leading to a reduction in the unskilled wage and an increased dispersion of income. But the same forces would be expected to increase production of labor-intensive goods in the exporting countries, and that in turn under similar conditions should increase the relative wages of unskilled workers and thus reduce income dispersion in those countries.

This does not seem to have happened. Wages of unskilled manufacturing workers in developing countries with rapidly growing exports do seem to have risen, and poverty has declined, but wages of skilled workers seem to have risen even more, contrary to expectation within the Heckscher-Ohlin framework. Chile, Colombia, Mexico, Turkey, and Venezuela, among others, have experienced increased wage dispersion based on education (Wood 1994; World Bank 2001).

A number of explanations are available, but all involve compromising the Heckscher-Ohlin framework in significant ways. Mexico experienced a substantial increase in wage dispersion in the late 1980s, following trade liberalization but before the North American Free Trade Agreement (NAFTA). The premium on education rose significantly, and the wages of nonproduction workers rose relative to the wages of production workers in the northern part of the country, where FDI tends to be concentrated. Since the United States is by far Mexico's largest trading partner, this is difficult to interpret in conventional Stolper-Samuelson terms, unless trade liberalization was concentrated in labor-intensive industries (which it was) and the relative price of labor-intensive products, mainly apparel, declined (which they did not; see Craig and Epelbaum 1996; Feenstra and Hanson 1997; and Hanson and Harrison 1999).

Argentina is a case in which foreign trade arguably had a strong influence on the distribution of income. The country has historically been a successful exporter of grain and beef, two products that also comprised "wage goods" of the Argentine population. A liberal trade policy could be expected to raise the domestic price of these wage goods and the rental returns to productive agricultural and grazing land. Land ownership was concentrated, so a liberal foreign trade policy would redistribute income from many workers, especially urban workers, to fewer farmers and landowners. It has been argued that this structural characteristic might explain Argentine protection against imports for decades following the 1930s (Diaz-Alejandro 1970; Findlay 1984; Leamer 1987). That explanation of course implicitly assumes that the rental income could not have been effectively taxed and used to finance public expenditure and even some domestic redistribution. One of the world's rich countries in 1950, Argentina grew at only about half the world's average rate during the next four decades.

Distribution of Income Across Countries

Complaints are frequent that the world distribution of income has become more unequal in recent decades, meaning that the gap between rich and poor has not been closing. How the gap is measured is not always specified, but sometimes the gap explicitly refers to the difference between average income in one or more poor countries and average income in one or more rich countries (see, for example, World Bank 2001, p. 51, which compares per capita GDP in the 20 poorest and the 20 richest countries). By this standard the gap is not only widening but will continue to widen for a long time, short of some global catastrophe. Growth of a robust 10 percent in a country with a per capita income of $1,000 means an initial annual increase of $100—the same as that in a country with a per capita income of $10,000 that grows at a meager 1 percent. Any growth higher than 1 percent will initially widen this gap between the two countries.

Economists usually focus on growth rates rather than arithmetic increments. If the second country grows at a reasonable 3 percent, the arithmetic difference in incomes in the two countries will continue to grow for 16 years. But in the long run higher growth rates win, and the poor country will have been converging on the rich country from the beginning in geometric terms—the ratio of the second country's income to that of the first will decline continuously (at unchanged growth rates). After 35 years income in the two countries will equalize at $28,000.

Have national per capita incomes been converging in this geometric sense? Put another way, have countries that were relatively poor 30 or 40 years ago experienced higher growth rates than those that were relatively rich? An extensive literature examined this question in the 1990s, stimulated partly by the growing availability of national income and related data for many countries, cumulated over a period long enough to examine economic growth empirically. Also contributing to this interest was the (re)discovery of endogenous growth theory, which posits that economic growth is not determined solely by growth in the labor force and the capital stock augmented by autonomous technical change, but also by economies of scale or endogenous technical change, brought about through learning-by-doing or growth-induced expenditures on applied research, so that growth itself sustains further growth (Romer 1986).

Barro (1997) reviews the results of this research and draws on a sample of more than 80 countries to test whether per capita income in 1960 had a negative impact on the growth in per capita income during 1960–90. Convergence, or catch-up, implies that poorer countries should on average grow faster than richer ones. Initially, the results show that poorer countries did not grow faster on average than richer ones; there is no statistically significant relationship. But poorer countries are found to have grown faster once several growth-relevant variables are controlled for. This has been called *conditional convergence*—conditional on having similar steady-state growth paths, which in the empirical work are assumed to depend on such variables as level of secondary education, life expectancy (as a proxy for general health), rule

of law, democracy (a rough proxy for freedom), fertility, government consumption (a proxy for tax and other policy distortions), and inflation (the last three entering with negative influence). Some convergence is evident after controlling for these basic factors, although it takes place slowly, at 2.5 percent a year (so that half of a given gap would be closed in 27 years). Investment rates are known to be highly correlated with growth rates, but they may be determined by the same underlying factors as growth, so Barro does not consider them a primary determinant of growth.

Barro is careful to point out that a tendency toward convergence does not automatically imply a move toward greater equality among nations. That depends also on the disturbances to which economies are constantly subjected. For a given distribution of disturbances, persistent convergence is consistent with a constant degree of inequality, just as regression of the heights of sons relative to their fathers toward the population mean does not imply that eventually all men will be the same height.

This analysis accords little or no role to foreign trade, investment, or trade policy. The only open economy variable in Barro's analysis that produces a statistically significant result is the terms of trade, which directly affect real income, not output. But empirically an improvement is conducive to higher growth, in part by stimulating more investment.

Openness and Growth

An extensive literature has also developed on the influence of openness or outward orientation on growth rates across countries. This literature is of interest here because differential growth rates among countries affect the global distribution of income over time and because growth rates affect the reduction of poverty, and possibly the distribution of income, within countries.

To take the second point first, Dollar and Kraay (2000) have demonstrated a robust relationship between average income in the lowest quintile and mean national income within 80 countries over a period of four decades. Concretely, poor people (defined as those in the lowest quintile) in rich countries are materially better off than poor people in poor countries, with high reliability. More relevant for the discussion here, growth in average per capita income (on a purchasing power parity basis) can reasonably be expected to raise average income in the poorest quintile roughly in proportion, indeed by proportionately slightly more. This general result does not seem to vary over time, between rich and poor countries, or between crisis and noncrisis periods (where crises are measured at five-year intervals). Moreover, this result does not seem to be very sensitive to the usual determinants of growth, such as those discussed in the preceding section.

In short, growth seems in general to be good for the poor, both in reducing their poverty and in maintaining their position in the income distribution. Empirical generalizations of course are not universal laws; examples can be found where income of the poorest fifth did not keep up with growth in average income. But these cases are not typical. Moreover, there is only one case in the sample where income of the poor grew significantly (12 percent) while average income fell by more than 1 per-

cent. The generalization could thus be rephrased to "growth is *almost* always necessary for serious alleviation of poverty, even if not always sufficient."

What, then, is the contribution of trade to growth? A number of researchers, using different approaches, have found growth to be enhanced by foreign trade, or openness, or trade liberalization (Dollar 1992; Sachs and Warner 1995; Edwards 1998; Ben-David, Nordstrom, and Winters 2000, based on earlier work; among others).

A general methodological problem arises in determining the impact of trade on growth because trade and output are determined simultaneously. Each researcher has developed surrogates for measuring the degree and character of openness, and each surrogate is open to disputation. Indeed, Rodríguez and Rodrik (2001) have provided a withering critique of the studies mentioned in the preceding paragraph, raising serious doubts about whether the authors have demonstrated their claim that pursuit of liberal trade policies has enhanced growth. Rodríguez and Rodrik persuasively find fault with the surrogates, with choice of data, or with specifications of the model to be fitted.

Frankel and Romer (1999) also find a significant impact of openness on levels of per capita income. To avoid the problem of simultaneity, they construct an index of trade possibility based on geographic factors and find that it is strongly correlated with per capita income. They also find that actual trade is positively correlated (r = .62) with trade possibilities and that trade possibilities enhance income through three diverse channels: greater stock of capital, greater stock of education, and higher total factor productivity. But they explicitly caution against using their results to draw inferences for trade policy, which brings different issues into play.

Drawing on the new growth theory, Ades and Glaeser (1999) conjecture that greater openness, by relaxing constraints imposed by the extent of the domestic market, should be associated with higher growth. More particularly, they hypothesize that initial levels of per capita income should have a greater (positive) impact on growth for more closed economies than for more open ones, since more open economies are less bound by domestic market size. Data for 1960–85 for 66 countries, half relatively closed and half open based on the ratio of trade to GDP, broadly confirm their hypothesis: the relationship of growth to initial per capita income is statistically significant for closed economies and insignificant for open ones.

In view of the discussion in the section above on the influence of trade on growth, it would be surprising if all instances of trade liberalization led to greater growth, as distinguished from a once and for all rise in output. It could happen, of course. The model of export-led growth, in either its demand or its supply (bottleneck easing) version, could be generally applicable. Or "liberal trade policy" may simply be a proxy for a complex of more liberal policies, with fewer distortions and less government intervention in the functioning of markets.

Rodríguez and Rodrik provide mainly a technical critique of a number of studies that have concluded empirically that liberal trade fosters growth. They suggest that alternative specifications and definitions of variables would be preferable. In fact, however, their critique generalizes to virtually all country cross-section regressions.

Thanks to the patient and persistent encouragement and pressure by international organizations on member countries to provide economic (and other) data in a standardized format, data in ever lengthening time series are now available for more than 100 countries. The data have proved irresistible to analysts wishing to test broad generalizations that earlier were supported only by theory backed by anecdotes or case studies. And they have been useful for debunking some of those generalizations, as Dollar and Kraay (2000) have done with the widely held view that modern economic growth leaves the poor behind.

But the countries for which observations are available have very diverse political arrangements, the only common element being their seats in the United Nations and other international organizations. Some are cohesive units for collective decision-making. Others are conglomerations of miscellaneous leftovers from European bargaining on imperial boundaries, unable even to maintain domestic order, the prime requisite for functioning government. Most lie between these extremes. So as far as formulation and execution of policy are concerned, the data are not drawn from the same universe, even though they may appear in the same accessible database.

In addition to data from the same (relevant) universe, proper analysis requires a general equilibrium framework, with joint determination of most economic variables. Even the policy variables, normally considered exogenous in economic modeling and estimation, may be endogenous. Some countries take International Monetary Fund/World Bank advice, while others offered that advice spurn it—or accept its rhetoric but not its implementation. In a comprehensive model this difference between acceptance and rejection should be explicable; but from an econometric perspective that fact implies specification error and biased estimates (see Lundberg and Squire [2001] on the problem of simultanety).

The key policy issue is whether for each country, starting where it is, some liberalization of trade (or foreign investment) would improve its economic performance. The answer lies not in cross-section country regressions, however carefully specified, but in detailed analysis of each country under study.

Foreign Investment, Growth, and Inequality

Far less empirical work has been done on foreign investment than on foreign trade, in part because data are neither so copious nor so detailed. Dobson and Hufbauer (2001) estimate conservatively that cumulative foreign investment (mainly FDI) contributed over 6 percent to the GDP of emerging market countries by 2000. This significantly outweighs the damage that foreign investment (mainly bank loans) may have contributed through banking or foreign exchange crises, which amounted to about 3 percent of the GDP of emerging markets (if half the estimated loss in GDP were attributed to foreign investment, an estimate the authors consider high).

The rich countries provided over $700 billion (in dollars of 1995 purchasing power) in economic assistance to poor countries during 1970–95. A World Bank study (1998) finds that economic aid alone did not foster economic growth—an appalling result, even allowing for the fact that much aid was given for political sup-

port to particular countries or governments, not necessarily to increase growth or reduce poverty. From the perspective of economic development, much aid seems to have been simply wasted. However, aid given to countries that pursue effective economic policies can boost economic growth significantly. Aid can contribute to growth in a policy environment that encompasses good management of economic policy and the setting of suitable development objectives. Aid alone cannot ensure the right policy environment; the government must desire economic development, or improvements in health or education, and act accordingly. Vigorous economic growth, in turn, always reduces poverty, even when it enriches some people more than others.

Borensztein, De Gregorio, and Lee (1998), examining the influence of FDI on economic growth in 69 developing countries for 1970–89, find, after controlling for other variables, that FDI makes a positive contribution. But the more significant finding is that this contribution interacts strongly with the amount of secondary school education. At the average level of secondary schooling in their sample (0.9 years for males over age 25 in 1980) "an increase of 0.005 in the FDI to GDP ratio (equivalent to one standard deviation) raises the growth rate of the host economy by 0.3 percentage points per year"(p. 125). This link to schooling is not present for domestic investment. The authors also find that FDI has a positive effect on domestic investment, not a negative one. Kant (1996) finds FDI to be associated with a significant reduction in capital flight as well.

Moran (1998), summarizing the work of others, reports that FDI in manufacturing is found overwhelmingly in highly concentrated industrial sectors (where competition is low). He summarizes three detailed studies on the impact of FDI on national income. Reuber (1973) found that nearly three-quarters of 45 subsidiaries in 30 (mostly rich) countries had production costs higher than those of their parents, suggesting that the firms could have satisfied host country demand from home country production, but did not because of restrictions on imports. These are circumstances in which FDI can actually lower host country GDP. Lall and Streeten (1977), on examining 88 subsidiaries in six developing countries, found that two-thirds had a positive effect on national welfare but that one-third had a negative effect. Encarnation and Wells (1986) evaluated 50 proposed direct investments in a single country, valuing inputs and outputs at world prices, and found that roughly two-thirds (the exact ratio depending on assumptions about shadow prices) would have increased the country's welfare, and roughly one-third would have reduced it. In all three studies, the negative effects arose in areas of low competition.

These studies are all dated, and in particular they predate the tremendous growth in FDI in developing countries (and in developed countries) in the 1980s and especially the 1990s. They also largely predate the extensive growth in FDI associated with locating different production processes for a single end product or class of end products in different countries, according to their costs of production. Some formerly poor economies began direct investment abroad, most notably Hong Kong and Taiwan in China. During this period considerable trade liberalization also

occurred as a result of the Tokyo (1979) and Uruguay (1994) Rounds of multilateral trade negotiations and the extension of free trade areas, most notably NAFTA and the association agreements of the European Union with prospective members. But the antidumping charges and rules of origin associated with regional trade agreements create new opportunities for trade diversion and hence the socially suboptimal location of foreign investment. Thus FDI should be reviewed again in the much altered current environment.

Apart from its effect on GDP, FDI can also have distributional and other local effects. It is widely accepted that firms with FDI in developing countries typically pay higher wages (and better fringe benefits) than domestic firms in the same industry and location. Whether this narrows or widens the distribution of income depends on all the other factors that influence the distribution of income, but in some cases it seems to have created an economic elite of favored workers. Foreign firms are also more likely to pay local taxes, except when (as often in developing countries) they have been granted tax holidays or other special revenue privileges (Hanson 2000).

Foreign firms may bid up the price of relatively skilled labor in the host country and thus bid such workers away from domestic firms. In the long run this may encourage educational attainment, but in the medium run it may worsen the economic condition of domestic firms and depress the local return to capital, with distributional implications. Aitken and Harrison (1999) find on examining data for more than 4,000 plants in Venezuela between 1976 and 1989 that foreign participation raises productivity in recipient plants, especially those with fewer than 50 employees. They also find, however, a negative effect on the productivity of domestically owned plants, thus providing no support for the argument that FDI creates positive spillovers for domestic firms. They summarize that "on balance, our evidence suggests that the net effect of foreign ownership on the economy is quite small" (p. 617). They report similar results from a comparable study of Indonesian firms, with the difference that the positive effects on joint venture firms seem more decisively to outweigh the negative ones on domestically owned firms.

It seems difficult to generalize the distributional impact of foreign investment. To the extent that it stimulates growth, it is likely to reduce poverty. But the impact on the distribution of income in a country will also depend on how much local wages rise, how much local returns to capital fall, and how income is initially distributed among the relevant groups.

Should We Be Concerned About Global Inequality?

It is often simply taken for granted that global inequality is undesirable and that an increase in global inequality is therefore undesirable. Rarely is anything said about why that is so or whether the way it was brought about makes any difference. Yet how global inequality is increased is important for how we evaluate it. In particular, global inequality brought about by an uneven process of economic growth, as each country finds its own way toward development, may not be undesirable—and is cer-

tainly less undesirable than global inequality brought about in other ways, such as through war and conquest, or preserved through lack of growth in richer countries.

Consider a world with two groups of countries: group A countries are rich and growing, and group B are poor and stagnant. The two groups are connected through mutually beneficial trade. Growth requires a complex array of conditions to be met, in particular a stable social system (low domestic turmoil) with widespread education and growth-supporting incentives for effort, saving, and risk-taking. Any small country can benefit by engaging in foreign trade to avoid some of the bottlenecks to growth that arise under autarky. And all countries can benefit from importing successful technology and management techniques, to avoid having to rediscover all improved techniques at home (although some rediscovery is probably useful and more likely to result in productive local adaptation).

Suppose that one by one, countries establish the conditions for growth and switch from group B to group A. During the switching transition they will experience an exceptionally rapid period of growth. What are the consequences of this simple process for the global distribution of income? Measured conventionally as the ratio of the top quintile or decile to the bottom quintile or decile or as the share of the bottom quintile in world income, the world distribution of income becomes increasingly unequal as more countries switch from group B to group A until some countries in the bottom quintile themselves begin the switch.

This is the history of the world economy, broadly characterized, during the last half century. In 1950 relatively rich countries were largely confined to western Europe and the former British regions of settlement: Australia, Canada, New Zealand, and the United States. Then some countries began to grow rapidly, making the switch from group B to group A. A number of countries now unambiguously in the rich category were relatively poor 50 years ago—Italy, Japan, and Spain, to name three large ones—although even then they were richer than some other parts of the world. Over the intervening decades many countries have made the switch, including the poorer European countries and a number of east Asian economies. Others look as if they have joined the process, including some Latin American countries, China, and possibly India.

A crude calculation of world income growth and inequality can be made by pretending that every resident of every country earns the average per capita income of that country (admittedly a gross simplification, in view of the substantial income disparities within countries and the large variation among countries in degree of disparity). Population estimates plus Maddison's (1995) estimates of per capita income can then be used to calculate median world income (50th percentile) and income at the 20th and 90th percentiles. On this basis median world income grew 3.5-fold, or more than 3 percent a year, between 1950 and 1992—no mean achievement.[11] Income at the 20th percentile grew 2.0 percent a year, a respectable rate by historical standards but well below the median. Per capita income at the 90th percentile grew 2.3 percent a year, slower than the median but faster than for the poorest group. The 90/20 ratio rose from 11.5 to 13.3 over this 42-year period, indicating less equal distribution of income in 1992 than in 1950, but the 90/50 ratio fell from

7.8 to 5.8, indicating a narrowing of (geometric) disparities between the well-to-do and the median world citizen.[12] These figures support the notion of gradual switching from group B to group A.[13]

Setbacks are still possible, but the process as a whole takes on an inexorable character. People want to be richer. We now know more or less how to achieve greater wealth, and the hard tasks are in the details of implementation. A major lesson of recent experience has been that ultimately the key to prosperity is a well-educated and disciplined but flexible populace, not, as used to be thought, a generous endowment of natural resources. Natural resources can be helpful in starting the process, but excessive reliance on resource rents can inhibit durable growth and prosperity.

The process, while inexorable, has also been uneven, because recognition and establishment of the conditions for growth have occurred in different countries at different times. The result has been an increase in global inequality on the measurements indicated above, or on their transformations, such as the variance of per capita income across countries, which has risen over recent decades.

But is this something we should worry about? It would be nice to have all countries move from group B to group A, and very likely that will eventually occur. But the necessary conditions cannot be imposed from outside; they must be discovered and embraced by each country or other collective decisionmaking entity. In the meantime, we certainly do not want to slow the switching process, even though for some decades that will imply growing inequality in global income distribution.

Some people may argue that this worsening will threaten the switching process. It would take this article too far afield to expound the possible mechanisms and to present the evidence against that argument. Apart from some possible local effects (for example, two contiguous countries, only one of them growing), I am not inclined to give the argument much credence.[14]

Growing inequality can be blunted or even reversed by income transfers from rich to poor. That suggestion implicitly lay behind many of the proposals in the 1970s of the Group of 77, a coalition of developing countries formed at the conclusion of the first United Nations Conference on Trade and Development in 1964. But ongoing income transfers (as distinguished from episode-related humanitarian assistance, technical assistance to foster development, or loans to help build infrastructure) create a climate of dependence and ultimately resentment that is not healthy for donors or recipients. And in any case income transfers on the vast scale necessary to affect the usual measures of global income inequality would, at least in the near future, be politically inconceivable.

Conclusions

The results of this survey of theory and evidence are inconclusive, perhaps leading to an agnostic view on the relationship of foreign trade and investment with world economic growth and its distribution. There are no compelling theoretical reasons to believe, in general, that trade promotes growth (as distinguished from an increase in real income), and the empirical work purporting to make a connection at the

country level has been heavily criticized on methodological grounds. The theoretical case that foreign investment should stimulate growth, and even diminish world and host country inequality of income, is stronger, but the history of foreign assistance, some of it supposedly targeted on improving growth, is disappointing. And FDI historically has been drawn by natural resources, trade barriers, and low domestic competition—which gives little confidence that FDI has enhanced growth or reduced inequality in income distribution. Nonetheless, some aggregate evidence credits FDI with a significant growth-enhancing impact, especially where adequate skills are locally available.

Despite the overall ambiguity of theory and evidence, it strains credulity to believe that trade liberalization did not play a significant role in the growth of the world economy in the second half of the 20th century. As a whole, this period offers the best economic performance in human history, far better than the often-cited second half of the 19th century. More people, and a higher proportion of them, were lifted out of poverty than ever before, as reflected in the sharp decline in the proportion of workers engaged in agriculture.

To be sure, factors other than trade contributed. Despite numerous conflicts, 1950–2000 was a relatively peaceful period. Thanks to the macroeconomic perspective of the Keynesian revolution in economic thinking, macroeconomic management was markedly better than earlier. And in the late 1940s the world installed a formal framework for economic cooperation among countries, embodied institutionally in the General Agreement on Tariffs and Trade (the GATT, absorbed into the newly created World Trade Organization in 1995), the International Monetary Fund, and the International Bank for Reconstruction and Development. Trade liberalization was of course an important product of this cooperative framework, especially the GATT. (The World Bank, however, financed many import-substitution projects, especially in the 1960s and 1970s.)

Given that trade grew twice as rapidly as gross world product, it is difficult to believe that trade was wholly a product of economic growth. Or, put another way, it is difficult to believe that the world economy would have grown as rapidly had trade barriers remained at the high levels of 1950.

It is of course possible to argue that the trade barriers of a country's trading partners are important to its growth, while its own barriers are not. That would be consistent with the export-led approach to growth, so long as the import barriers do not restrain exports through any of the channels by which they might do so—overvalued exchange rate, macroeconomic imbalance, diversion of resources from production for export, or obstruction of imports crucial to export performance. That case can perhaps be maintained for any single country, or group of economically small countries, who thus become free riders on a liberalizing world economy. More questionably, it could be tenable for all developing countries so long as the rich countries grow, maintain open markets, and continue to dominate the world economy. But future growth of the world economy will depend increasingly on developing countries, especially with the aging of Europe and Japan. Thus trade liberalization by the larger and more rapidly growing developing countries is increasingly necessary for

others to thrive. Both logic and empirical evidence suggest a strong link between economic growth and alleviation of poverty.

While cross-country studies linking trade liberalization to economic growth have been persuasively criticized on methodological grounds, I would nonetheless offer the rebuttable presumption that a country wanting to develop should tie itself to the world economy. It should ensure that exports remain competitive in price and quality and that domestic production not be severely insulated from foreign competition, which otherwise would result in misuse of scarce capital (including foreign capital) and rent-seeking behavior by businesspeople who should be concentrating on improving their businesses.

The presumption is rebuttable in that circumstances in a particular country at a particular time might lead to unwanted and unacceptable consequences from such an opening, possibly of a distributional nature.

Finally, inequality itself, especially global inequality, should not be a focus of great interest or research. Undesirable consequences may well flow from greater inequality in particular circumstances in particular locations, especially if the growing inequality lacks legitimacy because of the way it was generated. The focus then should be on the most efficacious ways to avoid or mitigate the undesirable consequences, or on the lack of legitimacy, not on the inequality as such. Greater global inequality, on the usual measures, is a natural consequence of uneven growth. Even growth is not possible since not all countries are ready to sustain it at the same time. Uneven growth is better than no growth. The key questions are whether people's lives are improving and whether they can look forward with hope to further improvement for themselves and their children. That is the perspective of most individuals, who are not concerned with aggregate statistics on global inequality.

Notes

1. Maddison (1998, 151) finds that China's growth over 1952–78 would be 4.7 percent a year using 1987 value-added weights, compared with an official growth rate of 6.1 percent. Chinese figures weighted manufactures much more highly than agriculture than would be warranted either by world prices of the 1950s or by postreform Chinese prices, imparting a significant upward bias to measured growth. Maddison also recalculated Chinese inflation, resulting in a further reduction in his measure of Chinese growth over this period to 4.4 percent a year.

2. Here *growth* is defined as the commonsensical increase in output over time, not in terms of growth rates of a hypothetical economy in steady-state conditions. That second measure, beloved of growth economists, is sensitive only to labor force growth and autonomous improvements in techniques of production.

3. Such that there is a divergence between market prices and social costs and benefits.

4. The "new growth theory" and its open-economy counterparts emphasize endogenous sources of growth: self-reinforcing economies of scale, external economies that stimulate production of competing or upstream and downstream firms, learning by doing, induced applied research, and the like: in short, any factors that prevent diminishing returns to capital from setting in. Any change in factor prices that can help stimulate one or another of these processes of course contributes to future growth, while changes that inhibit these processes

will detract from growth. See Romer (1986), Helpman (1990), Grossman and Helpman (1991), Krugman (1995), and Gomory and Baumol (2000). Greater openness can contribute to this process by increasing the extent of the market, permitting economies of scale, of whatever nature, to be enjoyed more readily. See Ades and Glaeser (1999).

5. A formal variant of this approach, focusing on supply rather than demand constraints, involves opening the von Neuman growth model, with its noted turnpike theorem, to foreign trade. This model emphasizes input-output relationships, and with elastic supplies of labor (indeed no factors in fixed supply) it can be shown that a maximum growth rate can be achieved by organizing production and reinvestment in a particular way for most of the journey between two points of time not too close together—moving the economy to the fast-paced "turnpike" for most of the journey. For an economy facing fixed world prices, opening the economy will in general result in a higher maximum growth rate. The reason is simple: the possibility of trade, of exporting some products and importing others, widens the range of transformation technologies and thus reduces bottlenecks to period-to-period growth. See Bliss (1989, 1230–34), and the references there cited.

6. Jones (1998) offers a nice example of how, within the Heckscher-Ohlin framework, specialization can lead to results the opposite of the conventional intuition. If two freely trading countries, A and B, each produce two of three possible goods with unique factor intensities under competitive conditions, labor growth in A will lead A to produce more of the most labor-intensive good and less of its capital-intensive good (the middle good in the world economy). Lower world supply of the capital-intensive good will raise its price (unless offset by an altered structure of demand). The higher price will lead B to produce more of that good, which is B's labor-intensive good, thus raising wages in B. Here an expansion of labor in A (and in the world) has led to an increase of wages in B, contrary to the usual expectation.

7. In the case of a small open economy with homogeneous factors of production and no nontradables, where commodity prices are fixed in the world market and techniques of production are unchanged, a capital inflow will raise national output and result in a shift in the composition of output toward the capital-intensive industry but will leave factor returns and hence the distribution of income unchanged: GDP will rise but GNP will not. This result flows from the Rybczynski theorem, a dual of the Stolper-Samuelson theorem. Relying on an analogous mechanism, capital accumulation has been invoked to explain the decline in agricultural employment in developing countries. See, for example, Martin and Warr (1993).

8. Sometimes trade policy discourages the purchase of local inputs, other than labor. Foreign firms are given duty-free import rights for their inputs, while comparable domestically produced imports are protected against competition for domestic sale. Import liberalization can then increase domestic value added. See Hertel and Martin (2000).

9. It is frequently claimed that average U.S. earnings showed no increase from the mid-1970s to the mid-1990s. This contention is difficult to reconcile with the fact that consumption per capita in the United States, in real terms, increased by 2.2 percent a year over the same period. For a partial reconciliation that casts doubt on the claim of no real increase in average earnings, see Cooper (2001). What is important here, however, is not the average level of U.S. earnings, but the indisputable increase in their dispersion.

10. See, for example, Lawrence and Slaughter (1993), Sachs and Schatz (1994), Wood (1994), Cooper (1996a), Cline (1997), and Fishlow and Parker (2000).

11. Median world income grew more rapidly than average (mean) growth in per capita income of 2.2 percent a year because population was growing considerably more rapidly in poor countries than in rich ones.

12. The median world citizen, on the assumption made here, was Indonesian in 1950 and Chinese in 1992. The 20th percentile was Indian in both years. The 90th percentile was British in 1950 and French in 1992. These positions were calculated from population data and per capita income in 1990 international (purchasing power parity) dollars in Maddison's (1995, p. 194–206) sample of 56 countries from all continents. Bhalla (2002) has made a

much more refined estimation of the world distribution of income since 1950. He finds that a Gini coefficient for the world drops significantly after 1980—i.e., moves toward greater world equality. More importantly, poverty (on the World Bank definition) in developing countries drops steadily from 63 percent of their population in 1950 to 35 percent in 1980 to under 10 percent in 1999.

13. A similar calculation for 1987–99, drawing on World Bank data, suggests that the 90/50 ratio dropped from 7.7 in 1987 to 6.6 in 1999. During this period and using this data set, however, the 90/20 ratio also dropped, from 17.1 to 10.2. The countries at the 20th, 50th, and 90th percentiles in 1987 were India, Indonesia, and Japan; in 1999 they were India, China, and France. Calculated from World Bank (1997 and 2001, annex table 1).

14. See Homer-Dixon (1999), and Zimmerman in Gurr (1980) for the influence of intra-country inequality on violence.

References

Adelman, Irma, and Sherman Robinson. 1989. "Income Distribution and Development." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*. Vol. II. Amsterdam: North Holland.

Ades, Alberto R., and Edward L. Glaeser. 1999. "Evidence on Growth, Increasing Returns, and the Extent of the Market." *Quarterly Journal of Economics* 114(August): 1025–46.

Aitken, Brian J., and Ann E. Harrison. 1999. "Do Domestic Firms Benefit From Direct Foreign Investment? Evidence From Venezuela." *American Economic Review* 89(June): 605–18.

Barro, Robert J. 1997. *Determinants of Economic Growth*. Cambridge, Mass.: MIT Press.

Ben-David, Dan, Hakan Nordstrom, and L. Alan Winters. 2000. *Trade, Income Disparity, and Poverty*. Geneva: World Trade Organization.

Bhalla, Surjit S. Forthcoming [2002]. *Imagine There is No Country: Globalization and Its Consequences for Poverty, Inequality, and Growth*. Washington, D.C.: Institute for International Economics.

Blanchflower, David G., and Matthew J. Slaughter. 2000. "The Causes and Consequences of Changing Income Inequality." In Albert Fishlow and Karen Parker, eds., *Growing Apart: the Causes and Consequences of Global Wage Inequality*. New York: Council on Foreign Relations.

Bliss, Christopher. 1989. "Trade and Development." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*. Vol. II. Amsterdam: North Holland.

Borensztein, E., J. De Gregorio, and J.-W. Lee. 1998. "How Does Foreign Direct Investment Affect Economic Growth?" *Journal of International Economics* 45(June): 115–36.

Brecher, Richard, and Carlos Diaz-Alejandro. 1977. "Tariffs, Foreign Capital, and Immiserizing Growth." *Journal of International Economics* 7(November), 317-22.

Camarota, Steven A., and Mark Krikorian. 2000. "The Impact of Immigration on the U.S. Labor Market." In Albert Fishlow and Karen Parker, eds., *Growing Apart: the Causes and Consequences of Global Wage Inequality*. New York: Council on Foreign Relations.

Cardoso, Eliana A., and Rudiger Dornbusch. 1989. "Foreign Private Capital Flows." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*. Vol. II. Amsterdam: North Holland.

Caves, Richard E. 1999. "Spillovers From Multinationals in Developing Countries: The Mechanisms at Work." Harvard University, Department of Economics, Cambridge, Mass.

Chenery, Hollis, and T. N. Srinivasan, eds. 1989. *Handbook of Development Economics*. Vol. II. Amsterdam: North Holland.

Cline, William R. 1997. *Trade and Income Distribution*, Washington, D.C.: Institute for International Economics.

Cooper, Richard N. 2001. 1996a. "Foreign Trade, Wages, and Unemployment." In Herbert Giersch, ed., *Fighting Europe's Unemployment in the 1990s*. Berlin: Springer-Verlag.

———. 1996b. "Is Growth in Developing Countries Beneficial to Industrial Countries?" In Michael Bruno and Boris Plescovic, eds., *Annual World Bank Conference on Development Economics 1995*. Washington, D.C.: World Bank.

———. 2001. "Comment on A. Sapir's 'Who is Afraid of Globalism: The Challenge of Adjustment in Europe and America.'" In Roger Porter, ed., *Efficiency, Equity, and Legitimacy: The Multilateral Trading System at the Millennium*. Washington, D.C.: The Brookings Institution.

Craig, M., and M. Epelbaum. 1996. "Why Has Wage Dispersion Grown in Mexico? Is It the Incidence of Reforms or the Growing Demand for Skills?" *Journal of Development Economics* 51: 99–116.

Diaz-Alejandro, Carlos. 1970. *Essays in the Economic History of the Argentine Republic*. New Haven, Conn.: Yale University Press.

Dobson, Wendy, and Gary C. Hufbauer. 2001. *World Capital Markets: Challenge to the G-10*. Washington, D.C.: Institute for International Economics.

Dollar, David. 1992. "Outward-Oriented Developing Economies Really Do Grow More Rapidly." *Economic Development and Cultural Change* 4(3): 523–44.

Dollar, David, and Aart Kraay. 2000. "Growth Is Good for the Poor." World Bank, Development Economics Research Group, Washington, D.C.

Eaton, Jonathan. 1989. "Foreign Public Capital Flows." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*. Vol. II. Amsterdam: North Holland.

Edwards, Sebastian. 1998. "Openness, Productivity and Growth: What Do We Really Know?" *Economic Journal* 108(March): 383–98.

Encarnation, Dennis J., and Louis T. Wells Jr. 1986. In Theodore H. Moran, ed., *Investing in Development: New Roles for Private Capital?* Washington, D.C.: Overseas Development Council.

Feenstra, Robert C., and Gordon H. Hanson. 1997. "Foreign Direct Investment and Relative Wages: Evidence from Mexico's *Maquiladoras*." *Journal of International Economics* 42: 371–93.

Findlay, Ronald E. 1984. "Growth and Development in Trade Models." In R. W. Jones and P.B. Kenen, eds., *Handbook of International Economics*. Vol. 1. Amsterdam: North Holland.

Fishlow, Albert, and Karen Parker, eds. 2000. *Growing Apart: The Causes and Consequences of Global Wage Inequality*. New York: Council on Foreign Relations.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(30): 379–99.

Freeman, Richard B. 2000. "The New Inequality in the United States." In Albert Fishlow and Karen Parker, eds., *Growing Apart: The Causes and Consequences of Global Wage Inequality*. New York: Council on Foreign Relations.

Gomory, Ralph E., and William J. Baumol. 2000. *Global Trade and Conflicting National Interests*. Cambridge, Mass.: MIT Press.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.

Gurr, Ted Robert, ed. 1980. *Handbook of Political Conflict*. New York: Free Press.

Hanson, Gordon H. 2000. "Should Countries Promote Foreign Direct Investment?" University of Michigan, Ann Arbor.

Hanson, Gordon H., and Anne Harrison. 1999. "Trade Liberalization and Wage Inequality." *Industrial and Labor Relations Review* 52(January): 271–88.

Helpman, Elhanan. 1990. *Monopolistic Competition in Trade Theory*. Special Papers in International Finance 16. Princeton, N.J.: Princeton University.

Hertel, Thomas, and Will Martin. 2000. "Liberalizating Agriculture and Manufactures in a Millennium Round: Implications for Developing Countries." *The World Economy* 23(4): 455–69.

Homer-Dixon, Thomas F. 1999. *Environment, Scarcity, and Violence*. Princeton, N.J.: Princeton University Press.

Jensen, J. Bradford, and Kenneth R. Troske. 2000. "Increasing Wage Dispersion in U.S. Manufacturing: Plant-Level Evidence on the Role of Trade and Technology." In Albert Fishlow and Karen Parker, eds., *Growing Apart: The Causes and Consequences of Global Wage Inequality.* New York: Council on Foreign Relations.

Johnson, Harry G. 1967. "A Note on Tariffs, Distortions, and Growth." *Economic Journal* 57(March): 151–54.

Jones, Ronald W. 1977. *"Two-Ness" in Trade Theory: Costs and Benefits.* Princeton International Finance Section, Special Papers in International Economics 12. Princeton, N.J.: Princeton University.

———. 1998. "Comment on Trade and Wages: A Malign Relationship?" In Susan M. Collins, ed., *Imports, Exports, and the American Worker.* Washington, D.C.: The Brookings Institution.

Kant, Chandler. 1996. *Foreign Direct Investment and Capital Flight*. Princeton Studies in International Finance 80. Princeton, N.J.: Princeton University.

Krueger, Alan B. 1993. "How Computers Have Changed the Wage Structure: Evidence From Microdata, 1984-1989." *Quarterly Journal of Economics* 108(February): 33–60.

Krugman, Paul. 1995. *Development, Geography, and Economic Theory*. Cambridge, Mass.: MIT Press.

Kuznets, Simon. 1966. *Modern Economic Growth*. New Haven, Conn.: Yale University Press.

Lall, Sanjaya, and Paul Streeten. 1977. *Foreign Investment, Transnationals, and Developing Countries*. Boulder, Colo.: Westview Press.

Lawrence, Robert Z., and Matthew J. Slaughter. 1993. "Trade, Jobs, and Wages: Great Sucking Sound or Small Hiccup?" *Brookings Papers on Economic Activity* 2. Washington, D.C.: The Brookings Institution.

Leamer, Edward E. 1987. "Paths of Development in the Three-Factor N-Good General Equilibrium Model." *Journal of Political Economy* 95(October): 961–99.

———. 1995. *The Heckscher-Ohlin Model in Theory and Practice*. Princeton Studies in International Finance 77. Princeton, N.J.: Princeton University.

Lundberg, Mattias, and Lyn Squire. 2001. "The Simultaneous Evolution of Growth and Inequality." World Bank, Development Economics Policy Research Department, Washington, D.C.

Maddison, Angus. 1995. *Monitoring the World Economy 1820-1992*. Paris: Organisation for Economic Co-operation and Development.

———. 1998. *Chinese Economic Performance in the Long Run*. Paris: Organisation for Economic Co-operation and Development.

Martin, Will, and Peter Warr. 1993. "Explaining Agriculture's Relative Decline: A Supply-Side Analysis for Indonesia." *World Bank Economic Review* 7(3): 381–401.

Moran, Theodore H. 1998. *Foreign Direct Investment and Development*. Washington, D.C.: Institute for International Economics.

Reuber, Grant. 1973. *Private Foreign Investment in Development*. Oxford: Clarendon Press.

Rodríguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In Ben Bernanke and Kenneth S. Rogoff, eds., *Macroeconomics Annual 2000*. Cambridge, Mass.: MIT Press for NBER.

Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94(October): 1002–37.

Sachs, Jeffrey D., and Howard J. Schatz. 1994. "Trade and Jobs in U.S. Manufacturing." *Brookings Papers on Economic Activity 1*. Washington, D.C.: The Brookings Institution.

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

Stolper, Wolfgang, and Paul Samuelson. 1941. "Protection and Real Wages." *Review of Economic Studies* 9(November): 58–73.

Wood, Adrian. 1994. *North-South Trade, Employment, and Inequality: Changing Fortunes in a Skill-Driven World*. Oxford: Clarendon Press.

World Bank. 1997. *World Development Report 1997: The State in a Changing World*. New York: Oxford University Press.

———. 1998. *Assessing Aid: What Works, What Doesn't, and Why*. New York: Oxford University Press.

———. 2001. *World Development Report 2000/2001: Attacking Poverty*. New York: Oxford University Press.

Comment on "Growth and Inequality: The Role of Foreign Trade and Investment" by Richard N. Cooper

Andrés Rodríguez-Clare

Richard Cooper's article reflects the conflict among academic economists today about foreign trade and growth. Most of us have the feeling that free trade policies lead to faster growth or significantly higher income levels, and yet we lack good theories for how this happens or evidence in support of that basic intuition.

Trade Policy and Total Factor Productivity

We should not be thinking about the impact of trade on growth, but rather about the impact of trade policy on income levels. The more reasonable and fruitful way to think about country growth is that in steady state all countries grow at the common world growth rate (thanks to technology diffusion), with countries' relative income levels determined by their policies (Klenow and Rodríguez-Clare 1997a). Thus, countries with more liberal trade policies would be expected, other things being equal, to have higher relative income levels, but not necessarily higher growth rates. Differences in growth rates are transitory phenomena that depend on differences between current income levels and steady-state levels, and thus it would be *changes* in trade policy that would affect temporary growth rates. This is also related to the finding by Easterly and others (1993) that policy is persistent whereas growth rates are not, making it impossible for differences in policies to explain differences in growth rates.

The relevant question then is how trade policy affects relative income levels. To advance this issue, it is important to understand the factors behind differences in relative income levels. After some debate there is an emerging consensus that it is differences in total factor productivity levels, rather than factor endowments, that explain most cross-country income dispersion (Klenow and Rodríguez-Clare 1997b; Parente and Prescott 2000). We can thus state the key question more precisely: how does trade policy affect relative total factor productivity levels?

Andrés Rodríguez-Clare is chief, Council of Presidential Advisors, in Costa Rica. He was previously associate professor of business economics, Graduate School of Business, University of Chicago.

Before we can have fruitful empirical research, we need a better theoretical understanding of how trade can have a significant effect on total factor productivity levels. We need calibrated models to see how large these effects can be. Calibration exercises of this kind have shown, for instance, that conventional trade theory implies very small effects of trade on income levels, because Harberger triangles are small.

Some recent theories look promising. Romer (1994) shows that when variety is endogenous, trade can have a more significant impact by affecting the variety of inputs available (Dupuit triangles can be large). In fact, using data from Costa Rica, Klenow and Rodríguez-Clare (1996) showed that lower trade barriers tend to increase the variety of goods imported, although the quantitative welfare implications are still not clear.

Cooper mentions another interesting avenue in his conclusion: the importance of competition for improving productivity among domestic firms. This insight, shared by most economists, has recently been developed theoretically by Melitz (1999), who shows that international trade leads low-productivity firms to exit and high-productivity firms to expand through exports. This finding is consistent with the evidence on the impact of trade on firm entry and exit behavior, but it remains unclear whether this channel is quantitatively significant. It would be interesting to calibrate this model to see the significance of the implied effect of trade on total factor productivity at the general level. Holmes and Schmitz (1998) and Parente and Prescott (2000) have conducted research along similar lines. Other potential channels for trade effects on income levels have also been explored theoretically, including the price and quality of capital goods (Jovanovic and Rob 1996; Rodríguez-Clare 1996; and Eaton and Kortum 2001) and technology diffusion (Keller 2001).

Limits of Cross-Country Empirical Research

Without this theoretical work we can keep running regressions without achieving real progress, as Cooper documents. Consider the research by Frankel and Romer (1999), which appears to show that trade leads to higher income levels. That study received considerable attention because it used geographic variables to generate an instrument for trade, thus getting around the endogeneity problem that has plagued this literature. But then Rodríguez and Rodrik (2000) showed that such geographic variables have a direct effect on income levels, thus negating the validity of the instrument proposed by Frankel and Romer. For instance, when distance from the equator is included in the regression, the coefficient of the trade instrument becomes insignificant in the income equation.

There are two important problems that make it difficult to learn much from this debate. First, Frankel and Romer (1999) find that "a rise of one percentage point in the ratio of trade to GDP increases income per person by at least one-half percent." But what is the mechanism through which trade has such a significant impact on income levels? Without knowing this, we have a correlation but not much under-

standing of the relationship between trade and productivity. Second, it is not clear that the methodology used by Rodríguez and Rodrik (2000) is a valid way to demolish Frankel and Romer's results. With such a small number of countries and such a large set of variables, it is always possible to find a variable that will eliminate another variable's significance level in a cross-country regression (Sala-i-Martin 1997). Moreover, what is the significance of distance from the equator? Why is it a relevant variable in a regression for income levels? Can we choose any variable without specifying the mechanism through which it has an effect on the dependent variable?

I am thus skeptical about the progress we can make from cross-country empirical research at this stage without more theoretical work about how trade affects total factor productivity levels. Still, I would not go so far as Cooper seems to go, when he argues against cross-country regressions on the grounds that the data are "not drawn from the same universe." This would have the extreme consequence of negating the validity of any cross-country regression, no matter how carefully done or how tightly integrated it was with theory. More significantly, it would render futile the search for a theory of development that could explain differences in income levels across countries.

References

Easterly, W., M. Kremer, L. Pritchett, and L. H. Summers. 1993. "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics* 32(3): 363–94.

Eaton, J., and S. Kortum. 2001. *Trade in Capital Goods*. NBER Working Paper 8070. Cambridge, Mass.: National Bureau of Economic Research.

Frankel, J., and D. Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3): 379–99.

Holmes, T., and J. Schmitz. 1998. "A Gain from Trade: More Research, Less Obstruction." Staff Report 245. Federal Reserve Bank of Minneapolis.

Jovanovic, B., and R. Rob. 1996. *Solow vs. Solow: Machine Prices and Development*. NBER Working Paper 5871. Cambridge, Mass.: National Bureau of Economic Research.

Keller, W. 2001. *The Geography and Channels of Diffusion at the World's Technology Frontier*. NBER Working Paper 8150. Cambridge, Mass.: National Bureau of Economic Research.

Klenow, P., and A. Rodríguez-Clare. 1996. "Quantifying Variety Gains from Trade Liberalization." University of Chicago, Graduate School of Business.

——— 1997a. "Economic Growth: A Review Essay." *Journal of Monetary Economics* 40(3): 597–618.

———. 1997b. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *1997 NBER Macroeconomics Annual*. Cambridge, Mass.: National Bureau of Economic Research.

Melitz, M. 1999. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." Harvard University, Department of Economics, Cambridge, Mass.

Parente, S., and E. Prescott. 2000. *Barriers to Riches*. Cambridge, Mass.: MIT Press.

Rodríguez, F., and D. Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." University of Maryland, Department of Economics, College Park. (http://ksghome.harvard.edu/~.drodrik.academic.ksg/papers.html).

Rodríguez-Clare, A. 1996. "The Role of Trade in Technology Diffusion." University of Chicago, Graduate School of Business.

Romer, P. 1994. "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions." *Journal of Development Economics* 43(1): 5–39.

Sala-i-Martin, X. 1997. "I Just Ran Two Million Regressions." *American Economic Review* 87(2): 178-83.

Comment on "Growth and Inequality: The Role of Foreign Trade and Investment" by Richard N. Cooper

Abdul Bayes

Of the many contemporary economic issues and theories related to trade, investment, and inequality addressed by Richard N. Cooper in his article, I would like to discuss just a few here.

What Determines Malleability?

Cooper recounts a parable of change in an economy to fix our attention on the possible consequences for inequality of a single significant shock. The initial ecological and economic equilibrium of his hypothetical traditional fishing villages is disturbed by the arrival of foreign ships engaged in (noncompeting) offshore activities. The ship owners select one of the villages to supply them with various food and nonfood items. The disturbance caused by the regular coming and going of the ships directly affects the traditional livelihood of the supplier village (but not that of the neighbors) by driving the fish away. At the same time opportunities are opened up by new demands for goods and services whose production requires skills (with a premium wage) that are not initially available in the village. The villagers vie for the new opportunities, and some of the villagers, especially the energetic ones, take steps to acquire the required skill. The question is: what is the new equilibrium, and how does it compare with the old one?

Cooper argues that if the new demand for labor, at the basic initial wage, exceeds the number of fishers, the new equilibrium leaves the village better off, albeit with a less equal distribution of income. But if the demand falls short, the village would be worse off, albeit with a more equal distribution of income. By and large the neighboring villagers would be better off, as will the ordinary workers and the energetic workers in the supplier village. The superior fishers, with nontransferable skills, might suffer a setback.

The analysis is analogous to that for many disturbances that economies are required to deal with. For example, trade liberalization hurts some industries but

Abdul Bayes is vice chancellor and professor of economics at Jahangirnagar University, Bangladesh.

helps others; immigrants depress wages but create new demand. The message seems to be the same everywhere: old opportunities are lost, but new ones emerge. In the process, young and more flexible workers tend to win while the aged and the inflexible face hardship. On average, however, the gainers would outweigh the losers.

Cooper's parable of change offers some pertinent lessons of practical importance. First, some trade is better than no trade. Trade is a two-way street that affects the relative prices of commodities and factors of production among partners engaged in exchange. Trade presents new challenges by disturbing the existing equilibrium, but at the same time it opens windows of opportunity. Malleability is especially important. In this connection, Cooper's message is crystal clear: "In every society some workers are more malleable than others. Malleable labor can adapt more easily to new opportunities." Local "elites" may have special talents, well suited to the pre-impact situation, but not easily adapted to change.

Since malleability matters most in reaching a new equilibrium, an important question is what determines the degree of malleability? There are many possibilities, but three appear to top the list: technology, training, and transport and communication. Technology usually follows from openness of the economy and the other two from investment in social and physical infrastructure. A country, therefore, should focus its policies and programs on the development of these factors to diffuse the negative impacts of the disturbance.

However, Cooper's conclusion about the outcome of the disturbance (graduation from traditional to modern activities) focuses mainly on economic factors. Recent experience in developing countries points to the importance of noneconomic factors in accelerating economic growth, such as law and order, educational attainment, private property rights, and sociopolitical factors. These deserve serious attention. Also, the sources of growth need to be considered. The loss of a traditional way of life can be costly in the future if those traditions tend to preserve and protect the environment (especially when natural resources are an input rather than an amenity). In that case, the same disturbance that delivers benefits today could turn out to be destructive tomorrow. Cutting trees might raise per capita income in the short run but reduce it in the long run. And finally, what would have happened had there been more than one significant shock to the villagers? And would a different path of growth have emerged had the foreign boats been involved in competing fishing activities?

"In Defense of Inequality"!

Cooper argues that as people acquire new, modern skills, income inequality may rise, perhaps for a long time, before it falls. He goes on to say that "inequality itself, especially global inequality, should not be a focus of great interest or research. Undesirable consequences may well flow from greater inequality in particular circumstances in particular locations, especially if the growing inequality lacks legitimacy because of the way it was generated."

These observations appear to be close to the views held by many economists. The following remarks by another economist (Welch 1999) are representative of this

perspective: "Inequality is an economic 'good' that receives too much bad press. All of economics results from inequality. Without inequality of priorities and capabilities, there would be no trade, no specialization, and no surpluses produced by cooperation. Incidentally, there would be no economics, and we would all be selling insurance! Actually not—without inequality, there would be nothing to insure."

The core of the concern in Cooper's article, as elsewhere, is legitimacy, a concern about how inequality is generated, not why. I concur fully with this view. There should be no inequality in access to opportunities. Inequality is destructive only when people, especially, those at the low end of the income ladder, view access to opportunities, mobility, and malleability as unfair. Policymakers and politicians need to guarantee equal access for all.

Cooper is correct in saying that new technology and new skills would initially widen inequality in the economy. Several studies have shown how and to what extent the technology revolution has created wage differentials in U.S. society. These studies also observe that the gap could be bridged through appropriate investments in training and other means to broaden access to new technologies (Caselli 1999, p. 98; Krueger 1993, pp. 54–56). If computer illiteracy, for example, is a cause of inequality, there is no point in cursing computers. Empirical research in Bangladesh has shown that the Village Pay Phone program, which makes mobile phones available to villagers, tends to generate more consumer surplus for the poor than for the nonpoor (Bayes, Von Braun, and Akhter 1999). From a policy perspective, then, the concern should be to prevent a widening of the opportunity gap between rich and poor by adopting, to the extent possible, broad-based technology and skills that are pro-poor. But it is not just the technology itself that will reduce inequality. Also important for income distribution are institutional arrangements and ownership of the technology.

Globalization and Global Inequality

Trade liberalization is the focus of intense concern today, vilified not only in developing countries but even in the developed countries that once so vigorously championed its virtues. Foreign trade and the influx of imports of manufactures from developing countries are blamed for the substantial widening of the distribution of income in the United States from the mid-1970s to the mid-1990s, for example. It is alleged that the imports of labor-intensive goods have put downward pressure on the wages of unskilled labor in rich countries. This outcome is exactly what would be predicted from the Stolper-Samuelson theorem. If the sun sets in the United States, it has to rise somewhere else—and vice versa. Under some conditions the relative wages of unskilled workers would rise in exporting countries, reducing income dispersion in those countries.

Empirically, however, there is very little evidence for the claim that cheap imports from developing countries affect the wages of unskilled workers in developed countries. Cooper (1996) concluded that only about 10 percent of the relative decline in wages of unskilled labor over the 1980s could be explained by imports. Other stud-

ies have also found that international trade has had only a small influence on the rising inequality of income in the United States.

Cooper deals eloquently with this issue, but mostly in relation to developed countries. But I expected more on the widespread allegations that trade liberalization is leading to the marginalization of many developing countries. One school of thought, found mostly in developing countries, alleges that trade liberalization is "of the rich" and "for the rich" but "by the poor" and that governments should therefore not open up the economy but rather should slow economic reforms and trade liberalization. By and large, trade liberalization and globalization are treated as a monster. Polemics are overtaking economics.

Trade liberalization has been associated with the marginalization of a large number of developing countries, including the least developed ones, but correlation is not the same as causation. During 1980–96 industrialized economies' share in world merchandize exports increased from 64 percent to 68 percent, that of the G13 economies soared from 10 percent to 21 percent, and that of the "other" developing countries shrank from 21 percent to 8 percent (Ghose 2000, pp. 285–86). So while there is marginalization, it might not have been caused by trade liberalization. Ghose (2000) argues that marginalization is partly explained by the failure to liberalize trade in agricultural commodities, which are major export items for many of the marginalized economies. Ghose argues further that in the absence of a substantial shift in the commodity composition of these countries, marginalization would have occurred whether liberalization occurred or not.

So why are so many countries being marginalized? There are many reasons, and some of them have already been mentioned in the discussion of the malleability of labor and inequality of income. A host of researchers blame the low level of development of infrastructure, both physical and social. Trade liberalization, they tend to say, has little to do with the downturn. Research in Bangladesh (one of the least developed countries) on business constraints faced by entrepreneurs ranks trade liberalization well below such factors as business environment and physical infrastructure (World Bank 1999, pp. 38–39).

In fact, both Cooper and Ghose (2000) suggest that trade liberalization, by inducing two-way trade in manufactures, contributed to the process of convergence between industrialized and developing countries. Nonetheless, it may be that the pace of liberalization might need to be slowed until other reforms catch up with trade reforms.

Convergence and Convulsions

There is no denying that over the past four decades or so incomes in some developing countries have been converging toward those in developed countries and that once several growth-relevant variables are controlled for, even more countries exhibit convergence. However, there are still only about 17 converging countries, accounting for about 74 percent of all developing country exports (87 percent of manufacturing exports) and 82 per cent of the foreign direct investment flows into

developing countries (Ghose 2000, p. 285). So saying that developing countries are converging is to commit the fallacy of composition.

The recent financial crises that swept some of these economic "tigers" has again thrown convergence into serious question. Cooper could have touched on the turmoil created by the crises and let us know whether such convergences are sustainable—that is, whether these economies have to depend on bailouts to survive or can rely solely on their own internal dynamics.

In Support of Growth

Cooper concludes with the observation that "uneven growth is better than no growth. The key questions are whether people's lives are improving and whether they can look forward with hope to further improvement for themselves and their children. That is the perspective of most individuals, who are not concerned with aggregate statistics on global inequality." Cooper appears to have jumped to this conclusion, which rests on very little evidence in his article. Still, there should be no disagreement about the necessity for growth. Growth is the key to poverty reduction, but growth is only a necessary condition. The sufficient conditions also need to be met for growth to become enduringly poverty reducing. Again it is perhaps true that individuals are not concerned with inequality (global or even domestic)—but only so long as inequality is not "illegitimate" and does not seriously affect inter-personal utility in any society.

References

Bayes, Abdul, Joachim Von Braun, and Rasheda Akhter. 1999. "Village Pay Phones: Insights from Grameen Bank Initiative." Research Report, Centre for Development Research, Bonn.

Caselli, Franasesco. 1999. "Technological Revolutions. " *American Economic Review* 89(2): 78–102

Cooper, Richard N. 1996. "Foreign Trade, Wages, and Unemployment." In Herbert Giersh, ed., *Fighting Europe's Unemployment in the 1990s*. Berlin: Springer-Verlag.

Ghose, Ajit K. 2000. "Trade Liberalization, Employment and Global Inequality." *International Labour Review* 139(3): 281–306.

Krueger, Alan B. 1993. "How Computers Have Changed the Wage Structure: Evidence from Micro Data, 1984–1989." *Quarterly Journal of Economics* 108(1): 33–60.

Welch, Finis. 1999. "In Defense of Inequality." *American Economic Review* 89(2): 1–17.

World Bank. 1999. "Bangladesh Trade Liberalization: Its Pace and Impact." South Asia Region, Poverty Reduction and Economic Management Unit, Washington, D.C.

Geography and International Inequalities: The Impact of New Technologies

Anthony J. Venables

Some writers have predicted that new technologies mean the "death of distance," allowing suitably skilled developing economies to converge with high-income countries. This article evaluates the prediction, arguing that geography is important for international interactions and for international income inequalities. Firms in remote locations—or served by poor infrastructure—can compete only if wages are low. There is empirical evidence that distance from established centers is a major determinant of per capita income levels.

New technologies will change the costs of distance in several ways. Some activities will become "weightless," so shipping costs become irrelevant. However, these will probably account for only a small percent of world GDP, and for most activities shipping costs on imported inputs and on exports of finished products will remain substantial. New technologies will facilitate management of remote operations and supply chains, as already seen in the growth of foreign direct investment and production networks. However, for many activities information is complex and hard to codify, so face-to-face communication will remain important. For some activities, such as just-in-time processes and products where demand changes fast, new technologies will actually increase the cost of distance, as being close to customers becomes more important. For these products new technologies provide information enabling suppliers to respond rapidly, providing they are close enough to consumers to meet rapid delivery times.

The analysis suggests that new technologies will allow some activities to relocate—those that are weightless or that are simple enough for relevant information to be easily codified and digitized. Other activities may well become more entrenched in established locations that have the advantage of dense networks of related activities. Activities that do relocate to lower-wage countries are themselves likely to cluster, suggesting that a relatively small group of countries will be beneficiaries. New technologies will therefore change economic geography by allowing some countries to

Anthony J. Venables is professor of international economics at the London School of Economics and director of the International Trade Program at the Centre for Economic Policy Research.

benefit from the relocation of some activities, but will leave many other countries outside this process.

New information and communication technologies offer many benefits to developing countries. Costs of establishing communications networks have been slashed, and with the savings comes the prospect of better provision of education, health care, and a host of other services. Some writers go further, arguing that information and communication technologies presage the "death of distance." In the words of Cairncross (2001, p. 16):

> To allow communications to work their magic, poor countries will need sound regulations, open markets, and, above all, widely available education. Where these are available, countries with good communications will be indistinguishable. They will all have access to services of world class quality. They will be able to join a world club of traders, electronically linked, and to operate as though geography has no meaning. This equality of access will be one of the great prizes of the death of distance.

This article evaluates this claim, asking, what are the prospects that information and communication technologies will lead to the death of distance? It finds that at present geography matters a great deal for economic interaction and for the spatial distribution of income. It considers how new technologies might change this, and what they might do for the location of economic activity and for international inequalities.

Conceptual Framework

The conceptual framework for addressing this question is based on the profitability of production in different countries, knowing that a change that increases profitability will tend to attract firms and bid up wage rates. The profitability of a location is determined by many forces: labor costs and efficiencies, the social infrastructure of the economy, and geography—location relative to sources of supply and to markets. The fact that firms tend to locate close to their markets creates a force for international inequality. Established economic centers offer large markets, attracting firms and hence supporting high wages—which in turn support the large market size. Pulling in the opposite direction are international wage differentials (or primary factor costs more generally). Obviously, the lower are primary factor prices, other things being equal, the more profitable is production in the country, a force for international equality.

The tradeoff between these forces provides a simple relationship between costs of distance and international inequalities. As the following section shows, there are international wage gradients, with wages falling as a function of remoteness from markets. Insofar as new technologies reduce the costs of distance, they might be

expected to flatten these gradients and reduce international inequalities. If trade were to become perfectly free—the limiting case of textbook international economics—distance would be dead, goods markets perfectly integrated, and factor price equalization would hold. Perfectly free international trade means that similar factors get paid the same price, regardless of location, although per capita income levels may differ as individuals own different amounts of human and physical capital.

This view of the effects of information and communication technologies is misleading, for at least two reasons. First, new technologies will have a mixed and complex effect on the costs of distance. Some activities can be digitized and supplied from a distance, but most cannot. Second, geography determines firms' profitability not only through ease of access of markets but also through access to a cluster of related activities. The propensity of economic activity to cluster is widely documented (for example, Porter 1990) and attributed to a range of forces. One is the development of dense local networks of suppliers of specialized goods and services for industry. A second is development of local labor markets with specialist skills, probably arising because of the training activities of other firms in the industry. A third is the benefit of being close to research centers and to the knowledge spillovers that firms derive from proximity to other firms: "the mysteries of the trade become no mystery; but are, as it were, in the air" (Marshall 1920 [1890], p. 271). Finally, it may simply be easier to manage and monitor activities in an established center where firms have local knowledge and can benchmark their performance on that of other firms in the same location.

How does new technology change these clustering forces? Some are likely to be weakened by new information and communication technologies. For example, proximity may come to matter less for the flow of knowledge between firms and for the supply of business services (at least to the extent that the relevant knowledge can be codified and digitized). But other clustering forces—such as those arising from labor market skills—are likely to be unaffected. Determining the overall effects of information and communication technologies on location and international inequalities therefore requires taking into account the fact that distance may die for some functions in some industries while remaining important for many other functions and activities. Thus some activities will no longer need to be close to consumers and will go in search of lower-cost locations. But low costs depend on wages, social infrastructure, and access to the benefits of a cluster of related activities. Consequently, some activities may tend to move to low-wage economies, while others become more deeply entrenched in high-wage economies.

These effects are illustrated by the experience of previous communications revolutions. The transport revolutions of the 19th century led not to the dispersion of economic activity but to its concentration—in relatively few countries and, within those countries, in large and often highly specialized cities. Lower transport costs reduced the value of being close to consumers, who could instead be supplied from cities in which production exploited the advantages of increasing returns to scale and agglomeration externalities. So too with new technologies we might expect to see changes in the economic geography of the world economy, but not necessarily changes toward the "integrated equilibrium" view of the death of distance.

This article develops these arguments in three stages. First, it shows that geography matters greatly for many economic interactions—be they trade, investment, or knowledge transfers—that are overwhelmingly local, falling off sharply with distance. The costs that cause interactions to fall off across space have major implications for world income distribution. Measures of distance based on the intensity of economic interaction between countries show that distance can account for a large part of international inequalities. Poor countries are poor, in part, because distance inhibits their access to the markets and suppliers of established economic centers.

Second, it considers the effects of information and communications technologies on the costs of international transactions. To do this, it looks more deeply at why distance is costly, examining the costs of searching and matching (identifying potential trading partners), moving inputs and outputs (direct shipping costs), monitoring and management, and time taken shipping to and communicating with distant locations. Information and communication technologies reduce some of these costs for some activities, but the effects are ambiguous and can in some cases increase rather than reduce the value of proximity.

Finally, the article examines the likely effects of these cost changes on the location of activities and hence on wages and income levels. Will centers of economic activity deconcentrate, with activities relocating to lower-wage economies? This will occur for some activities, but for others concentration in central regions may well be reinforced. Furthermore, activities that relocate will tend to cluster in relatively few new locations. Thus new technologies may change the pattern of inequalities in the world economy but not necessarily reduce them. In this way new information and communication technology may be like previous rounds of infrastructure development, such as that for canals, railways, and road networks, which permitted deagglomeration of some industrial activities but probably reinforced rather than diminished centralizing tendencies (Leamer and Storper 2000).

Does Distance Matter?

Almost all economic interactions fall off very rapidly with distance. Some of the reasons for this are presented later, after a simple outline of the facts.

Distance and Economic Interactions

The standard framework for quantifying the effect of distance on economic interactions is the gravity model, which relates interactions between a pair of countries to their economic mass and to a measure of the cost of the interaction between them. This framework has been applied in a number of different contexts, most of all to trade flows. Thus, if y_{ij} is the value of exports from country i to country j, the gravity relationship takes the form

$$y_{ij} = s_i m_j t_{ij}^\theta$$

(1)

where s_i denotes exporter (supplier) country characteristics, m_j denotes importer country characteristics, t_{ij} is a set of between-country factors measuring the costs of trade between the countries, and θ is the elasticity of trade flows with respect to these between-country factors. This between-country term is typically proxied by distance and perhaps also by further between-country characteristics such as a treaty relationship or a common border, language, or history. Exporter and importer country characteristics can be modeled in detail, including income, area, population, and geographica features such as being landlocked. However, if the main interest is the between-country term, t_{ij}, then s_i and m_j can simply take the form of dummy variables whose values are estimated for each country.

Extensive data permit the gravity trade model to be estimated on the bilateral trade flows of 100 or more countries. Studies find an elasticity of trade flows with respect to distance of –0.9 to –1.5. The decline in trade volumes with distance implied by this relationship is extremely steep. If $\theta = -1.25$, then relative to 1,000 kilometers volumes are 82 percent lower by 4,000 kilometers and 93 percent lower by 8,000 kilometers (table 1).

Similar methods have been used to study other economic interactions. Portes and Rey (1999) study cross-border equity transactions during 1989–96, using data for 14 countries accounting for around 87 percent of global equity market capitalization. Their main measure of country mass is stock market capitalization, and their baseline specification gives an elasticity of transactions with respect to distance of –0.85. This indicates again how much—controlling for the characteristics of the countries—distance matters (column 2 of table 1).

Other researchers have studied foreign direct investment and technology flows. Data limitations mean that the set of countries is once again quite small, and the estimated gravity coefficient is smaller, although still highly significant. Di Mauro (2000) finds an elasticity of foreign direct investment flows with respect to distance of –0.42. Keller (2001) explores the effect of distance on technology flows, looking at the dependence of total factor productivity on research and development (R&D) stocks (cumulated R&D expenditures) for 12 industries in the G-7 countries for 1971–95. The R&D stocks include both own-country stock and foreign-country stocks weighted by distance. Both own- and foreign-country stocks are significant determinants of each country's total factor productivity, as is the distance effect, with R&D stocks in distant economies having much weaker spillover effects on total factor productivity than do R&D stocks in closer economies (last column of table 1).[1]

Distance and Real Income

If, as argued above, distance matters greatly for economic interactions, how does that feed into the distribution of income across countries? Several mechanisms might be at work, including the effects of investment flows and technology transfers. Here, to illustrate the effects, attention is focused just on the way that trade flows and the implicit trade costs demonstrated by the gravity model can generate international income gradients.

Table 1. Economic Interactions and Distance
(Flows relative to their magnitude at 1,000 kilometers)

Distance (kilometers)	Trade ($\theta = -1.25$)	Equity flows ($\theta = -0.85$)	Foreign direct investment ($\theta = -0.42$)	Technology[a]
1,000	1	1	1	1
2,000	0.42	0.55	0.75	0.65
4,000	0.18	0.31	0.56	0.28
8,000	0.07	0.17	0.42	0.05

a. Includes own-country stock and foreign country stocks weighted by distance ($\exp[-\theta \, \text{distance}_{ij}]$).
Note: See text for details.
Source: Author's computations based on Redding and Venables (2000) for trade, Portes and Rey (1999) for equity flows, Di Mauro (2000) for foreign direct investment, and Keller (2001) for technology.

The effect of distance on factor prices is seen through a simple example. Suppose that country 1 represents the high-income countries from which country 2, a developing country, imports intermediate goods and to which it exports manufactures. The cost of producing manufactures in country 1 is given by $c(w_1,r_1,q)$, where w_1 and r_1 are the unit costs of labor and capital and q the cost of intermediate goods.[2] The developing country has to import the intermediate good, and imports are subject to trade costs at proportionate rate t.[3] (These trade costs consist of a number of different elements, discussed in the following section.) Trade costs at rate t mean that the price of intermediates in country 2 is tq, so country 2's units costs are $c(w_2,r_2,tq)$, given its factor prices, w_2 and r_2.[4] It sells in the developed country market but faces trade cost factor t in shipping to this market.

For country 2 to compete with production in country 1, the following equation must therefore hold:

$$(2) \qquad c\left(w_1,r_1,q\right) = tc\left(w_2,r_2,tq\right).$$

Country 2's wages (expressed as a proportion of country 1 wages) are shown in figure 1 as a function of trade costs, computed from this relationship with the assumption that $r_2 = r_1$. The curves can be thought of as illustrating the wage gradient for different countries at increasing distances (increasing trade costs) from the center. In the three cases illustrated, two-thirds of value added is labor and one-third capital. In the upper line there are no intermediate goods, while in the other two lines intermediates account for 25 percent and 50 percent of country 1's costs. The figure shows how rapidly wages get squeezed at more remote locations with higher trade costs. Thus if trade costs are 30 percent of the value of output ($t = 1.3$) and intermediate inputs are 50 percent of costs (bottom curve), wages drop to about one-tenth their level in the center. Trade costs of 30 percent are not that high (the median cif/fob ratio for all countries reporting bilateral trade is 1.28). Furthermore, if the price of capital were higher in more remote locations ($r_2 > r_1$), wages would be depressed further.

To establish in fact the importance of distance for international inequalities that is suggested theoretically in figure 1 requires generalizing it to many countries and to the full set of trade relationships between them. Market access of country i,

Figure 1. Relationship of Trade Costs and Wages, at Different Costs of Intermediate Goods

Wages relative to central wages

No intermediates

Intermediates 25% country 1 costs

Intermediates 50% country 1 costs
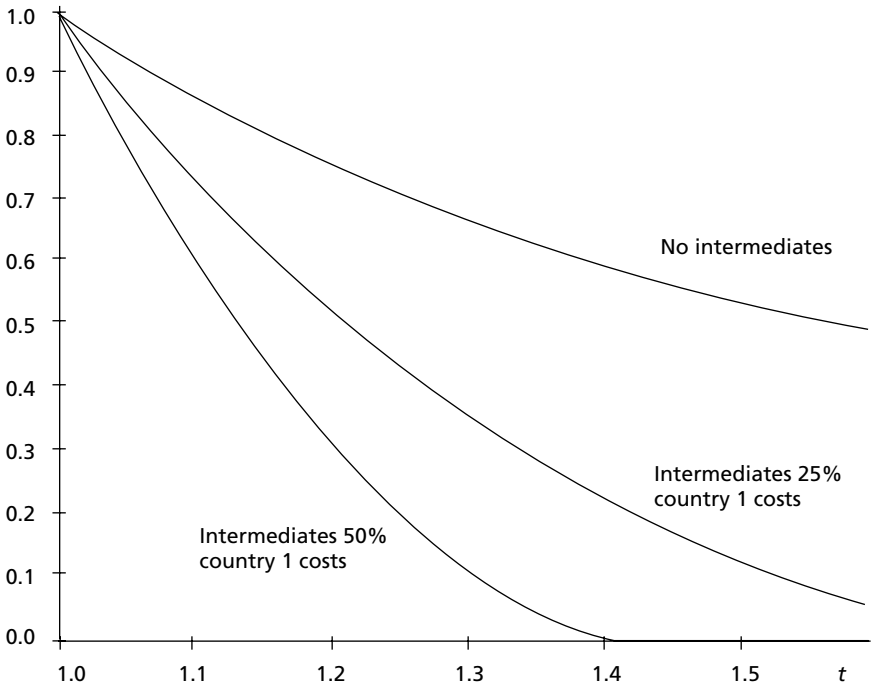
Note: Two-thirds of value added is labor and one-third is capital.
Source: Authors' illustration.

$MA_1 = \sum_j m_j t_{ij}^\theta$, replaces simple measures of transport costs. Recall that m_j measures the economic mass of an importer country, and t_{ij}^θ the rate at which its effect falls off with distance. MA_i is therefore a measure of country i's access to demand from all countries. It provides a generalization of the old idea of market potential (Harris 1954), which takes GDP as economic mass and the reciprocal of distance as the measure of spatial decay. Analogously, the supplier access of country i is defined as

$$SA_i = \sum_j s_j t_{ij}^\theta$$

where s_j represents the economic characteristics of exporting countries, such as manufacturing output, and SA_i can be used to measure country i's access to suppliers of intermediate goods. Thus a high value of SA_i means that country i is close to exporting countries and so has relatively cheap access to intermediate goods.

These concepts can now be used to express the rate of return to production in country i as a function of its wage and its market and supplier access:

(3) $r_i = R\left(w_i, \sum_j m_j t_{ij}^\theta, \sum_j s_j t_{ij}^\theta \right) = R\left(w_i, MA_i, SA_i \right).$

Suppose that economic activity locates in a manner that equalizes the rate of return across countries. Equation 3 then constitutes a set of equations linking each country's wage to its market and supplier access, generating an estimating equation of the form

(4) $$w_i = \alpha + \varphi_1 MA_i + \varphi_2 SA_i + u_i .$$

The final term, u_i, is an error term to which are assigned, for the moment, all other influences on wages. Redding and Venables (2000) estimate this relationship in a two-stage procedure using a cross-section of data on 101 developed and developing countries.[5] At the first stage a gravity trade model (equation 1) is used to estimate m_j, s_i, and θ, from which measures of market access and supplier access can be constructed for each country. The full specification of market access and supplier access requires including each country's own market and supply, as well as the effect of all foreign markets and suppliers. Here, only the foreign market access effects, $FMA_i = \sum_{j \neq i} m_j t_{ij}^{\theta}$, and foreign supplier effects, $FSA_i = \sum_{i \neq i} s_j t_{ij}^{\theta}$, are discussed (Redding and Venables [2000] deal with the full case).

At the second stage, equation 4 is estimated econometrically. Before looking at regression results, it is instructive to look at a scatter diagram that plots the log of foreign market access and the log of GDP per capita, a proxy for manufacturing wages (figure 2).[6] The diagram shows the importance of market access in determining wages—the empirical analogue of figure 1. Clearly, there is a strong positive association between foreign market access and per capita income. There are outliers such as Australia, Hong Kong (China), Japan, New Zealand, Singapore, and the United States. For two of these, Japan and the United States, sheer population mass means that domestic market and supplier access are extremely important relative to foreign access. For the rest of the sample the relationship holds within as well as between regions. Thus there is a European wage gradient passing from the core countries down through Spain and Portugal to Greece. And there is an Eastern European gradient, lying below the Western European gradient, indicating that these countries have lower per capita incomes than their location alone would justify. Similar gradients can be pulled out for other regions.

The regression results using these data are reported in table 2. The estimated coefficient for foreign market access alone is positive and highly statistically significant; foreign market access explains 35 percent of the cross-country variation in income per capita (column 1 in table 2). The results are similar for foreign supplier access alone (column 2). The theoretical specification calls for the inclusion of both market access and supplier access. Although separately identifying the coefficients on these two variables is difficult given the high degree of correlation between them, theory suggests a restriction across the two coefficients based on the relative shares of labor and intermediates in costs. The estimates here assume that the intermediate goods share of costs is 50 percent higher than the labor share (column 3). Once again, results are highly significant, with the measures explaining 36 percent of the variation in the cross-country income distribution.

Figure 2. Relationship of GDP per Capita and Foreign Market Access for 101 Countries

ln GDP per capita (US dollars)

ln FMA
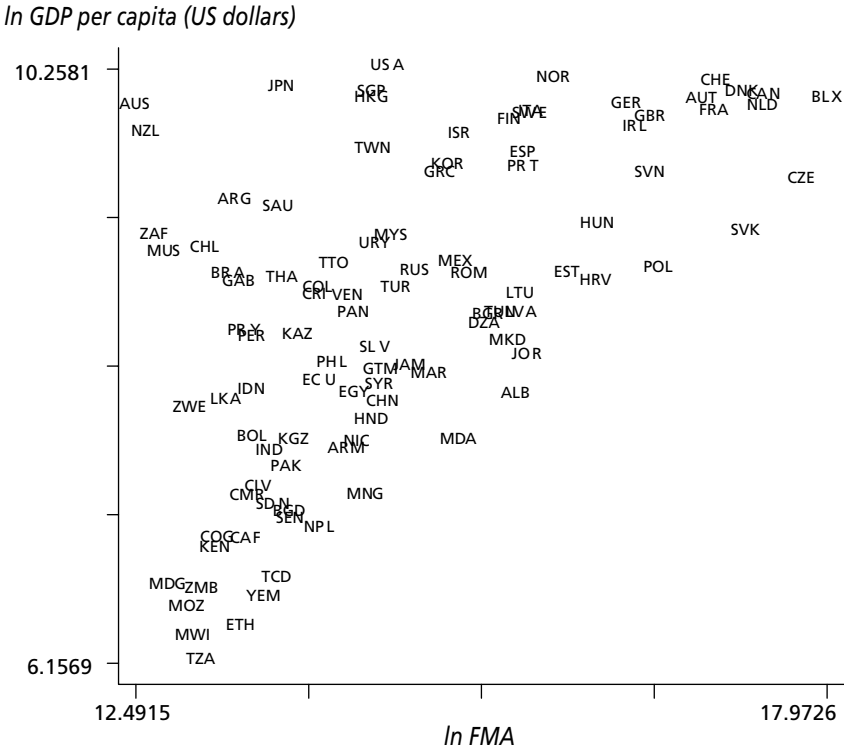
Note: See appendix table A.1 for list of countries.
Source: Authors' computations based on data from Redding and Venables (2000).

Of course, geography is not the only cause of cross-country variations in income. The final column of table 2 includes other variables, such as those used by Gallup, Sachs, and Mellinger (1999).[7] Endowments of hydrocarbons per capita have a positive and significant effect, as would be expected, while the proportion of land in the tropics has a negative but insignificant effect. Former socialist rule and involvement in external wars have negative and significant effects. Gallup, Sachs, and Mellinger (1999) have argued that malaria can have a pervasive productivity-reducing effect, and indeed the variable measuring the prevalence of malaria (a dummy variable taking a value of 1 in countries where malaria is endemic) has a significant, negative, and quantitatively important effect. Together with the foreign market access measure these variables explain around two-thirds of the cross-country variation in per capita income. From the current perspective, the main point is that the foreign market access measure remains highly significant, making the point that distance matters for per capita income, as suggested by the theory.

Table 2. Regression Results for GDP per Capita and Foreign Market and Supplier Access, 1996

Variable	(1)	(2)	(3)	(4)
Log foreign market access	0.476		0.319	0.277
	(0.076)			(0.063)
Log foreign supplier access		0.532	0.182	
		(0.114)	(0.040)	
Log hydrocarbons per capita				0.026
				(0.016)
Fraction of land in geographic tropics				−0.139
				(0.253)
Prevalence of malaria				−1.496
				(0.268)
Socialist rule, 1950–95				−0.743
				(0.156)
External war, 1960–85				−0.344
				(0.170)
R^2	0.346	0.377	0.361	0.671
$F(\cdot)$	52.76	57.05	54.60	55.63
Number of observations	101	101	101	99

Note: Dependent variable is log GDP per capita · First-stage estimation of the trade equation using Tobit. Numbers in parentheses are bootstrapped standard errors (200 replications).
Source: Author's computations, based on data from Redding and Venables (2000).

What Determines Distance Costs, and How Are They Changing?

The section above argues that geography is an important determinant of per capita income. Despite the presence of large cross-country wage differences, it is not profitable for firms to relocate, moving away from markets and suppliers. This section looks in more detail at the determinants of the costs of distance and at the effects of new technologies on these costs. This can best be addressed through the following thought experiment. A firm is considering where to source its supplies or where to locate its production. How is the decision to outsource to a low-wage economy deterred by distance, and how might information and communication technology mitigate this deterrent effect?

The distance effects can be divided into four main elements. First, making any sort of trade involves finding a trading partner, a process of search and matching that turns on the availability of information. Second, inputs and outputs have to be transported. These depend on country and commodity characteristics, and there is some evidence that they are changing. In "weightless" activities new technologies set these costs essentially at zero, but such activities amount to only a small percentage of total expenditure. Third, the supply chain has to be managed; for outsourced supply this involves a process of information exchange and monitoring, and for own investment it involves management of the entire project. And fourth, new technologies, by speeding up aspects of production and management, affect costs by affecting time in transit. This could either increase or decrease the benefits of proximity and the costs of distance.

Searching and Matching

A major reason why transactions fall off with distance is that firms simply know less about potential trades with people on the other side of the earth than about potential trades with their neighbors. Relatively little is known about the magnitude of these information barriers, although a number of researchers have attempted to establish their existence. For example, Rauch and Trindada (1999) use a gravity trade model to show how the existence of ethnic Chinese networks seems to increase trade volumes.

It seems likely that new technologies—the Internet in particular—significantly reduce search and matching costs. The Internet means that distance ceases to be important in advertising, and business-to-business exchanges facilitate search and matching across space. From my desktop a search engine will produce about 10,300 matches for the search string "garment+export+china+ltd", at least the first 10 of which are trading houses or Chinese firms offering supply. The most heavily researched examples of searching and matching through the Internet have a national rather than international focus. For example, in the U.S. automobile market in 1999 more than 40 percent of buyers used the Internet to seek out price and model information—although only 3 percent of sales were made on the Internet (Cairncross 2001, p. 113). This example makes a point that many dotcom companies have discovered and that surely applies even more in an international context: the Internet is excellent for acquiring information, but information is a necessary but by no means sufficient condition for completing a trade.

Moving Inputs and Outputs

An international transaction requires moving outputs and traded inputs across space. This can be done by different modes—surface, air, or for some activities, digitally. How large are these costs, and in what ways—and for how large a share of trade—are new technologies expected to reduce them?

Data on shipping costs indicate a very wide dispersion of transport costs across commodities and countries. Thus in 1994 freight expenditure was only 3.8 percent of the value of imports for the United States, but 7.3 percent for Brazil and 13.3 percent for Paraguay (Hummels 1999a, from customs data). These values incorporate the fact that most trade is with countries that are close, and in goods that have relatively low transport costs. Looking at transport costs unweighted by trade volumes gives much higher numbers; thus the median cif/fob ratio, across all country pairs for which data are available, is 1.28 (implying 28 percent transport and insurance costs). Looking across commodities, an unweighted average of freight rates is typically two to three times higher than the trade-weighted average rate.

Estimates of the determinants of transport costs are given in Hummels (1999b) and Limao and Venables (2001). These studies typically find elasticities of transport costs with respect to distance of 0.2–0.3. Limao and Venables find that a common border substantially reduces transport costs, that overland transport is seven times

more expensive than sea transport, and that being landlocked increases transport costs by approximately 50 percent. Infrastructure quality (as measured by a composite of index of transport and communications networks) is important; for example, while the median cif/fob ratio is 1.28, the predicted value of this ratio for a pair of countries with infrastructure quality at the 75th percentile rises to 1.40.

How are transport costs changing? In looking at the evolution of the costs of ocean shipping, air freight, and transmission of digitized information, there are three main points to notice (figure 3). First, the costs of sea transport declined during the 1940s and 1950s, but since then there has been no trend decline, although there have been substantial fluctuations driven largely by oil prices. This seems superficially surprising, but less so after noting that the variable reported is shipping cost relative to a goods price index. Thus there has been technical progress in shipping, but it has been no faster than the average in the rest of the economy. Second, the cost of air freight fell more and continued to fall for a longer period, but this too essentially bottomed out in the 1980s. Third, the most dramatic fall has been in the cost of transmitting digitized information, which can now be regarded as close to zero. For investigating international inequalities the important question is: what share of world expenditure is now "weightless" and can be digitized and transmitted at close to zero cost?

This question is hard to answer, because it is typically particular economic functions that can be digitized, rather than the whole production sectors that are the

Figure 3. Changes in Transport and Communication Costs, 1940–90
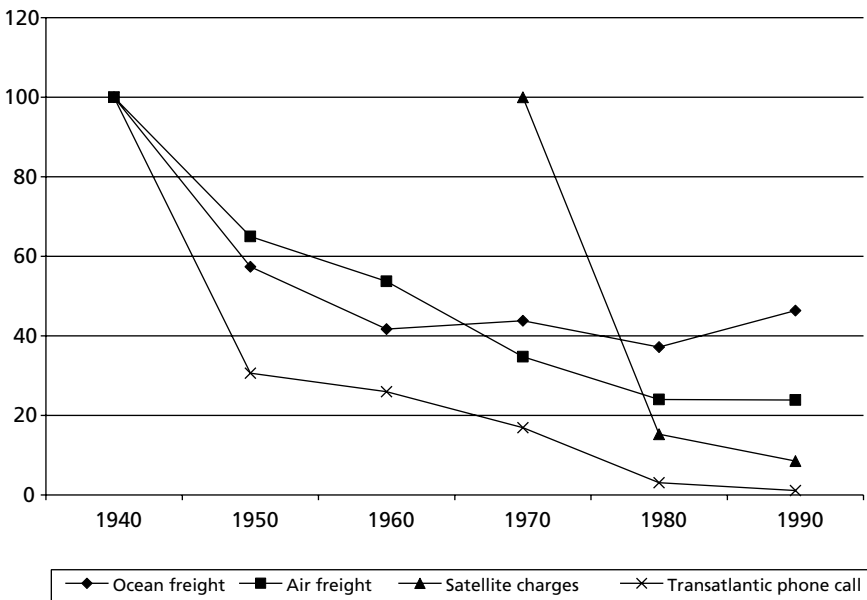
Source: Baldwin and Martin 1999.

basis for data collection. There are numerous examples of activities that have been digitized and relocated. Airline ticketing services and the backroom operations of banks are the standard ones. Call centers, architectural drawings, transcription of medical notes, and cartoons and computer graphics for the film industry are further possibilities.

One way to get a quantitative estimate is to look sectorally. These numbers look rather small. Consider U.S. household consumption of information and communication-based products and services. By 1998, 50 percent of Americans had a personal computer and 30 percent were regular Internet users. But total consumption of information and communication-based products and services, including voice telephony, was only 2.4 percent of consumer expenditure, a large part of which is devoted to upkeep of the network, a largely nontradable activity (Turner 2001). On the supply side the U.S. Bureau of Labor Statistics foresees employment in the information and communication industry growing from 3.7 percent of the U.S. total in 1998 to 4.9 percent in 2008, with the increase concentrated almost entirely in computer processing and software services (Turner 2001). The Organisation for Economic Co-operation and Development (OECD 1999) estimates that all software- and computer- related services in 1996 accounted for 2.7 percent of U.S. GDP—and half that in other OECD countries studied. Software products and computer services combined accounted for just 0.8 percent of U.S. exports in 1996.

Other sectors contain functions that are information technology enabled, such as call centers (customer interaction centers), medical transcriptions, and finance and accounting services. An estimated 17–24 percent of the cost base of banks can be outsourced ("Back office to the world" 2001), a share that seems low for an activity that is fundamentally weightless.

Another way to get a feel for the magnitude of these activities is to look at the recent experience of the highly successful Indian software and information technology-enabled services sectors. The total output of software and related services in India in 2000 was around $8 billion, of which exports were $4 billion. Information technology-enabled services had exports to the United States of $0.26 billion in 1999 that are predicted to grow to $4 billion by 2005 (*The Economist,* 5 May 2001). With total Indian exports of $45 billion in 2000, these are substantial activities for India but constitute less than 1 percent of total U.S. imports of around $950 billion.

Although it is difficult to quantify the share of the economy that is, or is likely to become, weightless, one fundamental point can be made. Codifying and digitizing not only enable an activity to be moved costlessly through space, but also typically lead to large productivity increases and price reductions. Thus the effect of information and communication technology on, say, airline ticketing has been primarily to replace labor by computer equipment and only secondarily to allow remaining workers to be employed in India rather than the United States or Europe. (Technology that can capture voice or handwriting will make Indian medical transcription obsolete.) This suggests that even if more activities become weightless, the share of world expenditure and employment attributable to these activities will remain slight—perhaps as little as a few percentage points of world GDP.

Monitoring and Management

Recent years have seen the rapid growth of both outsourcing and foreign direct investment, with the associated development of production networks or production chains.[8] Foreign direct investment has grown faster than either income or trade. The growth of production networks has been studied by a number of researchers. One way to measure its growth is by looking at trade in components. Yeats (1998) estimates that 30 percent of world trade in manufactures is trade in components rather than final products. Hummels, Ishii, and Yi (2001) chart trade flows that cross borders multiple times, as when a country imports a component and then reexports it embodied in some downstream product. For 10 OECD countries they find that the share of imported value added in exports rose by one-third between 1970 and 1990, reaching 21 percent of export value.

Both foreign direct investment and outsourcing involve, in somewhat different ways, a fragmentation of the structure of the firm, as production is split into geographically or organizationally different units. From an international perspective this fragmentation offers the benefits of enabling particular stages of the production process to be moved to the lowest-cost locations—labor-intensive parts to low-wage economies, and so on. However, as well as involving potentially costly shipping of parts and components, this fragmentation creates formidable management challenges. Product specification and other information has to be transferred, and production schedules and quality standards have to be monitored. Do new technologies reduce the costs of these activities?

To the extent that pertinent information is codifiable, the answer is likely to be yes. The use of information and communication technology for business-to-business trade is well documented, although it is reported to often reduce rather than increase the number of suppliers a firm uses.[9] In mass production of standardized products, designs can be relatively easily codified; the production process is routine, daily or hourly production runs can be reported, and quality data can be monitored. Dell Computers offers the classic example of the use of new technologies to outsource to order, getting components from suppliers at short notice. However, it is instructive that Dell's business practices, while held up as a model, have not been widely emulated ("A revolution of one" 2001). They work because personal computers are made almost entirely from standard parts, available from many sources, so there is no need to order special components in advance. And consumer customization falls within very narrow limits—speed and memory, but not color or trim. The product range and set of options are vastly less complex than those for a motor car.

For many activities the pertinent information cannot be codified so easily. There are two sorts of reasons for this. One is the inherent complexity of the activity. For example, frequent design changes or an ongoing process of product design and improvement (involving both marketing and production engineering) may require a level of interaction that—to date—can be achieved only by face-to-face contact.

The second reason has to do with the fact that contracts are incomplete, and people on each side of the contract (or in different positions within a single firm) have

their own objectives. It is typically expensive or impossible to ensure that these peoples' incentives can be shaped to be compatible with the objectives of the firm. This issue is the subject of an extensive economics literature. Part of the literature has its origins in analysis of the boundaries of the firm (Coase 1937), asking what transaction are best done within the firm and what by the market. Following Williamson (1975, 1985), this analysis is typically modeled as a tradeoff between the efficiency gains of using specialist suppliers (or suppliers in locations with a comparative advantage or low labor costs) and the problems encountered in writing (enforceable) contracts with them. Another part of the literature looks at the problems of incentives in organizations, asking how employees can be induced to meet their firm's objectives.[10]

While new technologies may reduce the costs of monitoring, it seems unlikely that the problems of incomplete contracts are amenable to a technological fix. What evidence is there? On the one hand there has been a dramatic increase in the outsourcing of activities to specialist suppliers in recent years, suggesting that difficulties in writing contracts and monitoring performance have been reduced. On the other hand a number of empirical studies point to the continuing importance, despite new technologies, of regular face-to-face contact. Gaspar and Glaeser (1998) argue that telephones are likely to be complements, not substitutes, for face-to-face contact as they increase the overall amount of business interaction. They suggest that, as a consequence, telephones have historically promoted the development of cities. The evidence on business travel suggests that as electronic communications have increased so too has travel, again indicating the importance of face-to-face contact. Leamer and Storper (2000) draw the distinction between "conversational" transactions (which can take place at a distance using information and communication technology) and "handshake" transactions, which require face-to-face contact. New technologies allow the dispersion of activities that require only conversational transactions but might also increase the complexity of the production and design process, increasing the proportion of activities requiring handshake communication.

Overall, then, it seems that there are some relatively straightforward activities for which knowledge can be codified, new technologies will make management from a distance easier, and relocation of the activity to lower-wage regions might be expected. But for more complex activities, monitoring, control, and information exchange still require contact that involves proximity and face-to-face meetings. Perhaps nowhere is this more evident than in the design and development of the new technologies themselves.

Time in Transit

New technologies provide radical opportunities for speeding up parts of the overall supply process. There are several ways this can occur. One is simply that basic information—product specifications, orders, invoices—can be transmitted and processed more rapidly. Another is that information about uncertain aspects of the supply process can be discovered and transmitted sooner. For example, retailers' electronic

stock control can provide manufacturers with real-time information about sales and hence about changes in fashion and overall expenditure levels. For intermediate goods improved stock controls and lean production techniques allow manufacturers to detect and identify defects in supplies more rapidly. These changes pose the interesting question, if some elements of the supply process become quicker, how does this affect the marginal value of time saved (or marginal cost of delay) in other parts of the process? In particular, if one part of the process that takes time is the physical shipment of goods, will time-saving technical changes encourage firms to move production closer to markets or allow them to move further away?

The importance of the costs of time in transit is highlighted by recent work by Hummels (2000), who analyzes some 25 million observations of shipments into the United States by air and sea (imports classified at the 10-digit commodity level by exporter country and district of entry to the United States for 25 years). Given data on the costs of each mode and the shipping times from different countries, Hummels is able to estimate the implicit value of time saved by using air transport. The numbers are quite large. The cost of an extra day's travel is around 0.3 percent of the value shipped for imports as a whole and 0.5 percent for manufacturing sectors. These costs are about 30 times larger than the interest charge on the value of the goods. One implication of these figures is that transport costs have fallen much more through time than suggested by freight charges alone. The share of U.S. imports going by air freight rose from zero to 30 percent between 1950 to 1998, and containerization approximately doubled the speed of ocean shipping, reducing shipping time by 26 days, equivalent to reducing shipping costs by 12–13 percent of the value of goods traded.

Given the magnitude of these costs, how might a time-saving technology influence the location of production? There are several mechanisms.[11] One relates to complementarity between shipping time and time for other parts of the supply process that are speeded up by information and communication technology. When activities are in parallel, with the total time taken set by the slowest part of the process, there is extreme complementarity. If information and communication technology speeds up one part of the process, the value of speeding up other parts is increased. Alternatively, when activities are strictly in sequence (say, transmitting information, followed by production and shipping), no complementarity is present. The total time taken is simply the sum of the parts. Reality is probably somewhere between these extremes, in which case there is some complementarity between activities, and information and communication technology increases the value of proximity.

A second mechanism is that the cost of time is increasing and concave in total time taken. This arises because of discounting. If information and communication technology speeds up the supply process, any further speeding becomes *more* valuable, as completion of the whole process occurs sooner and is therefore less heavily discounted. It also arises because information and communication technology may make it possible to react more quickly to changes in demand. Suppose that a firm produces a fashion-sensitive product, and under the old retail stock-control technology it was impossible to detect consumer response to this season's fashion until

it was too late to change production for the season. The firm produced all stock in advance and could do this in a low-wage location. Under the new retail stock control the firm can learn about sales and consumer response to fashion instantaneously and so has the opportunity to change production plans in line with fashion preferences. However, this opportunity is valuable only if production can be shipped to the market very rapidly. The new technology therefore causes production to move closer to the market, to exploit the advantage of more rapid market information.

An example is the highly successful Spanish clothing chain, Zara ("Floating on air" 2001). Zara uses real-time sales data, can make a new product line in three weeks (the industry average is nine months), and commits only 15 percent of production at the start of the season (the industry average is 60 percent). It also does almost all its manufacturing (from fabric dyeing through the full manufacturing process) in-house in Spain, with most of the sewing done by 400 local cooperatives (other firms in the industry engage in extensive outsourcing).

Other examples occur in intermediate goods supply, where new technology might make it easier to detect faults rather than to detect changes in fashion preferences. The supplier would then want to move production closer to the market and cut delivery times so that fewer faulty items were in the delivery chain.

Evidence comes from studies of just-in-time technologies, which have allowed much improved stock control and ordering, and a consequent movement of suppliers toward customers. In a study of suppliers to the U.S. automobile industry Klier (1999) finds that 70–80 percent of them are located within one day's drive of the assembly plant; any closer location is limited by the fact that many suppliers serve several assembly plants. He also finds evidence of increased concentration of supplier plants around assembly plants since 1980, a timing consistent with the introduction of just-in-time production methods. The leader in the application of just-in-time techniques is Toyota, whose independent suppliers are on average only 59 miles from its assembly plants, to which they make eight deliveries a day. By contrast, General Motors' suppliers in North America are an average of 427 miles away from the plants they serve and make fewer than two deliveries a day. As a result, Toyota and its suppliers maintain inventories one-fourth the size of General Motors', when measured as a percentage of sales (Taylor III and Kahn 1997).

Where Will Activities Move?

The preceding section suggests that information and communication technology will change the costs of distance in quite different ways for different types of activity. For many activities both face-to-face contact and proximity to markets or a cluster of related activities will remain important. These are activities for which complexity makes it difficult to codify information and write complete contracts, uncertainty makes rapid response to changing circumstances important, or transport costs remain important. Other sorts of activities can be fully digitized (the weightless activities) or may be sufficiently simple that information flows required in production control and monitoring can be codified and implemented remotely.

Activities for which face-to-face contact and proximity to markets or a cluster of related activities are important are likely to remain spatially concentrated—and their concentration might even increase because of complementarities in the value of time and the possibility of spatially separating some activities from more routine parts of the supply process. Suppose that financial services require both frontroom operations (which tend to cluster) and backroom operations (which are intensive in medium-skilled labor and office space). If the front- and backroom operations have to be located together, the overall clustering force might be quite weak—firms that are not in London, Tokyo, or New York lose out on the benefits of being in a cluster, but have the benefits of cheaper labor and office space. But once the backroom operations can be separated from the frontroom, the agglomeration forces on frontroom activities become overwhelming. All these activities will therefore be further concentrated by new technologies. It is perhaps to be expected then that financial services—in some ways a prime example of a weightless activity—are strongly concentrated in a few centers, with no prospect that technology will cause the dissolution of these centers.

What about the more routine and codifiable activities? These now have the possibility of moving out of established centers, but where will they go? They might spread rather evenly through many locations, bringing modest increases in labor demand in many countries. Or if there is some propensity for these activities to cluster, they might move to just a few countries. The propensity may be quite weak—the point is simply that as activities leave established centers in search of lower-wage locations, it is likely that a location that has some similar activities will look more attractive than one that has none.

The effects of reductions in trade costs in a world where manufacturing is internationally mobile but subject to some clustering forces can be illustrated with a variant of the new economic geography models of Fujita, Krugman, and Venables (1999). Suppose that there are many countries, arranged in a linear world with a well-defined center and pair of peripheries. Each country is identical (apart from in its location), being endowed with the same quantity of two factors of production (labor and land). There are two production activities. One is called "agriculture," although it can be interpreted as a wider aggregate of all the perfectly competitive sectors of the economy. This sector uses labor, land, and manufactures to produce a perfectly tradable output. The other sector is manufacturing. This sector uses labor and manufactures to produce manufactures, operating with increasing returns to scale in a monopolistically competitive market structure.

The manufacturing sector has forward and backward linkages, as manufacturing firms use inputs from other manufacturing firms and supply outputs to other manufacturing firms. These linkages encourage agglomeration, so that typically manufacturing operates only in the central locations, while peripheral locations are specialized in agriculture. The wage implications are illustrated in figure 4. At an initial position with high trade costs, low-wage countries have agriculture only, as do a corresponding set of countries on the other side of the center (concealed in the diagram). Wages are much lower in these countries than in industrialized countries,

Figure 4. Trade Costs, Real Wages, and Location
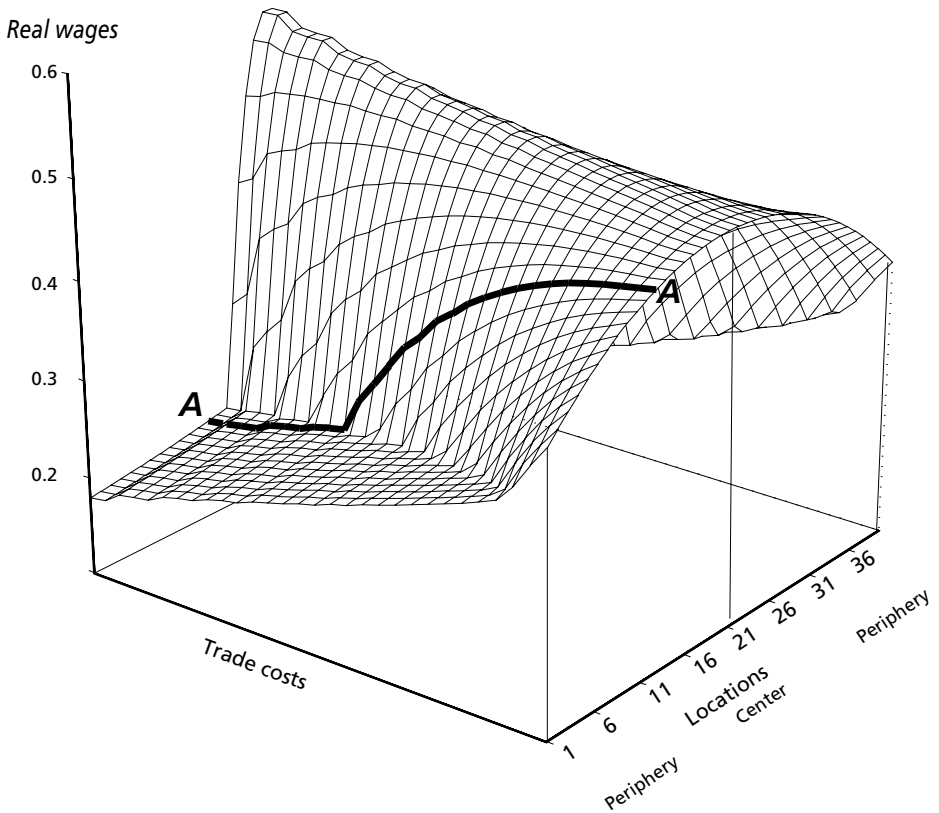
Source: Authors' illustration.

and wages peak in the central region, which has the best market access and best supplier access.

At lower trade costs (a movement to the right in figure 4) it becomes profitable for some firms to relocate to lower-wage countries. But these countries are relatively close to established centers, and as they attract industry, a process of cumulative causation commences. Forward and backward linkages between firms in the country mean rapid "takeoff" for these countries, as indicated by the steepness of the wage gradient. The bold line AA illustrates the wage path of a country located midway between center and edge as transport costs fall. This country is initially in the periphery, with no manufacturing and low wages, but lower trade costs cause manufacturing to spread away from the center, industrializing this country and causing the rapid wage growth illustrated.

The point of this example is that even for activities that can relocate from established centers, the presence of (weak) agglomeration forces means that they will move to just a subset of possible new locations. As a consequence some coun-

tries will experience a rapid increase in labor demand and wages, while others remain in the periphery, essentially untouched by the process. New technologies change the pattern of inequalities in the world economy but do not uniformly decrease them.

The predictions of this theoretical model seem to be broadly in line with what we know about recent sectoral relocations. Much software production has left the United States—but only to concentrate largely in Ireland and Bangalore. At a broader level there has been growth of production networks, with production of components outsourced to lower-wage countries. But this growth of vertical specialization and parts and components trade is concentrated in a few countries neighboring existing centers—in Asia, Europe, and America.

The growth of trade in production networks and its geographica concentration are illustrated in table 3, which looks at countries' exports of telecommunications equipment (final equipment and parts and components), a set of commodities for which outsourcing to lower-wage countries has grown rapidly. The 68 countries in the sample are divided according to initial (1983–85) per capita incomes. The share of low-income countries in world trade in telecommunications equipment rose from 5 percent in the early 1980s to 19 percent in the late 1990s. The share of telecommunications in exports exhibits a very skewed distribution: telecommunications equipment production and trade have become important for just a few low-income countries (they account for more than 10 percent of exports for one country and 6.6–10 percent for another) while remaining unimportant for the vast majority. This pattern is repeated in other sectors, generally with the same set of countries being the main exporters.

Table 3. Exports of Telecommunications Equipment, Final and Parts, 1983–85 and 1995–97
(number of countries)

	1983–85			1995–97		
Share	Low-income countries	Middle-income countries	High-income countries	Low-income countries	Middle-income countries	High-income countries
Telecommunications' share in country's exports						
<3.33 percent	36	9	14	32	7	11
3.3–6.6 percent	1	1	3	3	4	5
6.6–10 percent	0	2	1	1	1	2
>10 percent	0	0	1	1	0	1
Countries' share in all telecommunications exports (percent)						
	5	12	83	19	11	70

Source: UN Comtrade database.

Conclusions

Speculating on the implications of new technology is a notoriously risky activity. Still, the analysis here suggests several conclusions. Some activities will become more deeply entrenched in high-income countries—and typically in cities in these countries. These activities will generally be complex—knowledge intensive, rapidly changing, and requiring face-to-face communication. But they will also include the supply of nontradables and of produced goods for which shipping is costly or time consuming. Other activities that are more readily transportable and less dependent on face-to-face communications may relocate to lower-wage countries, and this will be an important force for development. However, since these activities may cluster, development is likely to take the form of rapid advance for a small number of countries (or regions) rather than a more uniform process of convergence. Although new technologies facilitate the relocation of these activities, the proportion of world GDP that can "operate as though geography has no meaning" (Cairncross 2001) is likely to be small.

New technologies will not mean the death of distance, but the contribution of these technologies to economic development will nevertheless be important. It will come primarily from allowing individuals greater access to knowledge, education, and basic services, not from rewriting the rules of economic geography.

Appendix Table A.1 Economies in Figure 2 And Table 2

1. Albania (ALB)	28. Estonia (EST)	55. Morocco (MAR)	82. Singapore (SGP)
2. Argentina (ARG)	29. Ethiopia (ETH)	56. Moldova (MDA)	83. El Salvador (SLV)
3. Armenia (ARM)	30. Finland (FIN)	57. Madagascar (MDG)	84. Slovak Republic (SVK)
4. Australia (AUS)	31. France (FRA)	58. Mexico (MEX)	85. Slovenia (SVN)
5. Austria (AUT)	32. Gabon (GAB)	59. Macedonia (MKD)	86. Sweden (SWE)
6. Bangladesh (BGD)	33. United Kingdom (GBR)	60. Mongolia (MNG)	87. Syria, Arab Rep. (SYR)
7. Bulgaria (BGR)	34. Greece (GRC)	61. Mozambique (MOZ)	88. Chad (TCD)
8. Belgium/Luxembourg (BLX)	35. Guatemala (GTM)	62. Mauritius (MUS)	89. Thailand (THA)
9. Bolivia (BOL)	36. Hong Kong China (HKG)	63. Malawi (MWI)	90. Trinidad & Tobago (TTO)
10. Brazil (BRA)	37. Honduras (HND)	64. Malaysia (MYS)	91. Tunisia (TUN)
11. Central African Republic (CAF)	38. Croatia (HRV)	65. Nicaragua (NIC)	92. Turkey (TUR)
12. Canada (CAN)	39. Hungary (HUN)	66. Netherlands (NLD)	93. Taiwan, China (TWN)
13. Switzerland (CHE),	40. Indonesia (IDN)	67. Norway (NOR)	94. Tanzania (TZA)
14. Chile (CHL)	41. India (IND)	68. Nepal (NPL)	95. Uruguay (URY)
15. China (CHN)	42. Ireland (IRL)	69. New Zealand (NZL)	96. United States (USA)
16. Côte d'Ivoire (CIV)	43. Israel (ISR)	70. Pakistan (PAK)	97. Venezuela (VEN)
17. Cameroon (CMR)	44. Italy (ITA)	71. Panama (PAN)	98. Yemen, Rep. of (YEM)
18. Congo, Rep. Of (COG)	45. Jamaica (JAM)	72. Peru (PER)	99. South Africa (ZAF)
19. Colombia (COL)	46. Jordan (JOR)	73. Philippines (PHL)	100. Zambia (ZMB)
20. Costa Rica (CRI)	47. Japan (JPN)	74. Poland (POL)	101. Zimbabwe (ZWE)
21. Czech Republic (CZE)	48. Kazakhstan (KAZ)	75. Portugal (PRT)	
22. Germany (DEU)	49. Kenya (KEN)	76. Paraguay (PRY)	
23. Denmark (DNK)	50. Kyrgyz Republic (KGZ)	77. Romania (ROM)	
24. Algeria (DZA)	51. Korea, Rep. of (KOR)	78. Russia (RUS)	
25. Ecuador (ECU)	52. Sri Lanka (LKA)	79. Saudi Arab. (SAU)	
26. Egypt (EGY)	53. Lithuania (LTU)	80. Sudan (SDN)	
27. Spain (ESP)	54. Latvia (LVA)	81. Senegal (SEN)	

Notes

1. To identify the channels through which technical knowledge is transmitted, Keller (2001) investigates not just distance between countries, but also the volume of trade between them, their bilateral holdings of foreign direct investment, and their language skills (the share of the population in country i that speaks the language of country j). Adding these variables renders simple geographic distance statistically insignificant; around two-thirds of the difference in bilateral technology diffusion is accounted for by trade patterns, and one-sixth each by foreign direct investment and language. However, all these variables are themselves declining with distance.

2. Of course, there are many intermediate goods, but here their prices are summarized in a single price index, q.

3. This is a trade cost factor, thus $t = 1.2$ means that trade costs are 20 percent of the value of goods shipped.

4. Technologies are assumed to be the same in all countries—geography is the only source of difference between countries.

5. They also derive the wage equation and the market access and supplier access from economic fundamentals, based on Fujita, Krugman, and Venables (1999).

6. A similar pattern is observed using data on manufacturing wages per worker. See Redding and Venables (2000) for further details.

7. Here only variables that can be reasonably regarded as exogenous are used, so, for example, measures of countries' human or physical capital stock are not included.

8. A good example of outsourcing is Nortel Networks, a Canadian company that specializes in high-performance communications networks. In 1998 it sold off its production plants to separate companies with which it now has long-term contracts, in order to concentrate on network installation and on production of the most sophisticated components (Cairncross 2001, p. 150).

9. British Airways expects to reduce the number of suppliers from 14,000 to around 2,000 as it implements on-line procurement (Cairncross 2001, p. 38).

10. See Holmstrom and Roberts (1998) and Gibbons (1998) for surveys of these two areas.

11. A simple model of this is presented in a fuller version of this article, available at http://econ.lse.ac.uk/staff/ajv.

Bibliography and References

"A revolution of one." 2001. *The Economist*, April 12.

Baldwin, R. E, and P. Martin. 1999. *Two Waves of Globalization: Superficial Similarity and Fundamental Differences*. NBER Working Paper 6904. Cambridge, Mass.: National Bureau for Economic Research.

Cairncross, F. 2001. *The Death of Distance 2.0: How the Communications Revolution Will Change Our Lives*. Cambridge, Mass.: Harvard Business School Press.

Coase, R. H. 1937. "The Nature of the Firm." *Economica* 4: 386–405.

Di Mauro, F. 2000."The Impact of Economic Integration on FDI and Exports: A Gravity Approach." CEPS Working Document 156. Center for European Policy Studies, Brussels.

Dickens, P. 1998. *Global Shift: Transforming the World Economy*. London: Chapmans.

Feenstra, R. C. 1998. "Integration of Trade and Disintegration of Production in the Global Economy." *Journal of Economic Perspectives* 12: 31–50.

"Floating on air." 2001. *The Economist*, April 19.

Fujita, M., P. Krugman, and A. J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, Mass.: MIT press.

Gallup, J. L., and J. Sachs, with A. D. Mellinger. 1999. "Geography and Economic Development." In B. Pleskovic and J. E. Stiglitz, eds., *Annual World Bank Conference on Development Economics 1998.* Washington, D.C.: World Bank.

Gasper, J., and E. L. Glaeser. 1998. "Information Technology and the Future of Cities" *Journal of Urban Economics* 43: 136–56.

Gibbons, R. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12: 115–32.

Grossman, G. M., and E. Helpman. 1999. *Incomplete Contracts and Industrial Organization.* NBER Working Paper 7303. Cambridge, Mass.: National Bureau for Economic Research.

Harris, C. 1954. "The Market as a Factor in the Localization of Industry in the United States." *Annals of the Association of American Geographers* 44: 315–48.

Holmstrom, B., and J. Roberts. 1998. "The boundaries of the Firm Revisited." *Journal of Economic Perspectives* 12: 73–94.

Hummels, D. 1999a. "Have International Transportation Costs Declined?" University of Chicago, Graduate School of Business. Mimeo.

———. 1999b. "Toward a Geography of Trade Costs." University of Chicago, Graduate School of Business. Mimeo.

———. 2000. "Time as a Trade Barrier." Purdue University, Department of Economics, Indianapolis, Ind. Mimeo.

Hummels, D., J. Ishii, and K.-M. Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics* 54: 75–96.

Keller, W. 2001. *The Geography and Channels of Diffusion at the World's Technology Frontier.* NBER Working Paper 8150. Cambridge, Mass.: National Bureau for Economic Research

Klier, T. H. 1999. "Agglomeration in the US Auto Supply Industry." *Economic Perspectives* 23: 118–34.
[http://www.chicagofed.org/publications/economicperspectives/1999/ep1Q99_2.pdf].

Leamer, E., and M. Storper. 2000. "The Economic Geography of The Internet Age." University of California, Los Angeles, Department of Management.

Limao, N., and A. J. Venables. 2001. "Infrastructure, Geographical Disadvantage, Transport Costs and Trade." *World Bank Economic Review* 15: 451–79.

Marshall, A. 1920 [1890]. *Principles of Economics.* 8th ed. London: Macmillan.

OECD (Organisation for Economic Co-operation and Development). 1999. "The Software Sector: A Statistical Profile for Selected OECD Countries." Paris.

"Outsourcing to India." 2001. *The Economist*, May 5.

Porter, M. E. 1990. *The Competitive Advantage of Nations.* London: Macmillan.

Portes, R., and H. Rey. 1999. "The Determinants of Cross-Border Equity Flows." CEPR Discussion Paper 2225. Centre for Economic Policy Research, London.

Rauch, J. E. and V. Trindada. 1999. *Ethnic Chinese Networks in International Trade.* NBER Working Paper 7189. Cambridge, Mass.: National Bureau for Economic Research.

Redding, S., and A. J. Venables. 2000. "Economic Geography and International Inequality." CEPR Discussion Paper 2568. Centre for Economic Policy Research, London.

Taylor III, A., and Kahn, Jeremy. 1997. "How Toyota Defies Gravity." *Fortune*, December 8.

Turner, A. 2001. *Just Capital: The Liberal Economy*. London: Macmillan.

Williamson, O. E.. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: Free Press.

———. 1985. *The Economic Institutions of Capitalism.* New York: Free Press

Yeats, A. 1998. "Just How Big Is Global Production Sharing?" Policy Research Working Paper 1871. World Bank, Development Economics Research Group, Washington, D.C.

Comment on "Geography and International Inequalities: The Impact of New Technologies" by Anthony J. Venables

Andrés Rodríguez-Clare

Anthony Venables's article offers clear thinking in an area saturated with speculation. The common wisdom is that falling transportation costs reduce the importance of distance. But things are much more complicated than that. True, falling transportation costs make it less costly for manufacturers to locate far from consumers. This could favor developing countries, since manufacturing could locate in developing countries to benefit from lower wages. However, manufacturing firms may prefer to remain in locations with high wages, to benefit from the agglomeration economies that arise from the networks of specialized input suppliers (*clusters)* found in such locations.

The question is thus whether new technologies that lower communication costs weaken agglomeration economies and thus lead manufacturing firms toward developing countries. This is precisely the question Venables examines, in a simple and useful way.

I have a few specific remarks. First, Venables presents data that appear to show that distance matters. For instance, in a regression of income per capita on measures of foreign market access and foreign supplier access (where distance is the key variable), these two access variables explain 36 percent of the variation in income per capita. But it seems important to know the results, as well, when other variables are included in the regression (see equation 5 in table 2).

In any case there should be at least some discussion of alternative explanations for the same phenomenon. For example, Acemoglu, Johnson and Robinson (2000) argue that today's geographic pattern of development reflects not the importance of distance, but—at least for an important set of countries—the way institutions developed in the colonial period. Along the same lines it seems important to verify that other implications of the distance-matters theory are consistent with the data. For example, is it true that distant countries, other things being equal, specialize in lighter and less bulky goods or in goods that use intermediate goods less intensively? Is it true

Andrés Rodríguez-Clare is chief, Council of Presidential Advisors, in Costa Rica. He was previously associate professor of business economics, Graduate School of Business, University of Chicago.

that for a given good, distant countries (those with lower supplier access) produce with a lower percentage of imported intermediate goods than countries with better supplier access? Are the implications of this model for trade between developing countries consistent with the data?

Second, before embarking on an analysis of the way new technologies affect the importance of distance, it would be useful to have a description of those new technologies and how they are being implemented at the micro level. Are we talking about the rise of the Internet, the falling cost of telecommunications, the advance of video conferencing, or wireless e-mail—or all of these technologies? Presumably, different technological advances will have different impacts on the importance of distance for development.

Third, it would be interesting to see whether the aggregate data show that distance is becoming more or less important. Certainly the lack of convergence across countries in the last decades does not appear to support the "death of distance." But Keller (2001) presents empirical evidence (although only for high-income countries) suggesting that distance had less of an effect in limiting the diffusion of technology in 1983–95 than in 1970–82. It would also be interesting to see whether economic activity is becoming more or less concentrated in the United States, for example, using the "dart board" approach proposed by Ellison and Glaser (1997).

Fourth, it seems important to note that even if new technologies make distance less important as a determinant of economic development, some developing countries could be negatively affected, at least during the transition. For example, the Central American countries could lose the advantage they now enjoy from being close to the U.S. market, facing stronger competition in that market from low-wage countries like China.

Finally, it seems important to discuss the policy implications of Venables's findings. At a first level, the advice that developing countries should promote investment in infrastructure (telecommunications, ports, roads, airports, railroads) seems to get additional support. At another level Venables suggests that countries should not expect automatic benefits from the new technologies. If developing countries continue doing the same things, not much will change. Venables suggests that developing countries will experience fast growth only if they implement policies that allow industrial clusters to develop.

References

Acemoglu, D., S. Johnson, and J. A. Robinson. 2000. *The Colonial Origins of Comparative Development: An Empirical Investigation*. NBER Working Paper 7771. Cambridge, Mass.: National Bureau of Economic Research.

Ellison, G., and E. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105: 889–927.

Keller, W. 2001. *The Geography and Channels of Diffusion at the World's Technology Frontier.* NBER Working Paper 8150. Cambridge, Mass.: National Bureau of Economic Research.

This Page Intentionally Left Blank

Health and

Development

This Page Intentionally Left Blank

The Role of Nongovernmental Organizations in Providing Health Care

Morten Rostrup

Nongovernmental organizations (NGOs) are a very heterogeneous group. Though generally considered a part of civil society, many receive substantial government funding. Some people argue that such arrangements tie humanitarian action to the foreign policy agenda of governments, in effect making such NGOs government subcontractors. Increasingly, NGOs are expected to work within the long-term development perspectives and peace-building activities of various "strategic frameworks." Médecins sans Frontières tries to distance itself from this way of thinking, believing that humanitarian action should be independent of initiatives that advocate models for development or social change. Using humanitarian action to promote peace and democracy or a free market-neoliberal agenda threatens a fundamental principle: humanitarian aid should depend solely on need, not on political agendas.

Nor should NGO assistance compensate for a failure to seek the political solutions needed to achieve real progress in health and society. Attention to poverty alleviation tends to focus on economic inequity but not on the political role in creating and sustaining inequity. Humanitarian assistance must not be manipulated to support a system that gave rise to the misery in the first place. Providing humanitarian assistance to populations in distress thus also involves moral dilemmas. One reason people die of diseases like AIDS, tuberculosis, and malaria is that life-saving medicines are too expensive. National and international authorities have abdicated responsibility in favor of the pharmaceutical industry. NGOs have a clear role in pushing for change and political responsibility.

Thus the fact that Médecins sans Frontières delivers health care in the world today is a sign of serious failure and suggests several questions: Should NGOs compensate for the state's retreat? Is it really an NGO's role to provide health care as part of a permanent solution? Aren't NGO activities lacking in continuity and coordination? Isn't the state's legitimacy being eroded by privatization of fundamental public services? And what does such a role for NGOs do to their transparency and accountability?

Nongovernmental organizations (NGOs) come from different national political cultures and belief systems and face different institutional pressures. Some have multiple mandates. Their world views, responsibilities, capacities, roles, and actions all differ. Some have a clear mission to build up certain systems and promote sustainable development and peace, based on specific political analysis. Some work from the bottom up, while others contract with their own national governments or are work as implementing partners for the United Nations system. Many NGOs operate only in their own country, while others are international. Thus NGOS are a very heterogeneous group.

Using Humanitarian Organizations to Achieve Other Objectives

The term used for these organizations—*nongovernmental organizations*—describes not what they are but what they are not. They are not businesses. They are viewed as voluntary and nonbureaucratic, as representing the private initiative of citizens taking affairs into their own hands, contesting the state, and holding it accountable. NGOs are also seen as providing an alternative to the state, using a self-help initiatives to fill in gaps left by the deficiencies of the state.

But the extent to which NGOs really constitute a part of civil society differs considerably. Some NGOs are genuine independent initiatives of civil society. They have a clear social basis, and they try to maintain their independence from the state. Other NGOs have emerged from political parties or other power groups. Even the classic humanitarian NGOs differ. They share the common objective of alleviating the suffering of victims of conflict, marginalization, discrimination, or oppression, but there are differences in ideology. Their implementation of humanitarian principles of neutrality and impartiality may differ, as may their modes of assistance of vulnerable populations. Based on the experience of its founding doctors during the Biafra crisis, Médecins sans Frontières has put the right of all people to medical assistance above concerns of state sovereignty. Oxfam defends justice in its operations, based on its early experience during the British blockade of Greece in 1942. CARE focuses more on technical aspects of aid.

Another important characteristic of NGOs today is their financial dependence on official donors. If an NGO receives 90 percent of its funding—or anything more than half—from government sources, can it be called "nongovernmental"? Some NGOs argue that they can maintain their operational independence despite substantial government support. Others argue that government funding necessarily links humanitarian action to the foreign policy agenda of governments and that such NGOs are really government subcontractors. Some NGOs argue, however, that economic ties are an efficient vehicle for lobbying and information sharing. From Médecins sans Frontières' perspective such financial dependence, along with increased demands for strict U.N. coordination in the field and implementation of strategic frameworks, threatens a core identity of humanitarian NGOs: their independence of action. The narrow, short-term view of humanitarian action as trying to preserve life and alleviate suffering while protecting human dignity is perceived

as politically incorrect. Criticism of this approach comes from both politicians and solidarity movements. Increasingly, humanitarian action is supposed to work under a certain framework of long-term development perspectives and peace-building activities.

Sierra Leone is a good example of this blurred vision of humanitarian assistance. A coordinator of the U.N. Office for the Coordination of Humanitarian Affairs (OCHA) argues in the "Consolidated Inter-Agency Appeal 2001 for Sierra Leone" (OCHA 2001, p. 1) that humanitarian assistance should "contribute towards lasting peace and economic development.…We must build an army of genuine humanitarians who will help to disseminate value systems crucial for the success of any peace process." Later the appeal comments that "U.N. agencies—working closely with [the National Commission for Reconstruction, Resettlement, and Rehabilitation], Government line ministries and non-governmental partners as well as [the United Nations Mission in Sierra Leone]—remain committed to not only providing immediate relief to the population, but also to investing time and resources in creating the conditions for a return to normalcy" (p. 13).The appeal also states that the humanitarian agencies agree that "assistance will be provided within the context of efforts to achieve sustainable peace" (p. 19).

The increasing use of humanitarian action as a first step in promoting peace and democracy threatens the principle that humanitarian aid should be provided solely on the basis of need not political agendas, whether national or multinational, as in the "strategic frameworks" of the United Nations. In this perspective humanitarian action should exist and be financed only as long as it contributes to higher political objectives such as peace, respect for human rights, or the promotion of good governance and democracy. Such initiatives all share a desire to obtain secondary benefits from humanitarian action. They force a system of principles and priorities on humanitarian action that is foreign to it. Humanitarian assistance must be provided to those in need without conditions. The beneficiaries, who suffer from intense need, have a right to such assistance, and the absence of conditions avoids any need to negotiate with those in need.

However, humanitarian action may have political effects without having political intent. Increasingly, humanitarian action is being used as an instrument of foreign policy to imply to constituencies that governments are active when they are not, that their actions are moral by association, or that they are pursuing other goals through humanitarian action, such as crisis containment or curtailment of refugee flows. Independent humanitarian actors are not opposed to political actors meeting the needs of their constituencies. But this must be done openly and transparently and through the proper channels.

Attempts to use humanitarian action to promote free market neoliberalism are increasing as well. A prominent example is the response of the international community to the socioeconomic crisis in Africa that began in the 1980s, prompted in part by the collapse of export prices (primary commodities) and the rising cost of essential imports such as oil. Recognizing that the state-led development model introduced after independence in the 1960s was failing, the International

Monetary Fund and the World Bank introduced structural adjustment programs to restore economic growth. The underlying philosophy was neoclassical liberalism, with its belief that private economic forces competing in free markets can maximize both individual benefits and public welfare. The ensuing rollback of the state was not, however, restricted to economic affairs, but extended to budget cuts affecting social services.

Critics of structural adjustment sharply denounced the state's withdrawal and cutbacks in social services expenditures. During the 1990s the World Bank seemed to acknowledge that the state should continue to play a vital role in socioeconomic development, discarding the extreme position that a smaller state is necessarily a better state. *World Development Report 1997: The State in a Changing World* called on the state to fulfill at least five fundamental tasks: establish a foundation of law, maintain sound economic policies, invest in social services and infrastructure, protect the vulnerable, and protect the environment, without which "sustainable, shared, poverty-development is impossible." However, despite the change in policy, problems with social sector funding continued. Helen Keller International reported growing health problems, including anemia and malnutrition, after Indonesia adopted a structural adjustment program during the economic crisis in 1997. Vaccinations for common childhood diseases were found to be too costly for poor families.

Along with these policy changes came an increased focus on poverty eradication, as the key objective of liberal economic reform efforts, and social services, as a necessary investment to increase productivity and combat poverty. And civil society—generally seen as comprising a free media, civic and nongovernmental organizations, trade unions, and possibly political parties—has taken on a new prominence.

But even with social services again recognized as a core responsibility of the state, requiring increased attention and funding, many concerns remain. Reducing and redefining the state's role and increasing the participation of NGOs and the local population still motivates many national governments and international organizations.

Attention to poverty alleviation tends to focus on economic inequity while ignoring the political role in creating and sustaining inequity and the need for political solutions—beyond good policy—to generate real progress in health and society. NGOs cannot compensate for this broader failure by providing technical or material assistance. They can shine a bright light on the need to politicize the understanding of human society, conflict, progress, and equity—in which economic analysis plays an important but not exclusive role.

For Médecins sans Frontières several crucial questions emerge from this background: Should NGOs compensate for the state's retreat, especially in social services? Is it really an NGO's role to provide health care as part of a permanent solution? Aren't NGO activities fragmented, lacking in continuity and coordination? Isn't the state's legitimacy being eroded by privatization of fundamental public services? And what does such a role mean for NGO transparency and accountability to beneficiaries?

Médecins sans Frontières—A Medical Humanitarian Organization

Médecins sans Frontières is a medical humanitarian organization. But humanitarian action is more than simple generosity. Médecins sans Frontières seeks to relieve suffering, to restore autonomy, to bear witness to injustice, and to insist on political responsibility. It strives to enable individuals to regain their rights and dignity as human beings. Toward these ends Médecins sans Frontières also works to provoke change and reveal injustice.

Thus Médecins sans Frontières is not just a service provider. It is not trying to replace political and local responsibility for developing political society and providing welfare services. On the contrary, it is trying to demonstrate the failure of the state to fulfill its responsibilities. Médecins sans Frontières sets up services and derives lessons for the construction of new models of service delivery, but it does not try to cover part of the national service network or to ensure access to complete and equitable services for some segment of the population outside of a national framework. It is not part of the liberal economic agenda to replace government responsibilities with private responsibilities. Rather, it tries to stimulate government and international responsibility. The fact that Médecins sans Frontières delivers health care in the world today is a sign of serious political failure, nothing else.

The definition Médecins sans Frontières uses to describe its framework was formulated in the early 1990s by Rony Brauman, who said that "Médecins sans Frontières helps the members of a society to survive a period of crises—defined as a disturbance of a previously existing equilibrium" (personal communication). This definition intentionally distanced Médecins sans Frontières from a more development-oriented way of thinking and from initiatives that promote models for development or for society.

The core activities of Médecins sans Frontières are practical work in the field. This is an effective approach for relief operations, and it is contrary to what might be termed a globalizing approach. It is not the mission of Médecins sans Frontières to eradicate poverty, as it is that of the World Bank ("To fight poverty with passion and professionalism for lasting results") or Oxfam ("To work with others to overcome poverty and suffering"), for example. The poverty objective for Médecins sans Frontières is to depoliticize it. Poverty is a lack not only of resources but also of political capital and voice. Humanitarian action does not problematize poverty or respond to poverty. Humanitarian action problematizes and responds to suffering—and explicitly recognizes the "abnormality" of that suffering and the duty of all human beings to respond to that suffering.

Powerful institutions like the World Bank, the International Monetary Fund, and the World Health Organization see the eradication of extreme poverty within 10 years as the foundation on which "Health for All" will be built. Thus, the *health problem* of today is defined as the absence of economic development, rather than individuals' lack of access to effective treatment. Major pharmaceutical companies seem to concur. While waiting for this prosperous future to arrive, there is apparently no need to offer effective pharmaceuticals at reasonable prices since the eco-

nomic conditions for using them are not in place. For Médecins sans Frontières this attitude represents an unacceptable response that will lead to the decline of medicine. It cannot accept a neoliberal order that excludes, marginalizes, and sacrifices the lives and dignity of millions of people in the name of some future economic benefit that will "trickle down" to the poor, given enough time. As long as there are patients in the field, Médecins sans Frontières has to insist on proper treatment.

Médecins sans Frontières, therefore, has a clear duty to struggle for the quality of care on the ground. The obligations of its doctors are clearly defined by the circumstances of their patients. Most patients in the developing world have few choices. Their precarious position results from the indifference, marginalization, discrimination, and violence they are subjected to. They do not have the luxury of choosing a new doctor if the first fails to meet their needs. It is therefore alarming to see that doctors themselves have increasingly internalized the failures of medicine, reducing their aspirations by accepting these constraints. They do not demand more. A medical humanitarian organization like Médecins sans Frontières must not fall into this trap or allow itself to become passive. It must hold firm to its mission: to provide quality health care, and to do it today, for those who need it most. Who will provide effective medicine and treatment to the poorest of patients if their own doctors do not demand it?

Consider the HIV/AIDS pandemic. The draft declaration for the special June 2001 U.N. General Assembly session on the HIV/AIDS pandemic clearly states that "prevention must be the mainstay of our response" (point 47). Treatment of people with HIV/AIDS is not given the same priority. Médecins sans Frontières considers this unacceptable. All people have the right to adequate medical care. HIV/AIDS is a medical condition, and life-prolonging and life-saving treatment exists. This treatment is feasible today, even in resource-limited settings. Prevention and treatment are mutually dependent and inherently linked, and no one is served by pitting one against the other. Winning the battle against the HIV/AIDS pandemic requires a global commitment from countries to comprehensive programs that provide a continuum of care, including fully integrated prevention and treatment. This goal should clearly be spelled out in the U.N. declaration of commitment.

Médecins sans Frontières is best known for its emergency interventions, and assisting the victims of natural catastrophes, epidemics, and armed conflicts is still an important part of its work. It is active today in more than 400 projects in almost 90 countries, relying on some 3,000 international volunteers a year and more than 15,000 local staff.

In January 2001 in Guinea, Médecins sans Frontières dealt with a major yellow fever outbreak. More than 1 million people needed to be vaccinated. At the time, Guinea was in the midst of one of the worst refugee crises in the world, with more than 100,000 refugees trapped in a war zone, with no possibility of fleeing and no access to health care.

Médecins sans Frontières is also active in countries in chronic conflict, such as Afghanistan, Angola, and Sudan, where it supports health structures with medical and logistical personnel and drugs. These activities are simple substitutes for gov-

ernment action in countries where government investment in the health system is clearly inadequate. In Angola Médecins sans Frontières documented a marked deterioration in medical and nutritional status in 2000, a clear result of government neglect. The report was based on more than 400 witness statements. Despite the country's rich resources, the government has consistently failed to invest in the well-being of its population. At the same time the international community is promoting a vision of peace just around the corner and a government that is making progress. This notion of a return to normalcy is dangerous because it may lead to inappropriate action. And it is totally wrong. War continues, and its consequences are visible in the field daily. Angola is one of the largest operations to date for Médecins sans Frontières, with more than 80 international staff supporting primary and secondary health structures in 9 of 18 provinces. Without that presence, a large part of the health system would collapse.

In more stable contexts Médecins sans Frontières can carry out innovative medical work, such as the mother-to-child HIV prevention program in South Africa.

Negative Effects of Health Care Provision by NGOs

NGO provision of health care may also have negative effects. Countries may come to depend on NGO provision, thereby slowing or halting government initiatives. Moreover, such aid may be manipulated to support the system that gave rise to the misery in the first place.

Médecins sans Frontières was the first independent humanitarian organization to enter the People's Democratic Republic of Korea in 1995. It ran into significant problems. Prevented from gaining access to the populations it wanted to assess, it was unable to identify the vulnerable or document a nutritional or health crisis. While allowed to distribute drugs to health facilities, it was unable to verify whether the population had free access to these health centers. Despite independent reports of major famine in some areas (areas to which Médecins sans Frontières was denied access), Médecins sans Frontières' feeding centers had very low numbers of malnourished children. Concluding that its assistance could not be given independently of influence by state authorities and that the most vulnerable were likely to remain so since food aid supported the very system that had created vulnerability and starvation among millions, Médecins sans Frontières left in the fall of 1998.

Humanitarian action must be based on the freedom to assess, deliver, and monitor assistance so that the most vulnerable can be assisted first. Because this was not the case in the People's Democratic Republic of Korea, where humanitarian actors were unable to serve those in need and were used in a political bargain, leaving was the lesser of bad options. While there likely was a real crisis, the government was trying to cover it up. And foreign governments were succumbing to nuclear blackmail and supporting Korea with vast quantities of aid. Aid should not be allowed to mask the causes of suffering or to be simply an internal or foreign policy tool that creates rather than counters human suffering.

During the genocide in Rwanda in 1994 roughly half a million people fled across the border into Zaire over a period of about 10 days. Initially they had no shelter, no clean water, no food, and no sanitation. Epidemics spread rapidly, causing death and disease on an unimaginable scale. Médecins sans Frontières and other organizations responded quickly to bring the epidemics under control. More than 1 million refugees settled in camps in the Democratic Republic of Congo and Tanzania. By mid-1994 humanitarian organizations had successfully controlled the epidemics and developed basic systems and supply lines for delivering food and other essential services. However, military groups soon began to reorganize, take control of refugee camps, retrain, and reequip.

As the need lessened and aid increasingly became coopted by a growing military structure that was guilty of the genocide in Rwanda in 1994, Médecins sans Frontières began to question its role and the perversion of humanitarian assistance. Médecins sans Frontières and other groups repeatedly called for the separation of those guilty of genocide from the legitimate refugees. But as Médecins sans Frontières tried to register the refugees, it was denied access by the camp authorities, and its attempts to deliver food directly to the people were forcibly prevented. Though adequate food was being delivered, there was still malnutrition. It was evident that there was diversion on a major scale by an organized and militarized authority responsible for the genocide. Again, Médecins sans Frontières had to withdraw even though there were still unmet medical needs.

These and other experiences clearly show that providing health care and other kinds of humanitarian assistance to populations in distress involves moral dilemmas. There are limits to the willingness to compromise humanitarian principles even when trying to provide access for people in need. Humanitarian organizations need freedom to independently assess the needs of the population; to gain unhindered access to them; to conduct, monitor, and evaluate the distribution of aid commodities; and to guarantee security for personnel and property. In many combat zones it may be very difficult to meet all these standards. In such situations the need for the humanitarian aid—and its effectiveness—must be weighed against the potential harm that the aid may do.

Infectious Diseases and Access to Essential Drugs

Gaining access to patients and avoiding negative effects from interventions are some of the major concerns Médecins sans Frontières faces today as a medical humanitarian organization. Overcoming the barriers to access to essential drugs is another.

An estimated 800 million people lack access to basic health care, 1.3 billion people (the extreme poor) live on less than $1 a day, and 2.6 billion lack access to safe and effective water and sanitation—the most basic indicator of access to health care. Among these people treatable or curable infectious diseases are the leading cause of death. Each year infectious diseases kill 14 million people, 90 percent of them in poor countries. One reason is that life-saving medicines are too expensive—because of patent protection, a lack of research on the diseases that afflict the poor most heavily,

or abandonment of drugs by manufacturers because the return on investment is too low. These factors are linked to the failure of national and international authorities to strongly support the right to access health care, an abdication of responsibility for the problem in favor of the multinational pharmaceutical industry, and the weakness of the mandates of international organizations such as the World Health Organization relative to the strength of organizations such as the World Trade Organization.

Malaria causes 300 to 500 million cases of illness and 1–2 million deaths every year, mostly among poor people in developing countries. Resistance to standard therapy is increasing rapidly, reaching 80–90 percent in some countries. New treatments are unavailable or unaffordable.

AIDS is another major health problem. Since the beginning of the pandemic in the 1980s, more than 20 million people have died of the disease, 40 million people now live with HIV, and 5.4 million people are infected every year. By 2020 half a billion people will be infected with HIV—some predictions are even higher. Most of the people with HIV and most of those who are likely to get HIV are in the developing countries. Treatment with patented antiretroviral drugs, which prolong life but do not cure AIDS, costs $10,000–15,000 a year. The cold fact is that only some 5 percent of people who are HIV positive have access to life-prolonging patented antiretroviral drugs. The 95 percent who do not are among the 2 billion poor people in the world who live on less than $2 a day.

These are the people that Médecins sans Frontières serves—the poor who have a need but no purchasing power and are therefore not a market for patented antiretroviral drugs. Entire African nations are on the verge of collapse, as doctors, teachers, military personnel, and civil servants succumb to AIDS. These nations are dying not of AIDS alone, but of market failure as well. Access to life-prolonging treatment is denied because of patent protection, the absence of public health infrastructure, and a lack of good-quality generic drugs. The availability of drugs is not the only issue—but it is the essential issue. Infrastructure and effective treatment delivery will never expand if there is not even the possibility of affordable drugs.

Sleeping sickness (trypanosomiasis) is yet another example. Production of eflornithin, one of the drugs developed to treat this deadly disease, was stopped because the patients who needed it could not pay for it. Again, the world's poor are not considered a market. They are people who have need, but not enough money. It is that simple.

Thus intellectual property rights and patent systems and uniform pricing worldwide exclude the poorest from access to health care. The role of NGOs is clear: to push for change and political responsibility in providing health care delivery. Based on its field experience, Médecins sans Frontières launched an international campaign in 1999, challenging politicians, pharmaceutical companies, the World Health Organization, and others to address this growing problem. There are solutions. Trade regulations offer possibilities for poor countries to bypass patent rights and produce their own drugs or import cheaper generic drugs. These options should be encouraged. In March 2001, 39 pharmaceutical companies and their trade organizations brought suit against the government of the Republic of South Africa to stop

plans to promote the use of generic drugs and cheaper imports to treat patients with HIV/AIDS. This was the only way the government could get affordable drugs for the millions of people who are infected and who will face an early death. Can we accept the notion that the interests of some pharmaceutical companies should prevail over the lives of millions? After major public pressure the pharmaceutical companies withdrew their suit, an important victory for poor people with HIV/AIDS.

A solution has also been found for the drug against sleeping sickness—but only incidentally, because a new market for the drug emerged. A drug that can save the lives of hundreds of thousands of people in Africa will be produced again because it turned out to be an effective drug for removing unwanted facial hair in women.

Lack of research and development (R&D) is another problem affecting the developing world. Despite the enormous private investment in drug research over the last quarter century, only 11 of the 1,223 new chemical entities approved during that period were for tropical diseases and most were the result of veterinary or military research. Similarly, of the 95,417 therapy-relevant scientific publications in 1995, only 182 concerned tropical diseases. Of these, 79 were on malaria, 34 on tuberculosis, and only 3 on African sleeping sickness.

The pharmaceutical market has been expanding rapidly in North America and Europe. In North America, with 5 percent of the world's population, the drug market is projected to rise to more than $160 billion in 2002, double the $80 billion of 1993. The market in Africa and Asia, by contrast, has remained the same while the population has doubled. Today, Africa and Asia make up 72 percent of the world's population, but Africa constitutes only 1 percent of the projected world pharmaceutical market for 2002. Thus the size of the R&D budget seems to be closely linked to the size of the market.

While pharmaceutical companies spend billions on R&D on diseases of concern to industrialized countries, the product development budget of the Tropical Disease Research Program (a joint program of the World Health Organization, the World Bank, and the United Nations Development Programme) has averaged just $10 million a year during the last decade. The numbers speak for themselves. R&D activities are responding less and less to the real clinical needs of the developing world. This is a crisis that must not continue.

Who is responsible for solutions?

Roy Vagelos, former head of Merck, has commented that we cannot ask industry to solve what is essentially a social problem. While the pharmaceutical industry must contribute to the search for solutions, Médecins sans Frontières believes that industry alone cannot solve the crisis or be allowed to set the rules. The lack of R&D is partly a result of market failure, but it is also a public health failure. Political leadership is crucial for ensuring that R&D does not serve only the needs of the wealthy.

Two years ago Médecins sans Frontières set up the Drugs for Neglected Diseases Working Group, an independent international team of biomedical scientists, tropical medicine experts, health economists, legal and regulatory specialists, and representatives from health NGOs, the World Health Organization, and industry. Its goal is to identify strategies to promote the development of new, effective, safe, afford-

able, and easy to use drugs. The group recommended defining a clear, need-driven research agenda for new medicines and vaccines, to assist policymakers, funding agencies, and the research community in setting the right priorities to address the needs of developing countries. This agenda will drive a coordinated effort to develop 10–20 new drugs over the next 10 years, with an estimated cost of $500 million to $2 billion. This amount is not beyond society's reach.

The group also recommended creating mechanisms to drive needed research in the private sector—for example, requiring that a small percentage of profits go toward developing essential medicines for neglected diseases. A fully subsidized system may be needed for diseases that are prevalent only in developing countries, and an equity pricing system should be used when the disease affects both rich and poor countries. For neglected diseases the group also suggested negotiating an international treaty to promote R&D on medicines and vaccines that are safe, effective, affordable, and easy to use. It should correct the imbalance between rights and obligations under current international treaties and agreements, such as Trade-Related Aspects of International Property Rights (TRIPs). It should guarantee that drugs for neglected diseases will be considered global public goods and address the relevant intellectual property issues.

Who Is Responsible for Health Care Delivery?

The solutions proposed by many U.N. agencies, governments, and multinational companies include donations or price reductions of patented drugs, public-private partnerships to support these initiatives, and corporate community programs to support public infrastructure and training programs. Are such programs the responsibility of corporations or of governments? Are drug donations or price reductions a sustainable solution for access for all? Are public-private partnerships viable solutions to the long-term responsibility of the state to ensure the right to access health care? Is it acceptable that some foundations should set and drive the international health agenda by virtue of the sheer size and power of their financial resources?

Where is the state in meeting these responsibilities? In the case of the HIV/AIDS pandemic these kinds of initiatives may allow pharmaceutical firms to sidestep the threat that compulsory licensing and generic drug competition represents to their profits. More important, such initiatives perpetuate the notion that private charity—an act of privilege—is a sound alternative to public responsibility for ensuring the right of access to health care. It allows politicians to respond with political platitudes and to use piecemeal private actions as a humanitarian alibi for the failure to achieve real access to health care for all.

NGOs, Médecins sans Frontières included, have been complicit in this humanitarian alibi. In many ways NGOs have become comanagers of misery with the state, providing a salve instead of a cure, allowing charity to mask duty, and failing to demand real political change in place of political platitudes. NGOs have failed to insist on political responsibility not just for the rich and the included, but for everyone—rich, poor, dispossessed, and excluded.

Now that the diseases and suffering of the poor are a threat to national security and the expansion of global markets, their plight has provoked political interest. This new interest, not economic or state security interests, should drive the agenda for access to health care for all, the included and the excluded. The economist Amartya Sen has argued that poverty is not just about economics, but also about a fundamental lack of freedoms. How NGOs choose to use their own liberty—how they define their vision, what actions they take, and how they use their voice—also matters. They have been too passive and too deferential to political platitudes and to partial and imperfect private initiatives. They must demand more and fight for the freedom of their beneficiaries, whose fundamental liberties are constrained. This is nothing but an outrage.

There are many who claim to speak for the poor—the World Bank, the International Monetary Fund, many U.N. agencies, and others. They are, in effect, coopting the voice of the poor. Médecins sans Frontières speaks only for itself. It does not pretend to speak "for" anyone—for victims of war, the marginalized, the excluded, the poor, or anyone else. Médecins sans Frontières speaks as itself, with its own voice, of its own direct experience of solidarity in its projects, of its own outrage, and of its own demands. And it is able to do this because it is operationally, politically, and financially independent. To repeat: Médecins sans Frontières is not part of the liberal economic agenda that seeks to substitute private actions for government responsibilities. Médecins sans Frontières tries in various ways, drawing on its field experience, to stimulate government responsibility and international responsibility. Our goal is not to exist.

References

OCHA (U.N. Office for Coordination of Humanitarian Affairs). 2001. "Consolidated Inter-Agency Appeal 2001 for Sierra Leone." Geneva.

United Nations General Assembly. 2001. "Declaration of Commitment on HIV/AIDS: Global Crisis, Global Action." United Nations Special Session on HIV/AIDS, 25–27 June 2001, New York. http://www.unaids.org/whatsnew/others/un_special/Declaration2706_en.htm

Comment on "The Role of Nongovernmental Organizations in Providing Health Care," by Morten Rostrup

Beatriz Zurita

It is informative to consider the empirical work of Médecins sans Frontières—the quintessential international nongovernmental organization (NGO) providing health care—when trying to explain policy considerations for service-oriented NGOs for alleviating inequality and poverty in developing countries.

NGO activities derive from the mandates of the organizations, respond to the interests of those who fund them, and comply with the regulatory framework in their country of origin. Morten Rostrup's article clearly presents the mandate of Médecins sans Frontières as a medical humanitarian organization that seeks to relieve suffering, restore autonomy, expose injustice, and insist on political responsibility and government accountability for human rights and dignity. Médecins sans Frontières has received international awards for its work, commending those who work for and support the organization.

In general, NGOs whose mandate is similar to that of Médecins sans Frontières provide social services as a means of achieving universal goals clearly supported by their constituents, such as saving and protecting lives—goals that are also international public goods. In the field, these NGOs have a variety of means for achieving these goals. Sometimes their activities have particular legitimacy and additional value, especially when carried out with local support and when they empower groups in greatest need. In other cases, NGOs carry out their work without buy-in by national and local groups. At most, these foreign NGOs may introduce new ideas into a society, but their activities are not likely to achieve sustainability even if they have positive short-term effects. Another problem is the poor regulatory framework in many developing countries, which allows for questionable and even illegitimate processes to remove barriers to important goals. In such countries, it is difficult to control or eliminate NGOs whose methods or activities are questionable or even illegal. Médecins sans Frontières could assist countries to develop a simple regulatory framework to help them stop the activities of such NGOs.

In its work, Médecins sans Frontières interacts with populations and governments that need its help, jointly developing a work plan. But the legal framework in which these NGOs work is not clear, especially since countries with the greatest need for assistance typically have the weakest laws and regulations to protect human rights—and even weaker enforcement.

It would be useful if Médecins sans Frontières were to help develop a minimum regulatory framework for the work of international NGOs. Médecins sans Frontières could explain how it complies with the regulatory framework and makes itself accountable to local groups, the governments of the countries in which it operates, its own government, and its supporters. Basic compliance with local and national regulations is necessary to avoid institutional pressures from well-established national political groups that would take advantage of NGOs to advance their political goals. In those cases Médecins sans Frontières has often elected to leave the country. If better regulations were available, good NGOs would not have to rely on questionable groups for support.

It is difficult to coordinate government and NGO activities, especially over the long term. When there is no local buy-in, institutional pressures increase. This typically occurs when the means used to reach goals are not compatible with the local or national culture. In this case, NGO activity can be counterproductive, making it more, not less, difficult to reach intended goals. It may even lead to the loss of lives.

An additional weak point for the work of NGOs is their accountability to those who support them financially or in kind. National governments, even in developed countries, tend to have poor regulatory frameworks for NGOs. That makes it very difficult for these organizations to demonstrate the transparency of their operations and their efficiency in achieving their mission. This becomes especially difficult when NGOs are small, engage in fragmented activities, and struggle to get appropriate support and legitimacy by local civil society. Long-term sustainability for NGOs is contingent on being embraced by local societies as one of their own.

A Patent Policy Proposal for Global Diseases

Jean O. Lanjouw

We are in the midst of a dramatic extension in the global reach of the patent system. Public concern over the price of HIV/AIDS drugs in Africa has focused attention on this new global system and generated a debate between those who support the establishment of strong patent laws to protect pharmaceuticals in developing countries and those who, in various ways, would weaken them. But the choice does not have to be limited to strong or weak. This article describes a policy that would improve on the current patent regime by acknowledging the differences in pharmaceutical markets in rich and poor countries and what they imply for optimal patent protection. It would lower the price of pharmaceuticals in developing countries for a selected group of global diseases—those that, like cancer, affect people everywhere—while allowing protection to increase where it is most likely to lead to the creation of new products for diseases that affect primarily people in developing countries. The proposal requires no changes in international treaties, no changes in the adoption in poor countries of patent systems that comply with membership requirements of the World Trade Organization, and only minor changes in the patent laws of rich countries. Because the policy would be largely self-enforcing and use existing institutions, it would cost very little to introduce or administer and thus would not divert spending from other important health and development projects.

We are in the midst of a global expansion in the extent to which pharmaceutical innovations are protected by the patent system. Previously, most developing countries treated such innovations as nonpatentable or at best offered only minimal protection for new manufacturing processes. Today, as a result of bilateral pressure and membership requirements of the World Trade Organization, developing countries are implementing new patent laws that look very similar to those in Europe and the United States, granting full protection to all inventions in this area.

The public attention now focused on patents and on the price of HIV/AIDS drugs in Africa has created an opening and a demand for creative thinking about ways to improve this new global system. Given the tradeoff between prices and innovation that is inherent in supporting research and development (R&D) through a patent system, can the system be structured to elicit the same amount of innovation at a lower welfare cost? In answering this basic question, it would be a mistake for international and domestic policy discussions to focus only on AIDS, despite its undoubted importance. The U.S. Patent and Trademark Office in 1998 granted more than 10,000 patents relating to pharmaceutical innovations, spanning thousands of diseases. No policy designed to address the current AIDS crisis is likely to be the best policy for the system as a whole, nor is it necessary to think in those terms. The AIDS epidemic is an international emergency of the first order. It can be treated as exceptional and deserves its own policies.

An analysis of the implications of extending protection to additional countries is closely analogous to that of the implications of granting protection for more years (see Nordhaus 1969; Deardorff 1992). Fundamental determinants of the optimal extent of protection are the degree to which the prospect of greater profits leads firms to increase investment in research and the degree to which greater investment results in innovation that benefits the public. These responses tend to decline at higher levels of R&D investment. Thus increasing protection can be expected to yield more benefit where incentives are initially low.

From this perspective, it is important to recognize that there are two very different and identifiable types of drug markets. Some diseases are important worldwide, being found in both poor and rich countries, and therapies for these diseases have global markets. Other diseases, such as malaria, are more specific, with almost the entire market for their therapies in the developing world (table 1).

There has been almost no investment outside the public sector in therapies for diseases specific to the developing world. Without patent protection in the developing world, there has been little prospect of profit anywhere and therefore little interest by firms in investing in therapies for these diseases (see Lanjouw and Cockburn 2001). The new patent regime may draw resources into creating drugs to prevent and treat diseases specific to poor countries.

Of course, even with effective patent systems, developing country markets may not be very attractive, given the prices they can support. Recent initiatives to "make a market" aim to put more money into these markets through a dedicated fund or tax credit to subsidize purchases of specified products (see Kremer 2001a, 2001b;

Table 1. Impact of Diseases Specific to the Developing World in Low- and Middle-Income Countries, 1998
(thousands)

Disease	Disability-adjusted life years lost	Deaths
Diarrheal diseases	72,742	2,212
Malaria	39,267	1,110
Measles	30,067	882
Pertussis	13,047	342
Tetanus	12,950	409
Syphilis	4,957	159
Lymphatic filariasis	4,698	0
Leishmaniasis	1,707	42
Schistosomiasis	1,696	7
Trichuriasis	1,287	5
Trachoma	1,255	0
Trypanosomiasis	1,219	40
Onchocerciasis (river blindness)	1,069	0
Chagas disease	588	17
Dengue	558	15
Japanese encephalitis	502	3
Leprosy	393	2
Polio	213	2
Diphtheria	181	5
Ancylostomiasis and necatoriasis	—	—

— Not available.
Note: The table shows diseases for which 99 percent or more of the global burden fell on low- and middle-income countries in 1990.
Source: For global burden, WHO (1996); for disability-adjusted life years lost and deaths, WHO (1999).

World Bank 1999). This type of policy is appropriate for stimulating private investment in research on disease therapies that have small markets in the West but are of great importance in the developing world.

But consider global diseases—those that are widespread not only in poor countries but also in rich countries. These diseases are the focus of the proposal described here. They have received less attention in development debates over intellectual property because they are not specific to developing countries. But that does not mean that they are not important causes of disability and mortality among the poor. Cancer, heart disease, and diabetes together account for 16 percent of the disability-adjusted life years (DALYs) lost in low- and middle-income countries (table 2). (Similar percentages were found using mortality.) This is four times the share of the DALYs accounted for by malaria in these countries.

Not only are "rich country" diseases important in poor countries, they appear to cut across the income spectrum. Data from a Pakistan health survey on the prevalence of strong risk factors for cancer and cardiovascular disease show that smoking is both widespread and significantly higher among poor males in Pakistan than among better-off males (table 3). Moreover, while those in the bottom half of the asset distribution have lower rates for the risk factors associated with cardiovascular disease, the rates are still high—with about a quarter suffering from hypertension and 15 percent having high cholesterol.

Table 2. Disability-Adjusted Life Years Lost as a Result of Selected Diseases, 1998
(percent)

Disease	Share of DALYs lost in low- and middle-income countries	High-income countries' expenditure-weighted share of global DALYs lost[a]
Cardiovascular disease	10	91
Cancer	5	94
Diabetes mellitus	1	96
Malaria	4	0

Note: Low- and middle-income countries have a weighted average annual GDP per capita of $1,250, and high-income countries, $25,510.
a. The weighted percentages are based on 1990 per capita drug expenditure in India and the United States (representing the poor and rich countries) times disability-adjusted life years (DALYs) lost in 1998.
Source: For disease statistics, WHO (1999); for expenditures, OPPI (1996).

Table 3. Prevalence of Risk Factors for Chronic Diseases by Level of Wealth in Pakistan, 1990–94
(percent)

Disease and risk factor	Rural		Urban	
	Low wealth	High wealth	Low wealth	High wealth
Cancer				
Smoking				
Males	35.5 (2.3)	33.7 (5.0)	57.0 (5.0)	33.0 (3.3)
Females	4.0 (0.7)	2.3 (1.2)	9.1 (2.1)	2.4 (1.0)
Cardiovascular disease				
Hypertension	22.0 (1.8)	52.1 (4.7)	29.7 (4.2)	46.0 (3.8)
High cholesterol	13.7 (1.8)	33.7 (5.7)	22.1 (3.7)	27.8 (4.0)
Percentage of population	42.0	6.0	8.0	9.0

Note: Wealth groups are defined by the number of assets owned, with fewer than three defined as low wealth and more than five as high. Assets include such items as fans, irons, radios, tape recorders, and televisions. A total of 18,315 people were surveyed and examined. See Pappas and others (2001) for more details. Figures in parentheses are estimated standard errors.
Source: Pappas and others 2001.

These data from Pakistan are unusual in having information from direct health examinations of the sampled individuals, rather than simply statements about disease incidence, together with at least some measure of household wealth. Other data give self-reported, and therefore less reliable, disease incidence, but include better measures of household wealth. Surveys in India, for example, found that of about 12,000 deaths in rural areas (over age 14), 11 percent of those occurring in the lowest 20 percent of the all-India wealth distribution were ascribed to cancer or heart disease. This rate is well below the 35 percent in the highest quintile, but still substantial (Deon Filmer, World Bank, personal communication, March 2001).

The evidence is not plentiful, but what evidence there is suggests that "rich country" diseases are widespread in poor countries and that they are important among the poor in those countries, not just the relatively rich. Still, almost all the potential market for therapies for global diseases is found in the West. Rough measures of the relative market size in rich and poor countries, based on disease incidence as measured by DALYs and weighted by estimated drug expenditures, show that almost all

the market for therapies for cancer, heart disease, and diabetes is in rich countries (see table 2). This is in stark contrast to the market for malaria therapies.[1]

Drug expenditure patterns suggest that poor countries contribute little to total world expenditure on drugs for global diseases (table 4, top panel). Yet they can still be a major source of demand in some therapy areas, such as parasitology. And a significant share of the total spending on drugs by poor countries goes to those for global diseases, though their spending is of little importance in world demand for such drugs (table 4, bottom panel). Among the largest developing country drug markets, six that together represent about 46 percent of the world's population account for less than 2 percent of total expenditure on drugs for cardiovascular disease (table 5).

Thus global diseases are worthy of attention for the following reason:

For therapies for global diseases, the profit derived from having a monopoly over sales in poor countries makes only a marginal contribution to the total worldwide profit of pharmaceutical firms and therefore only marginally increases their incentive to invest in research. At the same time, in a poor country even a small price increase due to such a monopoly can greatly reduce the number of people able to purchase patented drugs and the welfare of those who do. This is particularly so since drug purchases in developing countries are largely paid for directly by consumers, without the benefit of insurance.

In this article I propose a policy that could improve on the current regime by acknowledging these differences in markets and what they imply for optimal patent protection. The policy would allow protection to continue increasing worldwide in most areas of pharmaceutical innovation, as envisioned in the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs) section of the General

Table 4. Drug Expenditure Patterns in Developed and Developing Countries, 2000
(percent)

Group or country	Spending as share of world total in therapy area			
	Cardiovascular	Anti-infectives	Parasitology	All therapy areas
Developed countries	95.7	92.3	65.4	93.6
Developing countries	4.3	7.7	34.6	6.4
Mexico	1.0	4.1	13.5	2.4
	Spending in therapy area as share of group or country total			
Developed countries	19.6	10.0	0.1	100.0
Developing countries	12.8	12.2	1.0	100.0
Mexico	8.0	17.5	0.9	100.0

Note: Based on expenditure in selected countries for the 12 months to October 2000. The developed countries included are France, Germany, Italy, Japan, the United Kingdom, and the United States. The developing countries are Argentina, Brazil, and Mexico. The choice of countries has no significance beyond the availability of detailed spending data.
Source: IMS HEALTH Global Services data [http://www.ims-global.com].

Table 5. Income, Population, and Drug Expenditure by Country, Late 1990s

Country	GDP per capita, 1998 (PPP U.S. dollars)[a]	Population, 1998 (millions)	Population as percentage of world total, 1998	Drug expenditure as percentage of world total, 1999	Estimated expenditure on cardiovascular drugs as percentage of world total, 1999[b]
Pakistan	1,715	131.6	2.2	0.30	0.12
India	2,077	979.7	16.7	1.13	0.47
Indonesia	2,651	203.7	3.5	0.27	0.11
Egypt, Arab Rep.	3,041	61.4	1.0	0.30	0.13
China	3,105	123.9	21.1	2.07	0.86
Philippines	3,555	75.1	1.3	0.39	0.16
Subtotals			45.8	4.0	1.85
Venezuela, República Bolivariana de	5,808	23.2	0.4	0.43	0.18
Colombia	6,006	40.8	0.7	0.43	0.18
Brazil	6,625	165.9	2.8	1.72	0.72
Mexico	7,704	95.8	1.6	1.59	0.66
South Africa	8,488	41.4	0.7	0.31	0.13
Saudi Arabia	10,158	20.7	0.4	0.38	0.16
Argentina	12,013	36.1	0.6	1.14	0.47

a. GDP per capita is converted to U.S. dollars using a constant purchasing power parity (PPP) index.
b. The estimated share of cardiovascular drug expenditure represented by a country is its percentage of total drug expenditure multiplied by the ratio of cardiovascular drug expenditure to total drug expenditure for Mexico: 1.0/2.4 = 0.41 (see table 4, top panel).
Source: For GDP and population, World Bank (2000); for expenditure, IMS HEALTH Global Services data [http://www.ims-global.com] and Anne Calbazana, IMS HEALTH Global Services, personal communication, April 2001.

Agreement on Tariffs and Trade (GATT), commonly known as the TRIPs agreement. In particular, and in contrast to other proposals being discussed, such as indiscriminate compulsory licensing, the policy would allow protection to strengthen for therapies for diseases specific to developing countries, where there is a clear argument to be made that new incentives are warranted. At the same time, it would effectively keep protection at its current level where an increase in profits is less likely to generate new innovation. To do this, the policy would require inventors to choose to avail themselves of protection either in the rich countries or in the poor countries, but not in both, whenever a patented product is for a global disease. Because the profit potential offered by rich country markets is far greater, firms would naturally relinquish rights in poor countries. Thus the policy would lower the price of drugs for global diseases and should be seen as a complement to policies that target diseases specific to poor countries.

The policy gives a feasible way to present patentees with the desired choice between protection in rich or poor country markets in the limited situations where their patents relate to products for specific global diseases. Economists and policy-makers have been reluctant to differentiate protection across types of innovation even though there is a strong theoretical basis for doing so (and Article 27 of the TRIPs agreement explicitly forbids discrimination by area of technology or by loca-

tion of invention or manufacture, except in specific and limited circumstances). There are good reasons for this. The information needed to decide how best to differentiate is in limited supply, and any differentiation must be on features both easily identified and hard to change or resources will be wasted as everyone tries to fit into the better class. The mechanism I propose has useful revelation and self-enforcement features that resolve these problems.

The Policy

The mechanism is remarkably simple to describe and to implement. But why it works is not directly obvious and requires some explanation. So before turning to the idea itself, I outline its attractive features.

Features

1. The policy does not contravene existing treaties (Paris Convention, Article 4bis; TRIPs agreement, Article 27).
2. The policy can be implemented unilaterally, although it would be most effective and acceptable to all parties if Japan, the European Union, and the United States moved together. (For simplicity, I comment below as though only the United States implemented the policy. The comments would be equally true for other rich countries, and one could read "France" or "Japan" in place of "the United States" if those countries participated.)
3. The policy would require no changes in developing countries' new patent systems or in the development of their enforcement procedures. In fact, better-functioning patent office and court systems in developing countries would only improve the working of the policy. At a time when there is concern with nurturing budding compliance with the TRIPs agreement, the fact that this mechanism would not "muddy the waters" seems a great advantage.
4. The mechanism would rely almost entirely on the quality and reliability of U.S. institutions, not on those in developing countries.
5. The policy would be fully controlled by the U.S. government. This is in contrast to the compulsory licensing sanctioned by developing country governments, which will find it difficult to resist pressure by local interests to expand coverage to all diseases.
6. The mechanism would not require information that is clearly not available. In particular, and crucially, it would not require examining patents to identify them as covering innovations for a particular disease. Such a task would be infeasible. The expense aside, the patent owner may not know the future uses of a patented innovation. The policy mechanism would induce firms to volunteer the link between patents and products when it becomes known and only as necessary.
7. No one would be told what to do. Incentives would be aligned to make use of the greater information that firms have about the relative size of global

markets for different products. Firms would behave as desired without outside control or monitoring.

8. Because the policy would use existing institutions and procedures, would be largely self-monitoring, and would not require the collection of information for each patent, it would cost very little to administer and enforce. As a result, the policy need not be seen as an alternative to other policies within the constraints of fixed health or development budgets.

The Mechanism

I first describe how the policy works in the simplest possible terms, leaving details to the discussion that follows. Assume, initially, that there are only

- two countries, the United States (representing a set of rich countries) and India (representing a set of poor ones);
- two diseases, malaria and cancer, the first representing a set whose therapies have no U.S. market, and the second, a set whose therapies have a very large U.S. market and a substantial but much smaller Indian market; and
- three companies—PharmaUS, CiplaIndia, and USGeneric—each representing a type of firm in the pharmaceutical market.

Bear in mind that patents are national in coverage. Obtaining protection in France requires an application for a French patent. Obtaining protection in Brazil requires an application for a Brazilian patent. Now when an innovation is made in the United States, the inventor is required to apply first for a U.S. patent. To make subsequent, foreign applications, the inventor must first obtain a foreign filing license from the U.S. Patent and Trademark Office. This rule is meant to protect military secrets, and variants of it are found in patent regulations elsewhere.[2]

The proposed policy is, very simply, to stipulate that when a patentee petitions for this license, he does so in language something like this:

I, the undersigned, request a license to make foreign filings for patent no. X, with the understanding that this permission will not be used to restrict the sale or manufacture of drugs for cancer in India by suing for patent infringement in India.

Requiring this declaration to obtain the license to file abroad is the entire policy. An existing provision in the patent law would be used to serve an unanticipated purpose. The mechanism would work because other features of the patent law and pharmaceutical regulation can also be turned to serve this new purpose. These are discussed below.

Why It Would Work—A Basic Outline

Consider the simplest situation. There is a cancer product based on a single innovation. This innovation is protected by a U.S. and an Indian patent, both owned by

PharmaUS. PharmaUS obtains marketing approval in both countries and sells the product. Now CiplaIndia (or USGeneric) enters the Indian market with its own version of the product. PharmaUS could do one of two things. First, it could do nothing. Under this choice PharmaUS would no longer obtain monopoly profit from its Indian patent, but its profits from the vastly larger U.S. market would not be affected. The competition introduced by the entry of CiplaIndia lowers prices in India, achieving the goal of the policy.

Second, PharmaUS could sue CiplaIndia for infringement of its patent in India, and would win. Nothing prevents the company from choosing to protect its monopoly profits in India on the basis of its patent there, in an Indian court, just as it would without the policy. But what happens then? At this point CiplaIndia—or, more likely, USGeneric—can infringe and defend itself with the claim that by attempting to stop CiplaIndia's sales of the cancer product in India, PharmaUS has rendered its U.S. patent unenforceable. By filing suit in India, PharmaUS has falsified the declaration it made to the U.S. Patent and Trademark Office to obtain the foreign filing license. Patentees have a duty to deal with this agency in good faith, and failure to do so is clear grounds for rendering a patent unenforceable.[3]

Now suppose that the innovation is for a malaria product. In this case the firm also has two choices when confronted by entry: it can do nothing, or it can sue CiplaIndia for infringement. Now, however, the suit gives no grounds for rendering the U.S. patent unenforceable, because the declaration made by PharmaUS to obtain the foreign filing license says nothing about malaria.

So what is the result? In the case of a patent for a cancer product, PharmaUS's two choices are effectively between protecting its profits in the United States or protecting them in India, but not both—just as desired. Given this choice, PharmaUS will not sue in India for infringements of cancer product patents because it will not want to jeopardize its U.S. patents. Knowing this, CiplaIndia will enter the market, and prices in India will fall. In the case of a patent for a malaria product, PharmaUS's two choices are effectively between protection in the United States (if it does nothing in the face of infringement) or protection in both the United States and India (if it does take action). Given this choice, it will sue in India for infringements of malaria product patents. Knowing this, CiplaIndia will avoid the suit by not entering the market, and the incentive for investment in malaria products will be maintained.

One might say, "With this policy, PharmaUS may not even bother to get a patent in India for its cancer product." This is true, and it is fine. PharmaUS will follow one of two strategies: it will continue to market its patented cancer product in India on a competitive basis, or it will leave the market to CiplaIndia and USGeneric.

Multinationals have followed both strategies over past decades in countries that have not granted them patent protection. Both developing country firms and developed country generics manufacturers have shown themselves to be adept at rapid imitation and entry. This was, after all, the point of pressing for the TRIPs agreement in the first place, as well as for domestic legislation to control entry by generics. Lanjouw (1998) presents evidence indicating that over the past two

decades major patented drugs typically arrived on the Indian market within seven years of their world launch and often much sooner. Watal (2000) suggests that arrival speed has increased. For 10 drugs launched in the United States after 1985, she finds an average time lag to availability in India of just two years. Thus there appears to be no reason to be concerned about which strategy the patentee chooses.

The mechanism is designed to be triggered by a lawsuit. Why go this route? Because when infringement suits are filed to prevent the sale of a *product*, it is on the basis of a set of *patents*. The desire to prosecute the suit successfully gives the patent-owning firm an incentive to correctly identify which patents it believes best protect the product. This incentive resolves the otherwise intractable problem of how to determine the use of particular patents. It allows the mechanism to work without requiring that a bevy of scientists try to identify patents that might someday be for cancer.

Parallel Imports and Low-Cost Sources of Supply

Firms have a legitimate concern about low-cost sources of supply and seepage across borders, particularly into their major markets. On the face of it, this proposal does not seem helpful in this regard, since its intention is precisely to encourage low prices for some products in poor countries. Firms may well object to it on these grounds. But we must have low-cost sources if we are to have any hope of ensuring adequate availability of drugs to poor people. The rich world shows no sign of willingness to supply aid at levels that would make purchases at U.S. prices feasible. Thus the only appropriate response to this concern of firms is to address the possibility of seepage. If firms confront substantial international arbitrage, they will naturally respond by selling at a uniform price—one likely to be far higher than even the monopoly prices appropriate to poor countries. And they may decide to forgo launching drugs in the poorest countries altogether. To prevent this, efforts should be directed toward helping firms to separate markets—regardless of whether the policy proposed here is implemented.

A first step in easing firms' concern might be legislative confirmation that the United States has no international exhaustion of rights doctrine, in keeping with the more recent federal Circuit Court interpretation of the law on exhaustion (see Adelman and others 1998). This would be a clear statement that holders of U.S. patents have the right to prevent products from coming into the United States from elsewhere, even if the products were originally sold by their own licensees or subsidiaries.

The bigger issue, however, is the enforcement of rights in this area. Drugs are small and lightweight, which makes it difficult to prevent products that have been sold cheaply in a country where consumers are poor from flowing back into markets where consumers are better off. The Internet may greatly exacerbate this problem by enabling consumers to purchase drugs directly from around the world. Once developing country firms have developed sufficient reputations for quality, one can

easily imagine hundreds of thousands of packets crossing borders by mail. Patentees will be hard pressed to identify such individual infringements and reluctant to enforce a separation of markets by suing their customers.[4] Internet sales also pose a safety threat to consumers. How is one to know that a Web-based pharmacy is actually in North Carolina and not a counterfeit operation operating from overseas? (See Hubbard [2000] for a discussion of efforts by the U.S. Food and Drug Administration [FDA] to combat this problem.)[5]

Resolving the enforcement problems will require better coordination and regulation of drugs at the source. Thus the participation of poor countries in efforts to prevent illegal movement of drugs across borders will be key. The proposal described here is designed to benefit developing countries and in a way that would be apparent to their populations. (This is in contrast to the TRIPs agreement, which, whatever its long-run benefits in the form of new products, has engendered much resentment in developing countries.) It would seem reasonable to expect that developing countries, in turn, would make efforts to ensure that drugs priced for their consumers actually get to their populations and do not escape as exports to rich countries.

This might be done in various ways. One possibility can be seen by analogy. The U.S. government taxes gasoline and diesel fuel at different rates, depending on the intended use. This differential taxation is difficult to enforce once distribution to users has occurred, since the taxed and untaxed fuels look the same. The solution has been to dye the untaxed fuel (see U.S. Department of the Treasury 2001). Health authorities in all countries already specify features of appearance and packaging for drugs. Poor countries that are candidates for inclusion under the policy could be asked to require that pharmaceuticals sold domestically be, say, lime green. This would make it simpler to check bulk movements, and alert consumers elsewhere that their drugs were not actually manufactured in North Carolina, as they had supposed.

There may be better ideas on how to use form and packaging to differentiate products; firms have much expertise in this area, and their advice would be valuable. But the point is clear. The fact that the policy encourages low prices in developing countries certainly implies the continued existence of low-cost sources of supply. But the policy also gives poor countries a positive reason to cooperate in resolving this looming, and extremely difficult, international enforcement problem. Seen from this perspective, the policy could help firms protect their more valuable markets.

Linkages

One of the stated advantages of the mechanism is its reliance on U.S. institutions. Yet it is triggered by a court case in India. This may seem surprising. But it is the filing of a suit that is the trigger—the policy's effectiveness does not rely on the subsequent legal proceedings in India. Using the Indian case for this purpose does raise two issues, however.

Linking Products to Diseases

First, there must be a clear procedure for determining, on the basis of U.S. institutions, whether the Indian product that is the subject of the suit corresponds to a particular disease. CiplaIndia or USGeneric will always have an incentive to claim that a disputed product is for cancer in order to render unenforceable the U.S. patent of PharmaUS, while PharmaUS will claim that all products are for malaria.

I suggest the following: All products marketed in the United States are approved by the FDA for specific indications. To render PharmaUS's patent unenforceable, USGeneric must apply to the FDA for an abbreviated new drug approval (ANDA) for the Indian product. This application would claim that the Indian product is equivalent to one already marketed in the United States with a cancer indication. This procedure is already followed for any generic on the expiration of a patent, so generics manufacturers are well versed in it. If the FDA issues tentative approval or a preliminary letter reporting bioequivalence, the case that the Indian product is for cancer is made.

At this point USGeneric or CiplaIndia will request final marketing approval from the FDA, since obtaining access to the U.S. market was the point of rendering PharmaUS's patent unenforceable. The bioequivalence report is the basis for that approval. Thus using the FDA's ANDA process for this purpose results in no net increase in resources expended by the companies or the government. And because the bioequivalence report has direct implications for the integrity of the U.S. system of safety regulation, the FDA has a serious interest in its quality.

Linking Patents to Patents

Second, the Indian patents supporting the suit need to be linked to their U.S. equivalents. Fortunately, this link is a standard output of international patent procedures. Following a first filing in the United States, a subsequent Indian application will typically refer back to the U.S. application to establish the owner's global priority over the innovation and the time limit for related foreign filings. The global links between patents covering the same innovation, to the extent exposed by this process, can be found in publicly available databases.

More Complex Settings

The simple situation described above, in which a single patent protects a single product, is rare. How would the mechanism work in more complex settings?

Single Patent, Multiple Uses

Suppose first that an innovation made by PharmaUS, and patented in both the United States and India, leads to a product found to be useful against two diseases: cancer and malaria. Suppose too that PharmaUS obtains marketing approval in the

United States for cancer and malaria indications but requests marketing approval in India only for the malaria indication. Now suppose that CiplaIndia or USGeneric enters the Indian market.

If PharmaUS files an infringement suit, the U.S. patent would be vulnerable because the Indian product is bioequivalent to a U.S. product approved for cancer. The disease indications claimed in the Indian marketing approval process are of no consequence. Given this, PharmaUS will refrain from enforcing its Indian patent regardless of how the product is used in India. Together with some profit from sales of the product in the United States for its malaria use, the valuable U.S. cancer market will be the source of support for R&D investment in dual-use products.

Of course, PharmaUS could protect markets in both countries by requesting marketing approval of the product in the United States only for the malaria indication. But this would prevent the firm from legally advertising the product's use for cancer to doctors and the public—and would therefore be an unattractive option when the cancer market is expected to be significant (exactly the outcome desired).

Multiple Patents, Single Use

Now let's return to the situation in which the pharmaceutical is useful only against cancer, but this time the drug requires several patents to produce. If each patent is owned by a different patentee, and each patentee is subject to the policy, this situation would not differ from the simple one first presented. But suppose that one of the patents is owned by PharmaUS, and the rest by nonparticipants. In this case the policy would affect only the single patent owned by PharmaUS and would be less effective as a result. This is one reason that joint adoption of the policy by Japan, the European Union, and the United States would be useful.[6] If the other patents were owned by CiplaIndia, the policy would shift remaining profits to inventors in India, supporting the development of research capacity there.

Finally, suppose that all the patents are owned by PharmaUS. If there are two subsets of the patents that are similarly effective in protecting the innovation, PharmaUS could sue on the basis of one subset in India and use the other to protect its market in the United States. In this case the policy would be ineffective. Whether such situations would significantly reduce the policy's overall effectiveness depends, of course, on how common it is for pharmaceutical innovations to be covered by sets of redundant patents. This deserves investigation. But in most instances, limiting the patents enforced in India to those not useful in protecting the U.S. market would substantially reduce protection in India, making it much easier for a competitor to sell a related product there without triggering an infringement suit.

Multiple Patents, Multiple Products

Next consider a situation with two patents and two products. Suppose that PharmaUS has a patent on a basic innovation that contributes to products for both cancer and malaria. In addition, PharmaUS has a second patent that protects an

adaptation of the basic innovation to make the product more useful against malaria. Producing the malaria product requires both patents, while producing the cancer product requires only the first. As we saw above, since the first patent relates to a cancer product, the firm will choose not to enforce it in India. But since the second patent does not relate to cancer, PharmaUS will choose to enforce that patent in both countries. Thus incentives to invest in research directed toward adapting innovations for uses specific to developing countries are maintained, and any profits from sales of malaria products in India now accrue solely to the developmental research that leads to their discovery.

Research Tools

Research tools are innovations used in doing further research, such as a process for inserting genetic material into cells. Because no product is associated with the use of these innovations, the patents covering them would not be directly affected by the policy. But the licensing fees that owners of research tools can charge depend at least indirectly on the size of the profits that those who use the tools can obtain on the resulting products (with "reach through" royalty contracts, which give the owner a percentage of the final product sales, this relationship is direct). Where patented research tools are important, the outcomes described above simply move back a step to those investing in creating new tools.

What Is "Cancer"? Where Is "India"?

The simplest example of the policy's application assumes that there is a single poor country, India, and a single disease, cancer, for which therapies have a predominantly rich country market. India and cancer are identified in the declaration for the foreign filing license. In reality, such a declaration would specify a set of diseases and a set of poor countries, and a procedure is needed to determine them.

The sets of diseases and poor countries could be specified by an expert committee. A better alternative, however, would be to devise a straightforward, transparent, and objective procedure to determine these groups. This procedure could be given to the U.S. Patent and Trademark Office, which would be asked to update the license declaration periodically. This approach would eliminate the need to convene committees or for the agency to make judgments, reducing the influence of interest groups. Periodic updating would allow countries to "graduate" as they grew richer and would allow the sets of diseases to change in response to new market conditions.

Before turning to the kind of information available for such a procedure, it is useful to clarify the goal: to identify a set of countries {P} and a set of diseases {D} such that, for each of the diseases in {D}, the percentage of the total potential profit that would come from the poor country markets is less than some threshold value z.

Clearly, the smaller the set {P} is, the larger the set {D} can be—and vice versa. Thus there is a choice to be made between having the policy benefit only the very

poorest countries (by lowering the prices of products treating a broad set of diseases) and having the policy benefit a wider group of countries (by defining the diseases more narrowly). The one requirement is to include a sufficient number of countries in {P} to cover the fixed costs of launching an imitative product in their competitive environments. This is not a particularly stringent condition, since the largest fixed cost in the pharmaceutical industry—the expense of discovery R&D and large-scale clinical trials—is not relevant to imitating entrants. The vibrant and competitive pharmaceutical industry in India developed entirely through imitative production (see Lanjouw 1998). A practical approach would be to first define several sets of increasingly poor countries {P} and then determine appropriate sets of diseases for each. Using several groups would lessen the "you're in or out" nature of the policy and help reduce lobbying efforts.

One issue is how to deal with products that are useful against many diseases. If most of the diseases relevant to a product are in the set {D}, the policy applies appropriately. Suppose, however, that only one of the diseases is in the set {D}. The policy would apply on the basis of that one indication, while the relevant market for the product in each country is actually the combined market for the diseases. It would be important to gauge the frequency of such situations and to consider, for example, whether using a classification system or aggregations of diseases might help to minimize them. To some extent profits from products for diseases not included in {D} could still be obtained by enforcing patents on adaptations (see the section on multiple patents and products). Nevertheless, this concern would suggest erring on the conservative side in defining the set of diseases {D}.

Implementing the decision criterion involves two main steps. The first is to measure profits. The second is to determine a reasonable threshold z. For the first, the most important problem is that profit figures are easily manipulated and there is no consistent, comprehensive source for such data. Moreover, the data that are available are not broken out by disease categories. The closest, and fortunately quite reasonable, approximation is the value of pharmaceutical sales. Data on sales are available from IMS HEALTH Global Services, a private database vendor, for disaggregated therapy classes and some 70 countries. In 1998 these countries accounted for 94.4 percent of world GDP measured in purchasing power parity terms (World Bank 2000; Anne Calbazana, IMS HEALTH Global Services, personal communication, April 2001). The value of sales of pharmaceuticals for a particular disease is directly related to what we want to measure—in contrast to the incidence of disease, another obvious contender. Because countries differ to a surprising extent in their use of drug therapies, cross-country statistics on disease incidence would give a very imprecise indication of the relative size of potential drug markets.[7] That said, relative gross sales figures do differ from relative profits in ways that will need to be taken into account in designing the decision criterion (see Lanjouw [2001] for a discussion).

The fundamental decision, of course, is the value of the threshold z. A small value for z—say, 0.02—implies that a disease class will fall under the policy if, for drugs in that class, expected profits from sales in the set of poor countries are less than 2

percent of total global profits. Increasing z would allow the policy to encompass more diseases and confer greater benefits on the poor, but would begin to significantly dampen incentives for research.

Other Policy Options

One response to the proposal outlined here is to ask, "Would it not be simpler for the developing countries to use existing provisions in the TRIPs agreement to lower their drug prices?" Most countries, rich and poor, control the prices of pharmaceuticals. The TRIPs agreement does not restrict such control. Moreover, it allows countries to issue compulsory licenses to attain public health goals. Compulsory licenses are nonexclusive licenses allowing domestic producers to use a protected innovation in return for reasonable royalty payments to the patentee.

The TRIPs agreement does put conditions on the use of compulsory licenses. These include requirements to treat requests for compulsory licenses on their individual merits, to consider a compulsory license only after negotiations with the patentee have failed, and to allow independent review of decisions on compulsory licenses (see Scherer and Watal [2001] for a detailed discussion). In addition, the output produced under a compulsory license must be primarily for domestic consumption.

This section briefly considers price control and compulsory licensing as well as a "compulsory license" variant of my proposal.

Across-the-Board Compulsory Licensing and Price Controls

If the only goal were to lower prices on products developed for rich country markets, either price control or compulsory licensing might be adequate. The problem with price control is that patentees would retain control over sales in the developing country market and if a firm viewed the controlled price as too low, it could simply keep its patented product off the market. Compulsory licensing avoids this problem by allowing domestic producers to sell a patented product. But this solution helps only in countries with some R&D and manufacturing capacity (under current rules, since no one can produce significant amounts for export under a compulsory license, there would be very limited sources of imports or none at all). Because of the procedural conditions noted above, reliance on a compulsory licensing system could also mean substantial delay in the arrival of new drugs on the market.

More important, neither price control nor compulsory licensing offers what the proposal here was designed to provide: a feasible way to allow competitive pricing in some areas while maintaining incentives for private firms to invest in research on diseases specific to poor countries. Private firms now do very little research on pharmaceutical products for the developing world (see Lanjouw and Cockburn [2001] for evidence). There is little doubt that the lack of patent protection in major developing country markets has contributed to this disinterest.

The public sector can be a source of research effort. But its resources are limited by the priorities of government sponsors (for example, just 0.8 percent of the 1999

U.S. National Institutes of Health budget went to tropical diseases, according to Lanjouw and Cockburn [2001]), and we probably should not expect an explosion of new public funding. Thus engaging the private sector could be of real benefit. With the extension of patent protection across all developing countries, we may see the private sector developing products of specific interest to them. How responsive firms will be is hard to predict. But compulsory licensing or stringent price control regimes that limit the returns to discovering new products to treat poor country health problems seem certain to prevent any beneficial redirection of research.

This problem does not arise when developed countries use compulsory licensing. Occasional and nonsystematic compulsory licensing, as practiced in the United States, for example, does not affect firms' R&D priorities, nor does blanket compulsory licensing when introduced by a country (such as Canada) with demand patterns similar to those of countries with strong patent regimes. Countries using blanket compulsory licensing can, to a large extent, free-ride on the incentives provided by those with strong patent regimes. By contrast, if developing countries were to implement comprehensive compulsory licensing, firms probably would purposefully avoid areas of special interest to those countries. There is no free ride for malaria.

Targeted Compulsory Licensing and Price Controls

Could compulsory licensing or price control regimes be structured so as to constrain most tightly the prices of products for global diseases while allowing larger profit margins for products for malaria? Several considerations suggest that the answer is no, at least not in a feasible manner. There are two main problems, one informational and one political.

The first problem is that compulsory licensing is meaningful only if it can be done quickly. Firms considering competitive entry will not even begin the investment that entry requires until they know that they will be able to proceed with production and sales. For this reason Scherer and Watal (2001), in a discussion of experience with compulsory licensing, commend the approach taken by the Canadian licensing board, which set 4 percent as the reasonable royalty payment for all such licenses. By doing this, the licensing board avoided having to investigate R&D costs and market conditions before setting each fee. The average approval time for licenses of only 10 months was possible precisely because no attempt was made to differentiate across products.

To differentiate effectively would require defining categories of products to receive different royalty or pricing treatments and then using a quick method for identifying the category for each product or set of patents. This leads directly to the difficult identification problems addressed above. Moreover, a differentiated compulsory licensing or pricing scheme would give firms every incentive to make categorizing patented products as hard as possible. The scheme would create clear opportunities for lobbying by firms and produce confrontations unlikely to contribute helpfully to the already acrimonious discussions between countries in this area.

Beyond the informational problem, the more difficult aspect of treating products for different types of diseases differently might well be political. Having seen a compulsory license granted for a product for a global disease with a "reasonable royalty" of 1 percent, those suffering from malaria might well object to a "reasonable royalty" of 30 or 50 percent being required of producers of malaria drugs, regardless of the sound economic logic. Domestic political pressure might make differentiation along the lines required by efficiency (that is, with higher royalty rates on patents for diseases specific to developing countries) untenable and result in a structure of incentives far from that suggested by the decision criterion.

My Proposal With a Royalty Payment

Under my proposal and for the specified set of global products, firms effectively obtain either full protection in the poor countries or no returns at all (a 0 percent royalty), depending on their choices. A variant would be to reformulate the declaration for a foreign filing license so as to enable firms to preserve monopoly rights in the rich countries while obtaining some return from the poor countries. For example, firms might declare that they will not prevent the manufacture or sale of drugs for cancer unless they obtain less than a 5 percent royalty.

Although this variant appears to be preferable, because it seems to strike a middle ground, it is not. From a firm's perspective there may be no difference between being held to a 0 percent royalty in 3 countries (my proposal) and being held to a 5 percent royalty in 10 countries. Of course, if the countries {P} and diseases {D} falling under the proposal did not change in going from a 0 to a 5 percent royalty, firms would prefer the 5 percent royalty. But this outcome would no longer accord with the decision criterion. With a 5 percent royalty, either more diseases or more countries should qualify—to the point at which firms would be indifferent between my proposal and this variant. Including more countries, which might be politically attractive, does not even require the positive royalty: the size of {P} can be increased as much as might be desired by reducing the set of diseases {D}.

An important aspect of my proposal is that the actions that make a U.S. patent vulnerable are crystal clear and immediate. Crystal clear because the punishment for falsifying the declaration for a foreign filing license is large and there should be no room for a patentee to do so by mistake. Immediate because patents are time limited. It is of no use to have a mechanism in which obtaining recourse takes so long that the U.S. patent is close to expiring anyway, because the threat of loss of the U.S. market would not inspire firms to behave as desired.

Under my proposal, proceedings to render a U.S. patent unenforceable can begin on the day a suit is filed in India. In contrast, a declaration reformulated as above would have to be shown to be false on the basis of the outcome of a suit in India—that is, only after CiplaIndia had successfully proven that royalties of at least 5 percent had been paid. Court proceedings can move slowly anywhere, but particularly in a developing country, so there would appear to be much scope for the patentee to delay the progress of such a case.

Conclusion

I have outlined a policy for lowering the price of pharmaceuticals in developing countries for important diseases while maintaining the R&D incentives of research firms. The policy would use aspects of patent law (such as the foreign filing license, rules of estoppel, and priority procedures), features of litigation and the drug approval process, and available data sources in ways not originally intended to arrive at a mechanism that serves its purpose.

The new rules would give firms new incentives, and in responding to these they would choose not to suppress competition in markets where the profit potential is small. Rarely would the procedure to render a patent unenforceable be invoked, because firms would alter their behavior to avoid this outcome. Never would an outside body have to make the difficult judgment about what a patent is for, because the patentee would have an incentive to provide this information whenever it is needed (in the event of an infringement suit). The policy would require no changes in international treaties and only minor changes to the legal codes of the developed countries that implement it. As a result, it would be straightforward to implement.

How beneficial would this policy be? This is a difficult question to answer given our vague understanding of the importance of any change in patent laws, including the major changes now under way as countries enter into compliance with the TRIPs agreement. But the "rich country" diseases to which the policy would apply account for a significant share of the disease burden in the poorest countries and weigh heavily on the poorest people in those countries. Clearly, allowing these countries to have competitive suppliers would lower prices for consumers. Without the policy, developing country consumers would face either the domestic monopoly price or a yet higher world market price if global pricing concerns make patentees reluctant to use tiered pricing.

The gain from allowing competition depends on the availability of substitute products and the demand conditions in the poor countries for drugs for these diseases. Data are available for estimating the detailed demand models needed to produce plausible projections of price reductions and their effect on consumer welfare in poor countries. This work remains to be done.

Not being exclusive to poor countries, the diseases to which this policy would apply are not viewed as "poor country" diseases and therefore have not received much attention in development debates over patent policy. They should. Rich countries, with some creativity in designing their patent systems, can use the excellence of their scientific research to give a big welfare boost to poor countries while supporting the full implementation of the TRIPs agreement in the developing world.

The policy can also be used in the developed world's self-interest. Large issues are at stake in the enforcement of intellectual property rights and safety regulations in a world of global Internet sales. Resolving these will require international cooperation and therefore a turn away from the polarized discussions of recent years.

Positive initiatives are needed to demonstrate that the developed world can be flexible and thoughtful in pursuing the interests of its own constituencies. This policy could provide such an initiative.

Notes

1. The figures in table 2 are provided to give an impression of the distinct differences in the global distribution of markets for therapies for the two types of diseases highlighted here. The data have some weaknesses and should not be taken too literally. For example, DALYs lost fall with pharmaceutical consumption, and for this reason the percentages in the second column underestimate the importance of rich country markets.

2. Other high-income countries whose patent law has some form of the domestic filing requirement for residents include Belgium, Denmark, Finland, France, Germany, Greece, Italy, Luxembourg, Portugal, the Russian Federation, Spain, Sweden, and the United Kingdom.

3. Forfeiture is not generally favored by courts as a remedy for breach of contract. The more usual remedy is damages. But rendering a patent unenforceable is the standard remedy in this context. It has been put into effect, for example, in cases where a patentee knowingly misrepresented prior art to the patent office or made a false declaration about the adequacy of the patent specification in revealing the invention. Note that the damage here would be to the integrity of the U.S. patent system, not to the developing country.

4. The physical movement of product does not appear to be a primary concern of the industry now. The bigger block to tiered pricing today is the reluctance of rich country consumers to tolerate lower prices in poor countries or what would appear to be their lack of awareness. For example, in the United States legislative efforts in 2000 to remove Food and Drug Administration controls on imports—efforts driven by anger over Canadian prices—do not distinguish between poor and rich source countries, nor did the public discussion note the potentially negative implications for poor countries. The political pressure and regulation resulting from such public attitudes cause prices in one country to spill over to prices in another—even if no product crosses national borders. Firms' natural response is reluctance to charge lower prices in poor countries.

5. In a statement to the U.S. Congress on 25 May 2001 William Hubbard, a commissioner of the FDA, stated that "Internet technology can obscure the source of the product.... [The agency] believes that illegal online drug sales pose a significant public health risk. Consumers...may be targets of unscrupulous business practices, such as the selling of unsafe, unapproved, expired, counterfeit, or otherwise illegal drugs. The sale of drugs to U.S. residents via foreign websites is an extremely challenging area.... FDA efforts are mostly limited to requesting the foreign government to take action" (Hubbard 2000, p. 5).

6. Joint adoption of the policy would also make it difficult for firms to avoid the policy by claiming to invent in subsidiary locations outside the United States. Well-developed case law relating to the identification of inventors limits firms' ability to simply choose any employee, based on convenience, to designate as the inventor.

7. There are two other problems with figures on disease incidence as well as mortality. First, they can be strongly affected by current drug consumption. Thus the larger the market is for drug therapies for a disease, the lower are the incidence of the disease and the related mortality (HIV/AIDS is a good example). Second, just as for profits, the data on disease incidence and mortality do not exist in the comprehensive and consistent form needed.

References

Adelman, Martin J., Randall R. Rader, John R. Thomas, and Harold C. Wegner. 1998. *Cases and Materials on Patent Law.* American Casebook Series. St. Paul, Minn.: Westgroup.

Deardorff, Alan V. 1992. "Welfare Effects of Global Patent Protection." *Economica* 59: 35–51.

Hubbard, William. 2000. "Statement Before the Subcommittee on Oversight and Investigations, Committee of Commerce, U.S. House of Representatives, May 25, 2000." [http://www.fda.gov/ola/2000/internetsales.html].

Kremer, Michael. 2001a. "Creating Markets for New Vaccines—Part 1: Rationale." In Adam B. Jaffe, Josh Lerner, and Scott Stern, eds., *Innovation Policy and the Economy*. Vol. 1. Cambridge, Mass.: National Bureau of Economic Research.

———. 2001b. "Creating Markets for New Vaccines—Part 2: Design Issues." In Adam B. Jaffe, Josh Lerner, and Scott Stern, eds., *Innovation Policy and the Economy*. Vol. 1. Cambridge, Mass.: National Bureau of Economic Research.

Lanjouw, Jean O. 1998. *The Introduction of Pharmaceutical Product Patents in India: "Heartless Exploitation of the Poor and Suffering"?* NBER Working Paper 6366. Cambridge, Mass.: National Bureau of Economic Research.

———. 2001. "A Patent Proposal for Global Diseases." Yale University, Department of Economics, New Haven, Conn. [http://pantheon.yale.edu/~lanjovw].

Lanjouw, Jean O., and Iain Cockburn. 2001. "New Pills for Poor People? Empirical Evidence After GATT." *World Development* 29(2): 265–89.

Malueg, David A., and Marius Schwartz. 1994. "Parallel Imports, Demand Dispersion, and International Price Discrimination." *Journal of International Economics* 37: 167–95.

Nordhaus, William. 1969. *Invention, Growth and Welfare.* Cambridge, Mass.: MIT Press.

OPPI (Organization of Pharmaceutical Producers of India). 1996. *Presentation to the Select Committee of Members of Parliament (Rajya Sabha) on Patents (Amendment) Bill 1995.* Mumbai.

Pappas, Gregory, Taslim Akhtar, Peter J. Gergen, Wilbur C. Hadden, and Adul Qayyum Khan. 2001. "Health Status of the Pakistani Population: A Health Profile and Comparison with the United States." *American Journal of Public Health* 91(1): 93–98.

Scherer, F. M., and Jayashree Watal. 2001. "Post-TRIPs Options for Access to Patented Medicines in Developing Countries." Harvard University, Cambridge, Mass.

U.S. Department of the Treasury. 2001. "Gasoline and Diesel Fuel Excise Tax: Registration Requirements." [http://ftp.fedworld.gov/pub/irs-regs/td8659.txt].

Watal, Jayashree. 2000. "Pharmaceutical Patents, Prices and Welfare Losses: A Simulation Study of Policy Options for India Under the WTO TRIPS Agreement." *World Economy* 23(5): 733–52.

World Bank. 1999. "Accelerating an AIDS Vaccine for Developing Countries: Issues and Options for the World Bank." Washington, D.C. [http://www.iaen.org].

———. 2000. *World Development Report 1999/2000: Entering the 21st Century*. New York: Oxford University Press.

WHO (World Health Organization). 1996. *Investing in Health Research and Development: Report of the Ad Hoc Committee on Health Research Relating to Future Intervention Options.* Geneva.

———. 1999. *World Health Report 1999.* Geneva.

Comment on "A Patent Policy Proposal for Global Diseases," by Jean O. Lanjouw

Rachel K. Gesami

Jean Lanjouw presents a compelling argument for redefining the debate about ways to improve on the current patent regime by recognizing the differences in markets and the implications of those differences for optimal patent protection. She puts forward a logical, plausible, and well thought out proposal for a policy that would help streamline the largely confused pharmaceuticals market in developing countries, as I illustrate through discussion of the Kenyan situation. Lanjouw blames the unavailability of drugs for global diseases to poor people in developing countries on the lack of a sound patent policy in developed countries.

The global diseases Lanjouw identifies are generally degenerative diseases (cancer, diabetes, heart disease), which require lifelong use of medicines and therefore burden not only the patients' health but also their family budgets. HIV/AIDS afflicts people in the same way, causing increasing impoverishment for households. The Joint United Nations Programme on HIV/AIDS (UNAIDS 2000) estimates that 70 percent of those infected with HIV—24.5 million people—live in Sub-Saharan Africa, home to only about 10 percent of the global population. In Botswana one in three people (35.8 percent) is infected. The AIDS epidemic has created 11 million African orphans, 90 percent of all orphans in the region. AIDS is expected to shorten life expectancy in Africa from 59 years to 45 between 2005 and 2010. Africa is also the continent most affected by malaria—with the highest morbidity rate (80–85 percent) and mortality (about 1 million deaths annually) due to that disease.

All these diseases, in addition to their pathogenic effects, are associated with huge economic burdens on households, with those affected directing large shares of their income and wealth toward treating and caring for the sick. Malaria contributes to a GDP loss of about 1.3 percent a year in Africa, beyond the costs at the household level. The economic burden of these diseases will lead to a worsening of poverty in Africa, which is already engulfed in poverty (Ali, Mwabu, and Gesami 2000).

Treatment for most of the diseases affecting Africa remains largely unavailable to African consumers. The current unavailability of drugs for global diseases to poor

Rachel K. Gesami is an economist, currently manager, External Liaison and Communications, at the African Economic Research Consortium in Nairobi, Kenya.

people in developing countries is blamed on market failure, resulting largely from patent policy protecting rights to innovations and technology in pharmaceutical development. The patents and associated pricing strategies guarantee returns and profits on innovations to the manufacturers, at the expense of those suffering from diseases.

Reducing prices for drugs to increase affordability and access to treatment not only relieves pain and suffering and prolongs lives. It also aids disease prevention by creating incentives for people to engage in responsible behavior and to seek prompt treatment once it becomes available. In the case of HIV/AIDS, the availability of treatment may improve the success of voluntary testing services, which are critical in preventing transmission of HIV. Indeed, evidence shows that the absence of treatment for AIDS-related conditions has precluded the success of voluntary testing and counseling programs in many parts of the world. The same principle applies for other diseases, such as malaria, that are having equally debilitating effects in Africa.

A policy that improves on the current patent regime along the lines proposed by Lanjouw could contribute immensely toward getting drugs to poor countries at affordable prices. Lanjouw has provided suitable, if complex, analogies to define proposals for dealing with the current patent policy regime.

In my comments I refer largely to the African situation to contextualize the proposals Lanjouw discusses, starting with a brief discussion of issues relating to HIV/AIDS and changes in the patent policy.

HIV/AIDS and Changes in the Patent Policy

Contrary to the suggestion that the HIV/AIDS crisis calls for separate policies, adopting a broad patent policy regime covering drugs and treatment for other, equally devastating diseases as well would be beneficial. While the crisis presented by HIV/AIDS warrants immediate attention, reacting to crises and emergencies diverts needed attention from many other diseases. Moreover, treating the HIV/AIDS crisis as special is as good as stigmatizing HIV/AIDS patients, a rampant phenomenon today. I therefore continue to refer to HIV/AIDS in my comments.

It might be appropriate here to note that HIV/AIDS is managed by two classes of medicines. The first consists of drugs for treating opportunistic infections associated with the syndrome. The second, and more important, category is the antiretroviral drugs, designed to knock out the virus.

Most of the drugs in the first class, except for fluconazole (Diflucan®), which is still under patent, are available as generics and included on the Essential Drugs List established by national ministries of health and approved by the World Health Organization. Those in the second class, however, which drastically improve patients' health and reduce the spread of infection by greatly reducing the viral load, are still under patent because they are recent innovations. As is well known, drugs under patent are more expensive than their corresponding generics. As a result, most governments of developing countries cannot afford to provide these drugs to HIV/AIDS patients. Nor can most of the HIV/AIDS patients afford these drugs pri-

vately. This has inevitably led to high mortality across all population groups (including workers) in these countries.

Although patented drugs for other diseases are also out of reach for developing country governments and consumers, the antiretrovirals are an extreme case. Even opportunistic infections associated with HIV/AIDS cannot be effectively managed because of the high cost of treatment. With the advent of HIV/AIDS, some drugs still under patent are being included on countries' Essential Drugs List because they are part of the basic treatment for HIV/AIDS. At the same time, increasing resistance of infectious agents to drugs has rendered the older and cheaper drugs inappropriate for disease management.

There is therefore a need for developing country governments to address the HIV/AIDS crisis, especially access to care and treatment. In particular, they need to define ways of making antiretroviral drugs available. One strategy, of course, is to design ways of reducing prices for antiretrovirals to levels within the economic capacity of these countries. In some instances drug prices have been reduced by more than 50 percent for purposes of competition, such as for Tagamet® (cimetidine). Thus price reductions are possible for the sake of developing country markets and for moral reasons.

These concerns extend to medications for malaria-type diseases, medications that few African households can afford. The mortality due to these diseases, which are essentially manageable, is enormous. Developing countries devote a much greater share of their drug expenditure to antiparasitic drugs than do developed countries (see table 4 in Lanjouw's article). But the expenditure shares on anti-infectives are comparable for developed and developing countries. Thus infections too are global diseases. A leading cause of death in Sub-Saharan Africa, infections mostly result in acute illnesses: either the patients are cured or they die. This is an absurd situation. For example, poor patients requiring an expensive lifesaving antibiotic (Fortum) usually die because they cannot afford the drug immediately and thus prevent the medical emergency. In contrast, patients in developed countries have easy access to lifesaving injections.

It is true that cancer, diabetes, and heart disease occur in developing countries, across the income spectrum. The drugs for global diseases such as cancer account for a big share of the per capita expenditure on medicines in developing countries. The drugs are overpriced in these markets: overpricing is known to often lead to higher prices for lifesaving medicines in Africa than in Europe and North America.

The pricing of drugs in developing countries is not the only issue; it needs to be remembered that well over 90 percent of Africans live below the poverty line and are uneducated. The poverty and lack of education in Sub-Saharan Africa prevent the purchase and appropriate use of the medicines. Consequently, for chronic diseases such as these, there are only two options: to have people die as a result of non-compliance with drug protocols or to design mechanisms for ensuring access to the drugs at nominal prices, especially through public sector health facilities, as has been done in Kenya under its cost-sharing system.

The Policy Proposal

The features of the policy Lanjouw proposes are generally attractive. But for developing countries, full control of the policy by the U.S. government would be oppressive. Because developing countries have a stake in the sourcing, supply, and sale of medicines, they obviously should also have a say in the control of the policy. Moreover, they should not exchange their sanctioning of compulsory licensing for this stand.

At the same time, access to medicines should not depend on local pressure, since it is not entirely true that such pressure can succeed in expanding coverage to all diseases. Instead, systems capable of responding to health emergencies need to be established. In established systems an emergency such as AIDS will stand out clearly and be addressed appropriately. A systematic response to emergencies removes the effect of local pressure, whose response is often ad hoc. Consider the analogy of a famine in which a particular region of a country is declared a disaster area and appropriate measures are taken to address its needs. The rest of the regions do not demand relief food just because some has been delivered into the country.

To illustrate how the policy would work, Lanjouw uses a complex analogy based on a contract between a U.S. pharmaceutical company, "PharmaUS," and the U.S. Patent and Trademark Office. India, with its 0.9 billion people, has been termed a small market. Most pharmaceutical companies would prefer to manufacture their products in India because labor and other inputs are cheap. Such manufacturing would be aimed entirely at lowering product prices to competitive levels. In Kenya, for example, India and other Asian countries have been a cheaper source of parallel imports of innovator products.

Lanjouw focuses on the patentee's possible loss of sales in developing countries in the event of patent infringement, illustrating that it sometimes might not be in the patentee's best interests to sue. That was the case in the much-publicized conflict in South Africa, where multinational pharmaceutical companies opted to drop an analogous suit.

It should also be pointed out that the benefit from patenting pharmaceuticals involves much more than the product's active ingredients. Even when a patent is infringed or expires, the brand name can still offer returns: the brand name remains protected, and "brand loyalty" is a big factor in business. Only the monopoly of the market is denied.

Parallel Imports and Low-Cost Sources of Supply

Lanjouw recognizes the apparent problem of seepage across borders as a major drawback of market separation. Indeed, the developing country governments concerned would need to participate fully in the enforcement of patent policies. Otherwise, the policies would be inoperative. The problem of seepage has already been noted in Kenya.

One solution is for governments at the point of manufacture to control the movement of the products. In Kenya such a system could be easily implemented, since the infrastructure exists for controlled distribution of medicines.

At the same time the destination governments would need to enact legislation to discourage inflows of unregistered medicines and step up market surveillance to detect such practices. Lanjouw's suggestion of instituting country-specific tags, such as color codes, is consistent with current practice. In India, for example, medicines produced for the domestic market are labeled "for sale in India only," while packages of drugs for export are appropriately labeled "for export."

Linkages

Lanjouw's argument relating to the claim by PharmaUS that all its products are for malaria and the claim by the generics manufacturers (CiplaIndia and USGeneric) that the disputed products are for cancer is questionable. Let's remember that the use of particular medicines is not open to arbitrary assignment. A cancer or antimalarial drug is globally known as such, and a chemical compound patented for a specific disease will therefore be widely understood to treat that disease. Such a scenario is unlikely to arise in health matters. Lanjouw's suggestion that all products marketed in the United States be approved by the Food and Drug Administration for specific indications is indeed the standard practice. Nonetheless, the argument she presents about patents, while rather abstract, is consistent with the basic principle of the proposal.

Lanjouw also explores a more complex situation in which a pharmaceutical company would specify, with the advice of an expert committee, a set of diseases and a set of countries in its basic declaration for a foreign filing license. But it would be difficult to group developing countries for this purpose, given minor political differences and diverse socioeconomic situations. Efforts at regional cooperation in Africa are known to have failed as a result of such differences, especially with civil strife and political conflict the order of the day. Moreover, if countries were to be grouped, seepage of products could occur across their borders, driven by price discrepancies. There is already a problem, though on a small scale, with seepage of low-priced medicines from Uganda into the Kenyan market.

The Kenyan Case

The government of Kenya has long recognized the need to ensure that drugs are affordable and can therefore be made available to its citizens. A government policy was formulated for these purposes in the 1970s in the form of an Essential Drugs List. This policy conformed well with the World Health Organization's health policy and had that organization's full support, and it is acceptable under the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs). The essential drugs typically are not under patent and are mainly generics. To reduce costs, the Kenyan government ensures that the essential drugs are locally manufac-

tured by an agent determined by a tendering process. The drugs are supplied to the central government stores for distribution to public health facilities.

Parallel importation of drugs is not allowed under Kenyan law, though it nevertheless occurs. This illegal importation of drugs takes two forms:

- People travel to particular countries (usually in Asia), buy small quantities of the drugs, and carry them back into the country in their baggage. Surprisingly, innovator products bought from retail outlets in those countries are far cheaper than the locally available products.
- The innovator company or an agent for a branded product imports its products from a market other than the registered source, at a lower price. This practice shows that multinational drug companies do not have fair price policies. Worse still, the companies or their agents sell the products at the official "European source" price—they do not pass the lower prices on to patients.

A major disadvantage of parallel importation in developing countries such as Kenya is that it becomes an avenue for counterfeits, destroying the market for innovator products. Parallel importation in Kenya is economically driven. It could be curbed if innovator companies marketed their products at realistic prices reflecting the country's per capita income.

Compulsory licensing (and controlled distribution), in contrast, could be beneficial to developing countries. Compulsory licensing has never been invoked in Kenya. But with the exorbitant prices of antiretrovirals, compulsory licensing for the manufacture of these drugs in Kenya may be inevitable. The Kenyan government may then be able to provide them as essential drugs. To prevent illegal seepage across borders, all the drugs would need to go to the central government stores for distribution. The government has mechanisms for monitoring the movement of such medicines, successfully used in programs for leprosy and tuberculosis and in the Kenya Expanded Programme on Immunization. Government control of the distribution of antiretrovirals could also greatly improve patient monitoring and the capture of epidemiological data on HIV/AIDS.

The Kenyan Government's Position on Patents

The constitution of the World Health Organization enshrines the following principle, binding on all its members, including developing countries:

> The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being.

This principle rests on the theory of health security, which requires that all individuals have access to health care that is affordable, relevant, and of good quality. It does not mean that states are obligated to guarantee health security for their population, as this is clearly impossible for most of the world's governments. But it does mean that every government has a recognized obligation under international law to

do its best to realize this dream. Access to essential medicines is clearly a vital part of establishing health security.

There is no doubt that the Kenyan government recognizes and accepts this responsibility. A member of parliament is quoted as asking, "How can we be denied access to drugs that prolong life when our people are dying?" (McNeil 2000). It is also clear that the government faces an uphill task in fulfilling this obligation. The most significant barrier is intellectual property rights, which allow artificially high prices for innovator products to compensate the inventor for the investment in research and development.

The Kenyan government has proposed an amended Industrial Property Bill that is likely to lead to an intellectual property rights system reflecting the need for health security in Kenya. All the proposed provisions conform to Kenya's obligations under the TRIPs agreement and other patent-related treaties to which it is a signatory.

The recent outcry in South Africa over patents is a clear indication of the mood of the people of Sub-Saharan Africa with respect to HIV/AIDS. In Kenya HIV/AIDS has been declared a national emergency—and this emergency calls for suspending regulations antagonistic to securing access to drugs. Section 80 of the Proposed Amendments to the Industrial Property Bill addresses the exploitation of patents by the Kenyan government through its authorized agent (Kenya, Industrial Property Office 2000). It identifies health and nutrition requirements as conditions for such exploitation. Under this section the patent owner is entitled to a royalty determined by the relevant Kenyan authority and commensurate with the economic value of the patent. The market is limited to Kenya.

Concluding Remarks

In these comments I have presented the example of Sub-Saharan Africa as justification for a changed patent policy. I have also addressed the practical aspects of the policy Lanjouw proposes. While I agree with the declaration for foreign filing licenses that forms the basis of Lanjouw's argument, I would like to emphasize that developing country governments need to participate in formulating such patent policies.

The Essential Drugs List of the World Health Organization can serve as a guide to the drugs that a developing country could produce under compulsory licensing in a health emergency. HIV/AIDS is such an emergency—in Sub-Saharan Africa and indeed globally. It is HIV/AIDS that appears to be inspiring the examination of the moral and ethical issues relating to patents.

A general message in Lanjouw's article is that patents must be respected. But the patent regime should help developing countries to manage their health systems. And it needs to reflect the importance of ensuring the availability of drugs in these countries. For their part, developing countries should pursue health care reforms to ensure that mechanisms are in place to cushion the effects of health crises on the poor, such as exemptions for the poor in government fee structures.

Lanjouw's article is weak in addressing the quality implications of innovator products and generics, although this is an important issue in patent regimes. In Kenya it

has been observed that patented medicines occasionally have quality problems. Even in developing countries drug manufacturing technology and quality assurance have generally improved significantly in recent times and shall continue to do so.

Finally, we should recognize that the successful formulation and implementation of any policy depend on the global political economy and need to be addressed in appropriate forums. Forcing countries to adhere to some policies may be expedient, but in moving toward a new patent regime it is essential that a participatory approach be employed. For a nation, the health of its people overrides any other considerations, especially when it faces a pandemic like HIV/AIDS. Otherwise, that nation may cease to exist.

References

Ali, Ali Abdir Gadir, Germano Mwabu, and Rachel K. Gesami. 2000. "Poverty Reduction in Africa: Challenges in the 21st Century." Collaborative Research Project Working Papers Series, CRC-2-12. African Economic Research Consortium, Nairobi, Kenya.

Gesami, Rachel K. 2000. "Health Care Financing in Kenya: An Empirical Analysis." Ph.D. dissertation. University of Nairobi, Department of Economics, Nairobi, Kenya.

Gunatilleke, Godfrey, and Aleya El Bindari Hammad, eds. 1997. *Health: The Courage to Care*. Geneva: World Health Organization.

Kenya, Industrial Property Office. 2000. "Proposed Amendments to the Industrial Property Bill." Nairobi.

Kenya, Ministry of Health. Various years. *Kenya Essential Drugs List*. Nairobi.

Kenya, Ministry of Health, Drugs and Poisons Licensing Board. 2000. "Drugs and Chemists Price List." July/August. Nairobi.

Kenya, Ministry of Health Task Force. 2001. *Strategies to Improve Access to Essential Medicines in Kenya, Report of Meeting Proceedings, Lion Hill Hotel, Nakuru, Kenya*. Nairobi.

Kibwage, I. O., E. Ayiemba, and C. O. Ondari. 1998. "Drug Utilization and Cost Patterns in Selected Health Care Facilities in Kenya." *East and Central African Journal of Management Sciences* 1(1): 11–14.

Kibwage, I. O., and J. K. Ngugi. 2000. "Sulphadoxine/Pyrimethamine Tablet Products on the Kenyan Market: Quality Concerns." *East and Central African Journal of Management Sciences* 3(1): 14–19.

Kibwage, I. O., C. O. Ondari, I. G. Mureithi, J. K. Thuranira, and J. Hoogmartens. 1998. "Analysis of Co-Trimoxazole Products on the Kenyan Market." *East and Central African Journal of Pharmaceutical Sciences* 1(2): 34–38.

Lanjouw, Jean O. 2001. "A Patent Policy Proposal for Global Diseases." Paper for Economic Studies Program, the Brookings Institution, Washington, D.C.

McNeil, Donald, Jr. 2000. "Prices for Medicine Are Exorbitant in Africa, Study Says." *New York Times,* 17 June.

Mwabu, Germano. 2000. "Poverty and Malaria in Africa: Research and Policy Agenda." University of Nairobi, Department of Economics, Nairobi, Kenya.

Scientific Media Services. 2000. "Chemists and Druggists Monthly Price List." January. Nairobi, Kenya.

UNAIDS (Joint United Nations Programme on HIV/AIDS). 2000. *Report on the Global HIV/AIDS Epidemic*. Geneva.

Comment on "A Patent Policy Proposal for Global Diseases," by Jean O. Lanjouw

Beatriz Zurita

Understanding and removing barriers to access to drugs, especially for developing countries, has been an elusive goal. As Morten Rostrup argues in his article, medicines to treat diseases prevalent in developing countries are too expensive mainly because of patent protection, lack of research and development, or insufficient production due to inadequate return on investments. Nonetheless, important health advances have been made over the past 20 years, thanks to the leadership of the World Health Organization (WHO) in promoting essential drugs and the production of generic drugs.

Although the essential drugs strategy was very successful—the number of people with access to essential drugs has doubled—WHO's Medicines Strategy for 2002–03 now recognizes that a third of the world still lacks access to drugs. Four factors are primarily at play: affordable prices, rational selection, sustainable financing, and reliable health and supply systems. The April 2001 Workshop on Differential Pricing and Financing of Essential Drugs, sponsored by WHO and the World Trade Organization (WTO), discussed the impact of tariff and nontariff trade barriers on developing country access to essential drugs and the impact of patent protection on the prices of essential drugs.

Jean Lanjouw's article is very timely. It proposes a compelling strategy to improve the current patent system while providing incentives to invest in research and development (R&D) for diseases that would otherwise be ignored. Pharmaceutical companies can continue to invest in R&D for diseases whose treatment is profitable in industrialized countries, even as incentives are strengthened for R&D in medicines for the health problems of greatest concern in the developing world.

Most of the discussion at the WTO–WHO workshop concentrated on drugs that are already available and on issues of price—producer costs, tariffs and taxes, distribution margins at the wholesale and retail levels, and demand and supply factors (such as differences in purchasing power, market structure, and exchange rates).

Beatriz Zurita is head of planning and development at the Medical Directorate of the Mexican Social Security Institute.

Competitive pricing is also related to patent issues, including protection of intellectual property rights, anticompetitive pricing, corruption, parallel imports of patented products, and export controls, and to other factors that affect entry and competition, such as the domestic generic drug industry or policies that facilitate imports of generics from the cheapest sources.

It would be important to determine the degree to which the market segmentation and differential pricing strategies proposed in the WTO–WHO workshop can complement a strengthened patent system. Specifically, how will the countries and institutions that are now moving to establish funds for drugs and vaccines and tax exemptions for corporate drug donations react to additional support through the patent system?

This is especially relevant in a context of low purchasing power, where most spending on drugs is out of pocket and where health systems are poorly organized and may not be able to cover the storage and distribution costs for donated drugs. This is the case in the new agreement signed between Merck and public institutions in Mexico. The agreement may face difficulties because of the different purchasing power of the institutions involved—the Social Security Institute spends more than twice as much as other public institutions and has more money to spend on drugs. The new agreement will change the market for HIV/AIDS drugs, and it is unclear who will benefit most from access to the drugs. Inequality in the country might even increase if the better-off can afford the distribution costs while the poor still do not have access to the drugs.

This Page Intentionally Left Blank

Health, Income, and Economic Development

Anne Case

There is a strong positive relationship between income and health throughout the world. If part of this association represents a causal effect from income to health, the maintenance and support of incomes become a potential policy instrument for promoting health among specific populations or groups. Policies for income support, such as transfers to poor people or the elderly, are instruments that should be assessed, along with the provision of health services, for their ability to improve health. Whether there is a causal link from income to health, and its size, are important research issues for those interested in health in developing countries. This article uses data from an integrated survey of health and economic well-being in South Africa to examine the impact of the old age pension on the health of pensioners and the prime-age adults and children who live with them. It finds evidence of a large and causal effect of income on health status, one that works at least in part through improved sanitation and living standards, in part through better nutritional status, and in part through reduced psychosocial stress. The pension is used to upgrade household facilities, and some of the improvements have health consequences. Governments interested in improving health status may find the provision of cash benefits to be one of the most effective policy tools available to them. Cash provides a yardstick against which other health interventions should be measured.

O   ther articles on health and development in this volume focus on health service delivery. It is generally implicit in such a focus that access to health services, and the quality of care administered, are the central determinants of health outcomes. It is a short step from there to taking improvements in the delivery of health care as a central focus for an organization like the World Bank. Such a focus is consonant with the World Health Organization's (WHO) (2000, p. 9) *World Health Report 2000*, which stresses the importance of health delivery in health out-

©2002 The International Bank for Reconstruction and Development / THE WORLD BANK

comes, claiming that "numerous studies beginning in the 1970s have consistently found that preventable deaths…have fallen at a faster rate than other deaths." The WHO report claims that where the connection between health delivery and health outcomes is weak, inefficient delivery is the cause.

In this article, without denying the importance of medical services, the focus shifts to the role of income in promoting health. This is an area largely untouched by economists, although it has long been a focus of public health research (see Adler and Ostrove [1999] for an overview). Within countries income is strongly correlated with health outcomes, and policy recommendations that provide for income transfers to the poor or the promise of increased earnings capacity may prove to be as important for health outcomes as those calling for additional funds for service provision, especially in settings where the capacity to deliver health services is weak.

An understanding of whether and how income generates better health is important for public health policy because the share of resources devoted to different policy options should depend on their relative effectiveness. This is not to gainsay the importance of public health campaigns to provide clean water, eradicate malaria, vaccinate children, deliver AIDS drugs in developing countries, or improve the quality of health services or access to them. However, weak links in the chain of public provision identified by many researchers (see Filmer, Hammer, and Pritchett [2000] for an overview) strengthen the case for considering alternatives and for quantifying the causal impact of income on health outcomes.

The efficacy of spending money to improve health delivery should be weighed against that of, say, improving school quality or increasing educational attainment to promote health. The relative merits of a new school or a new clinic should also be weighed against those of increasing the incomes of the poor as a method of improving health outcomes. Poor people may spend part of this additional income on medical care or on goods associated with better health—more nutritional food, better housing, or safer places to live, for example. They may also derive health benefits directly from money, for example, by relieving the stress and susceptibility to infection associated with the daily trials of coping on inadequate income.

Health improvement through income redistribution has become a strategy debated in the mainstream. The Acheson (1998) report, *Independent Inquiry Into Inequalities in Health Report*, commissioned to study health inequalities and ways to reduce the growing gaps in illness and death in the United Kingdom, recommended "policies which will further reduce income inequalities" (p. 36) as a mechanism to improve the health of the poor. An understanding of the merits of these different mechanisms is of first-order importance.

This article provides estimates of the causal impact of income on health outcomes in one transition economy, South Africa. South Africa's experience is interesting for many reasons. South Africa is a transition economy with both a "developed" country population and a "developing" country population. Access to first world medical care, although largely limited by race and wealth, is available to some of South Africa's poor. (Historically many poor people in Cape Town used Groote Schuur Hospital—home of the world's first successful heart transplant—as their local clinic,

while those in rural areas struggled with poorly provisioned clinics.) South Africa has also witnessed many innovative policy experiments since the change of government in 1994, including the provision of large pensions to all elderly citizens. The old age pension is used here to estimate the causal effect of income on health.

The Health Gradient

People in wealthier countries live longer and have lower rates of illness, on average, than do people in poorer countries. Within countries, wealthier people live longer than poorer people, and as countries become wealthier average life expectancy increases. This phenomenon is not limited to the bottom end of the income or wealth distribution. Indeed, the *gradient* in health status—the phenomenon that relatively wealthier people have better health and longevity—is evident throughout the income distribution.

Evidence on the relationship between income and health comes from historical analyses, cross-country comparisons, and country analyses using microeconomic data. (The literature is vast, and only a small part of it is mentioned here, to provide context.) McKeown (1976) and Fogel (1994) have argued that improvements in longevity in the 19th century were driven not by advances in medicine or public health but by improvements in nutrition, largely a result of higher incomes. Other researchers, including Preston (1975, 1980), Szreter (1988), and Easterlin (1998, 1999), present historical and cross-country evidence on shifts in the health production function, shifts they attribute less to income or income growth than to public health efforts (particularly sanitation, vaccination, and vector control) and to advances in health technologies, such as antibiotics.

At the microeconomic level researchers have used household or individual-level data to document a significant positive relationship between income and measures of health status in both developed and developing countries. For developed countries both a channel from health status to income and feedback from income to health are expected (Smith 1999; Adler and others 1994). For developing countries much of the research on the connection between income and health has focused on infant mortality, for which the feedback from health to income is quite limited (see, for example, Gwatkin [2000] and Wagstaff [2000]).

Even if one accepts that public health and health technology have reduced mortality in rich countries over time and in developing countries since World War II and that better delivery of health services is an important goal in poor countries, it is important to explore the use of income transfers as a tool of health policy. If individual income is causally related to individual health, and if organizational capacity, political will, or control of corruption hampers service delivery, income transfers may do more for the health of the poor than would committing more money to an ineffective health delivery system.

Difficulties Quantifying the Causal Effects of Income

Whether and how money can be exchanged for better health have proved difficult to quantify, particularly for working-age adults, not least because pinning down the

effect of income on health is difficult when health simultaneously influences income. In addition, the mechanisms through which money is translated into health may vary with the level of development; no single mechanism (or set of mechanisms) may be at work everywhere. Finally, until recently there have been few sources of data with the details about both health and income needed to allow reasonable micro-level examinations of the phenomenon.

If people are observed at only at one point in time and if illness causes them to work fewer hours, they will be observed to have low health status and low income. One would not want to conclude that the current lack of income caused the illness. Quantifying the impact of income on health requires a tool that separates the correlation working from health to income from that working from income to health. This is not always possible, nor is it always possible to rule out a role for third factors—variables that determine both income and health status.

The causes of poor health status in the developing world may have little in common with those in industrial countries. In developing countries the risks may be primarily from infectious disease, lack of clean drinking water, and inadequate diet. In industrial countries the risks may be primarily from chronic diseases associated with lifestyle and reduced physical activity. The causal links between health and income may be specific to the diseases people face, and for this reason, quantifying those links may be especially difficult in transition economies, where people living in the same household may be struggling against both infectious disease (the "unfinished agenda" of developing countries) and chronic disease and violence (the "'emerging agenda" of industrial countries; Kahn and others 1999); obese women may be living near or with malnourished children (Case and Wilson 2001). The challenge of comparing across levels of development may be rewarded, however, if such comparisons bring to light universal mechanisms at work. (Psychosocial stress is apt to be found in every environment, to take one example; see Marmot [1999].)

In many developing countries health authorities see little need for health surveys, apart from measuring access to health facilities, and the surveys that are conducted tend to focus on reproductive health. On the public health side there has been little appreciation of the need to incorporate direct health measures into survey methodology. And on the economic side there is too great a focus on household surveys and too little on individuals, even though health is a characteristic of individuals, not households. Few economic surveys ask questions about health status, and few focus on the whole life cycle, looking at children, adults, and the elderly. This makes it difficult to examine interactions between these groups and the intergenerational transmission of health or to predict the effects of aging. Moreover, because children and the elderly are less likely to work than adults, the lack of attention to them deprives researchers of a tool for disentangling the links between income and health.

Two of the most important sources of data for household- or individual-level analysis in developing countries, the Demographic and Health Surveys and the Living Standards Measurement Study surveys, were not designed to analyze the interaction between income and health. The Demographic and Health Surveys contain no information on household income and focus primarily on reproductive

health. The Living Standards Measurement Study surveys generally contain detailed information on consumption and income but very limited (if any) information on health status. Integrated household surveys that collect data for individuals on economics and health can be used to explore the welfare consequences and interactions of different deprivations. Economists, in collaboration with physicians and other social scientists, can do much to find out how poverty and ill health, separately and in interaction, determine well-being.

The South African Integrated Family (Langeberg) Survey

A team of researchers in the United States and South Africa developed an instrument for linking information on individual and household economic well-being with the health and mental health status of household members, with an eye toward identifying causal links in the income-health gradient. The instrument includes questions on household resources, control over those resources within households, physical health of adults and children, mental health of adults, access to medical services, and a broad set of questions on other aspects of well-being, including social integration, exposure to violence, and the death of family members and friends. The survey was used in 1999 to collect data from a racially stratified random sample of 300 households (1,300 individuals) in the Langeberg health district in the Western Cape, an area that contains a mix of Black, White, and Coloured communities.[1] (See Case and Wilson [2001] for additional details on the survey.)

Measured along most economic and health dimensions, Blacks are less well off than are Coloureds, and Coloureds are less well off than are Whites (table 1). The mean monthly income of adult respondents was 511 rand (R) for Blacks, R936 for Coloureds, and R2,968 for Whites—or roughly a sixfold difference in monthly incomes for Blacks and Whites. The differences are even more stark when calculated as per person income figures (as shown in the lower panel in table 1). Blacks on average live in larger households (4.4 members) than Whites (2.8 members), with larger numbers of children (1.7 for Blacks and 0.8 for Whites).

Every adult in the survey was asked: How would you describe your health at present? Would you say it is excellent, good, average, poor, or very poor? Answers were scored from 1 (excellent) to 5 (very poor). Poor self-reported health has been shown to be a powerful predictor of mortality, even when controls were used for current health status and behaviors. Recent work has also found that self-ratings of poor health are a significant predictor of changes in functioning among the elderly. (See Idler and Kasl [1995] for results on changes in functioning and for extensive references on the studies of self-reported health and mortality.)

Blacks reported poorer health on average: 19 percent of Black adults reported themselves to be in poor or very poor health, true of only 8 percent of Coloureds and Whites. This is true even though Whites in the Langeberg health district (a popular retirement area for Whites) are older on average than are Blacks or Coloureds: the mean age was 49.5 years for White adult respondents, 37.6 years for Black adults, and 38.7 years for Coloured adults.

Table 1. Summary Statistics for the Langeberg Survey 1999

Means	Black	Coloured	White
Individual data, ages 18 and older			
Respondent's income (rand per month)	511	936	2,968
Indicator: respondent works for money	0.397	0.608	0.529
Self-reported health status (1 = excellent, 5 = very poor)	2.83	2.31	2.23
Indicator: asthma	0.087	0.085	0.048
Indicator: tuberculosis	0.091	0.099	0.034
Indicator: cancer	0.008	0.005	0.047
Indicator: heart trouble	0.072	0.064	0.146
Indicator: stroke	0.021	0.029	0.058
Indicator: high cholesterol	0.162	0.160	0.220
Indicator: diabetes	0.035	0.075	0.071
Indicator: emphysema	0.041	0.064	0.067
Indicator: both chronic and infectious disease	0.039	0.035	0.008
Age	37.6	38.7	49.5
Indicator: female	0.508	0.528	0.539
Number of observations	224	336	128
Household-level data			
Total household income per member	326.8	572.8	2,141.8
Indicator: no income pooling	0.127	0.174	0.077
Indicator: woman has a say[a]	0.679	0.756	0.717
Household size	4.44	4.85	2.84
Number members ages 0–17	1.73	1.89	0.78
Number members ages 18–54	2.20	2.51	1.23
Number members ages 55+	0.38	0.40	0.83
Number of observations	100	125	63

Note: The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era. Sample means are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race) and the clustering of observations (by enumerator area).
a. Has a value of 1 if a woman is reported to have a say in the way the household spends its resources.
Source: The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru].

In the Langeberg district most adults reported having access to some branch of the health care system and having been examined by a health care professional.[2] Blacks and Coloureds were more likely to report that a doctor or nurse or health care professional had told them that they have asthma or tuberculosis (see table 1). Roughly 4 percent of Black and Coloured adults reported both tuberculosis and a chronic disease (cancer, heart trouble, stroke, diabetes, or emphysema), true for less than 1 percent of Whites.

With respect to household decisionmaking, 13 percent of Black households report that members do not pool their incomes. Roughly two-thirds of all households report that a woman is among those who have the most say in decisions about spending in the household. Pooling and the position of women in the decisionmaking hierarchy have been shown to be important in determining household resource allocation. (See Alderman and others [1995] and references therein.)

Evidence on Health Status and Income from the Langeberg Survey

For all three races there is a negative and significant relationship between a respondent's income from all sources and health status (table 2), measured using the five-point scale introduced above. For all three races a doubling of income is associated

with a two-tenths of one point betterment of health status (lower numbers are associated with better health). That the gradient is strong for all three races, whose levels of income vary markedly, is consistent with results presented by Adler and others (1994, p. 15), who provide evidence that "the association of SES [socioeconomic status] and health occurs at every level of the SES hierarchy, not simply below the threshold of poverty."

The health of Blacks deteriorates more rapidly with each year of age than does the health of Coloureds. Blacks on average reported a 0.03-point worsening of health with each year of age, while Coloureds reported a 0.02-point worsening. Whites "age" more slowly: one year of age worsens reported health by 0.015 point—just half that reported for Blacks. For Blacks a doubling of income (a move, for example, from the median to the 67th percentile of the income distribution) is comparable to turning back the age clock by seven years: both are associated with an improvement in self-reported health status of 0.2 point.

The gradient in adult health is not eliminated when controls are added for education. In fact, educational attainment is not strongly correlated with health status for Blacks or Coloureds. For Blacks there is a small and significant relationship between education and health status in some specifications. For Coloureds there is no significant correlation in any specification. For Whites education and income both appear to be significantly correlated with health status, and the inclusion of education reduces the impact of income on health status by one-third (from 0.3 to 0.2 point). The differential effects of education across races are difficult to interpret here: the quality of South African schools varied markedly by race in the apartheid years (see Case and Deaton [1999]). As school quality improves for Coloured and Black children, education may show a stronger impact on health status.

Table 2. Income and Health Status

Controls	South Africa						United States	
	Black		Coloured		White		Black	White
Log own income	−0.175	−0.163	−0.147	−0.157	−0.324	−0.205	−0.193	−0.200
	(0.068)	(0.065)	(0.037)	(0.036)	(0.072)	(0.061)	(0.004)	(0.002)
Age	0.029	0.025	0.021	0.022	0.015	0.013	0.019	0.017
	(0.005)	(0.004)	(0.002)	(0.002)	(0.006)	(0.006)	(0.0002)	(0.00008)
Education		−0.036		0.010		−0.095	−0.052	−0.072
		(0.018)		(0.012)		(0.035)	(0.001)	(0.0005)
Number of observations	122	122	250	250	86	86	83,427	544,256

Note: The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era. The dependent variable is health status (1 = excellent to 5 = very poor). Numbers in parentheses are standard errors. South African regressions are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race) and the clustering of observations (by enumerator area). Income for the National Health Interview Survey is total household income. Both the U.S. and South African samples are restricted to adults ages 18 and older.
Source: Columns 1–6, The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru]; columns 7-8, the U.S. National Health Interview Survey 1986–95.

The gradient in health status in South Africa is identical to that in the United States. The last two columns of table 2 present results for Blacks and Whites in the United States using 10 years of data from the National Health Interview Survey.[3] For both Whites and Blacks a doubling of income is associated with an improvement of 0.2 point in self-reported health status. As is true in South Africa, the health status of Blacks in the United States deteriorates faster with age than does that of Whites, and education is more protective of health for Whites. (The precision of the U.S. estimates, which allows differentiation between the Black and White coefficients, comes from the sample size: there are half a million White observations in the U.S. estimate.)

The fact that the gradient in health status is identical between the United States and South Africa would seem to rule out some factors as an explanation of the relationship. Access to and quality of health services differ markedly between the countries, as do the quality and average levels of education. In what follows, this article examines for the South African results whether the correlation is due to the impact of health on earnings, the impact of money on health, or potentially some third factor that affects both health and income. Some of the mechanisms can be ruled in, and ruled out, using the data collected in the Langeberg Survey.

The Impact of Health on Income

Part of the correlation between income and health in South Africa is due to the reduced earnings potential of South Africans who have been chronically ill. Chronic illnesses have a large and significant effect on reported health status, as illustrated in column 1 of table 3, which presents results of a regression of respondents' self-reported health status on responses about whether a health professional has informed them that they have specific chronic conditions. This regression, which also includes controls for age, gender, and race, shows that asthma, tuberculosis, and cancer have large and significant effects on health status, with each worsening the reported status by more than half a point. Conditions associated with obesity (high cholesterol and diabetes) each worsen health status by three-tenths of a point on average.

That these chronic illnesses are related to labor force participation is seen in column 2 of table 3, which presents results on reasons for retirement. The Langeberg Survey asked each older adult (ages 55 and above) whether they reduced the hours or changed the type of work they were doing as they got older. If the response was positive, an open-ended follow-up question asked why the change had occurred. Poor health was the modal response; being old, or old enough to receive a pension, was the second most frequent response. The dependent variable in column 2 is defined only for respondents who report retirement. It is an indicator variable that the retirement occurred because of poor health. For the 74 older adults who reported retirement, retirement due to poor health is significantly correlated with the same chronic conditions that led to the largest reduction in health status—asthma, tuberculosis, and cancer. The health conditions are jointly significant deter-

Table 3. Health Conditions and Employment

Controls	All adults: Dependent variable is self-reported health status	Retirees only: Dependent variable = 1 if retirement was due to poor health
Asthma	0.582 (0.129)	0.212 (0.127)
Tuberculosis	0.538 (0.186)	0.306 (0.227)
Cancer	0.811 (0.219)	0.780 (0.213)
Heart trouble	0.177 (0.120)	−0.106 (0.097)
Stroke	0.076 (0.219)	−0.242 (0.104)
High cholesterol	0.342 (0.118)	0.019 (0.111)
Diabetes	0.288 (0.124)	−0.153 (0.120)
Emphysema	0.162 (0.162)	−0.148 (0.121)
Black[a]	0.782 (0.142)	0.275 (0.188)
Coloured[a]	0.249 (0.119)	0.427 (0.146)
Age	0.014 (0.003)	−0.003 (0.004)
Female	0.100 (0.051)	−0.212 (0.082)
Number of observations	632	74

Note: Numbers in parentheses are standard errors. Regressions also include a constant term. Both regressions are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race), and the clustering of observations (by enumerator area).
a. The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era.
Source: The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru].

minants of reporting that retirement was due to poor health (F-test = 10.56, p-value = 0.0000). This is prima facie evidence of a channel from poorer health to lower income, working through the effect of chronic disease on labor force participation.

The Impact of Income on Health

That a channel exists from health to income does not imply the absence of a channel from income to health. However, it does suggest caution in separating the effects. To investigate whether income has a causal effect on health requires identifying a source of income that is not itself determined by a respondent's health status. For South Africa this challenge is met by the state old age pension. Women ages 60 and older and men ages 65 and older are eligible for a monthly cash transfer if

they do not have an employer-based pension. Take-up rates for the state old age pension among Blacks and Coloureds are on the order of 80 percent (Case and Deaton 1998). White participation is only about 10 percent, because access to private pensions precludes take-up of the state pension for most Whites. Thus analysis here is restricted to Coloured and Black respondents.

In many communities in South Africa, where unemployment is as high as 40 percent, the state pension is the only stable source of income. It is also a large sum of money. At the time of the Langeberg Survey the old age pension was R520 per month, which is equal to the median income for adult Coloured respondents and is more than twice the median for Blacks (R220). The presence in the household of an age-eligible member is used as the marker for the pension, rather than receipt of the pension, to avoid modeling the timing of pension take-up.

Pension income should benefit prime-age adults and children living in a pensioner household, but only if incomes are pooled. In the Langeberg Survey 16 percent of Black and Coloured households (20 percent of adult respondents) reported that income was not pooled in their households. Pooling is not correlated with total household income, household size, the presence of pensioners in the household, or the gender of the pensioner. However, pooling is significantly more likely the smaller the age difference between the oldest and youngest adults in the household and the fewer the number of generations. Pooling is significantly less likely if a young adult (male or female) in the household reported working for money. This finding is consistent with pilot surveys in South Africa between 1996 and 1998, which found that younger working household members were often reluctant to pool resources with older members, choosing instead to make their own way (even while living under the same roof with their kin).

Thus information on both pension receipt and income pooling is used to identify a causal effect of income on health status. Figure 1 plots the conditional expectation of health status, given age, for Black and Coloured adults living in households that pool income. The conditional expectations are calculated using a Fan (1992) locally weighted regression smoother, which allows the data to determine the shape of the function, rather than imposing (for example) a linear or quadratic form. Health status by age is presented separately for respondents who are currently living with a pensioner and for those who are not. For both groups self-reported health status generally worsens with age. (The improvement observed in self-reported health status between ages 20 and 35 for respondents living with pensioners is due in part to the higher probability of there being multiple pensioners in households with respondents 35 years old than in those with respondents 25 years old. This is explored in table 4 below.)

Pension income protects the health of all adult members in households that pool income (see figure 1). For respondents in their 20s and continuing through adulthood, a gap in self-reported health status develops in favor of respondents living with a pensioner. For pensioners in households that pool income, health continues to worsen with age at roughly the same rate it does for members between the ages of 40 and 60. This does not imply that the pension does not improve the health of pensioners. If a woman were the first person in her household to become eligible for

Figure 1. Self-Reported Health Status and Age for Respondents in Households That Pool Income, With and Without Pensioners, Black and Coloured Respondents

Self-reported health status

Age
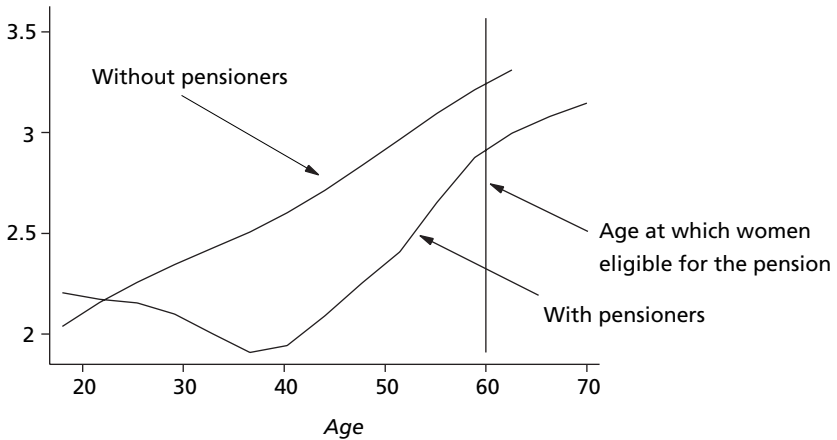
the pension, she would be expected to fall on the upper "without pensioners" curve at age 59. At age 60, when she receives her pension, her health would be expected to improve, on average, and she would be expected to move to the lower "with pensioners" curve. Not only would her health be expected to improve, but so would that of other members of her household.

The results are sharply different for nonpooling households (figure 2). Again, there is a pronounced worsening of health with age. But there is much less of a difference in the health status of respondents, at any given age, with and without pensioners in their households. In contrast to the results for pooling households, the self-reported health status of respondents not living with a pensioner is slightly better. Once a respondent in a nonpooling household reaches pension age, health status stops declining altogether and begins to improve. Results presented in figures 1 and 2, then, suggest that the pension protects the health of all members when the household pools income and protects the health of pensioners only when the household does not pool.

Would the health of pensioners in pooling households be better on average if they lived in nonpooling households? Comparisons of health status for all respondents living with a pensioner in households that pool income and those that do not again show the protective effect of pension income on the health of all members of a pooling household (figure 3). At age 60 women coming into their pensions from a pooling household start life as a pensioner with better health than do those coming into their pensions from a nonpooling household. Not until age 70 does the ameliorating effect of the pension for pensioners in nonpooling households bring their health status into line with that observed for pensioners from pooling households.

Table 4. Pension Income, Income Pooling, and Health Status, Black and Coloured Adult Respondents

Controls	Household income pooled			Household income not pooled			All households	
Someone in household is eligible for a state pension	−.395 (.145)	−.389 (.186)		.261 (.161)	.351 (.132)		−.264 (.086)	
One person in household is eligible for a state pension			−.320 (.145)			.339 (.134)		−.146 (.080)
Two or more people in household are eligible for a state pension			−.725 (.214)			1.418 (.275)		−.534 (.227)
Respondent is eligible for a state pension		−.015 (.300)	.102 (.303)		−.618 (.419)	−.710 (.433)		−.113 (.305)
Respondent is female	.358 (.104)	.358 (.110)	.347 (.116)	.066 (.196)	.069 (.203)	.066 (.204)	.315 (.089)	.314 (.098)
A woman has the most say in household spending	−.076 (.110)	−.076 (.112)	−.078 (.115)	−.269 (.207)	−.206 (.199)	−.215 (.202)	−.101 (.111)	−.098 (.115)
"Female" interacted with "A woman has the most say in spending"	−.218 (.113)	−.219 (.117)	−.208 (.128)	.169 (.367)	.168 (.362)	.181 (.363)	−.175 (.075)	−.169 (.078)
Household does not pool income							.051 (.087)	.025 (.089)
Number of observations	416	416	416	88	88	88	504	504

Note: The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era. The dependent variable is health status (1 = excellent,…5 = very poor). Numbers in parentheses are standard errors. All regressions are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race) and the clustering of observations (by enumerator area). Included in all regressions but not shown in the table are respondent's age, an indicator that respondent is Coloured, and age interacted with an indicator that respondent is Coloured.
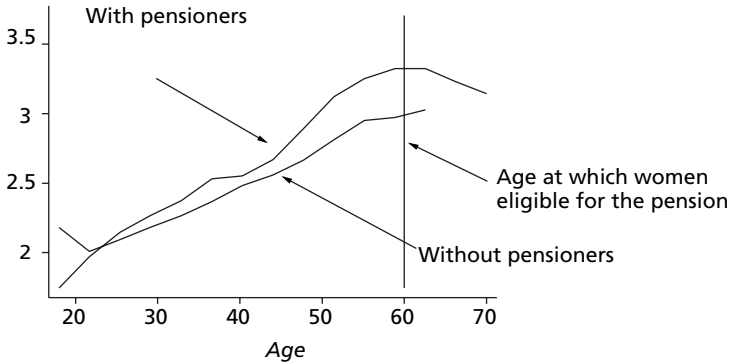*Source:* The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru].

Figures 1 through 3 provide evidence that pension income is protective of health status. For pensioners living in nonpooling households, the beneficial effects of the pension accrue only to the pensioners and only on receipt of the pension. However, for pensioners living in income-pooling households, the beneficial effects of the pension accrue to all members.

The size of this effect, after controlling for other characteristics of the household and its members, is presented in table 4 for Black and Coloured respondents in households that pool income (the first set of columns) and those that do not (the second set). In households that pool resources, the presence of a pensioner improves

Figure 2. Self-Reported Health Status and Age for Respondents in Households That Do Not Pool Income, With and Without Pensioners, Black and Coloured Respondents
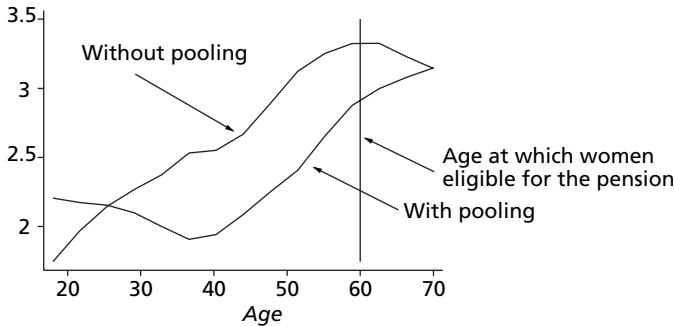
Self-reported health status

Source: The Langeberg Survey 1999.

Figure 3. Self-Reported Health Status in Pensioner Households With and Without Income Pooling, Black and Coloured Adult Respondents

Self-reported health status

Source: The Langeberg Survey 1999.

self-reported health status on average by 0.4 point (column 1). No additional protection of the pension income accrues to the pensioner: an indicator that the respondent is a pensioner has a small coefficient (0.015) that is insignificantly different from zero (column 2). In contrast, being a pensioner in a nonpooling household has a large effect on self-reported health status (column 5). This difference in outcomes for pensioners in households that pool and those that do not appears in many of the results presented below.

When pensioner households are split into those with one pensioner and those with two or more pensioners (each pensioner receives R520 per month), the coefficient on having two or more pensioners in a pooling household is more than twice as large as that observed for one-pensioner households (last column in each panel of table 4). Again, being the pensioner provides no extra protection in households that pool. In households that do not pool, pensioners' health is 0.7 point better than that of other member.

There may be unobservable differences between pooling and nonpooling households that are correlated with health status and with the presence of pensioners. If so, these may bias the estimated impact of pension income on health status. To examine whether this is driving the findings, results are reported for all Black and Coloured respondents from both pooling and nonpooling households (last two columns of table 4). A significant effect of pension income on outcomes remains for all members, an effect that is larger the greater the number of pensioners in the household. The results are attenuated relative to others reported in table 4 because households in which the pension helps all members (pooling households) are combined with those in which the pension protects only the elderly.

In pooling households women's health is significantly worse than men's health. However, women's health receives some protection if the woman is reported to have the "most say" in how household money is spent. For pooling households the variables indicating that a woman has the most say in spending and that indicator interacted with whether the respondent is a woman are jointly significant in all specifications.

Data on children's height (an indicator of long-run nutritional status) and weight (an indicator of short-run nutritional status) were used to test for the impact of pension income on children's health (table 5). Regressions were run separately for Black and Coloured children because of possible differences in growth patterns. The sample was restricted to children born after the old age pension system was fully implemented (January 1, 1994). A complete set of quarter-since-birth indicators was used to control for age.

For both Blacks and Coloureds the presence of a pensioner is positively and significantly correlated with children's height. The presence of one pensioner is associated with an additional 3 to 4 centimeters of height. On average Black and Coloured children were 8 centimeters taller with each year of age from ages zero to six, so the additional height associated with the presence of a pensioner is roughly that associated with an additional half year of age. For Blacks the presence of two pensioners has an effect on height more than twice that of one pensioner. (There were no young Coloured children in two-pensioner households.) The presence of a pensioner is also associated with increased weight (1 kilogram) for Coloured children, but is not a significant predictor of weight for Black children. That the pension is protective of children's height, and thus of their long-run nutritional status, may prove important in a country where many AIDS orphans will live with elderly relatives in the next decade.[4]

Table 5. Children's Heights and Weights, Black and Coloured Children

Controls	Height in centimeters				Weight in kilograms			
	Black		Coloured		Black		Coloured	
Someone in household is eligible for a state pension	8.07 (4.35)		4.54 (2.21)		−.055 (2.44)		1.10 (.316)	
One person in household is eligible for a state pension		3.21 (1.89)		4.54 (2.21)		.074 (4.35)		1.10 (.316)
Two or more people in household are eligible for a state pension		13.9 (6.20)				−.211 (.498)		
Child is female	−7.83 (2.87)	−7.12 (2.97)	−1.10 (1.34)	−1.10 (1.34)	2.67 (9.86)	2.65 (10.42)	−.584 (.316)	−.584 (.316)
Number of observations	37	37	44	44	37	37	44	44

Note: The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era. Sample is restricted to children born after 1 January 1994, when pension system was fully implemented. Numbers in parentheses are standard errors. All regressions include a complete set of quarter-of-birth indicators (quarter 1, 1994, through quarter 2, 1999). Regressions are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race) and the clustering of observations (by enumerator area).
Source: The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru].

Turning Money into Health

The results of the previous section provide evidence that pension income protects health status. What are some of the mechanisms by which money generates health? This section examines four potential channels: medical care, water and sanitation, nutrition, and psychosocial stress.

THROUGH MEDICAL CARE. One way in which money might generate health is through its effects on health care. Higher incomes may allow people to spend more time and money seeking out health services. A pensioner may be able to afford a consultation with a private doctor, for example, or a longer taxi ride to a better-equipped clinic.

Black and Coloured respondents were asked whether they had been to any of a number of health services for medical care in the previous three months—a day hospital or community clinic, mobile clinic, government hospital or outpatient clinic at a hospital, private hospital or clinic, private doctor, chemist shop, traditional healer, or community nurse who visits at home. Respondents' answers to each of these questions were regressed on an indicator of whether there was a pensioner in the household and whether the respondent was a pensioner, together with the respondent's age, an indicator that the respondent is Coloured, age interacted with the Coloured indicator, and indicators that the respondent is female and that the

respondent lives in an urban area. Neither having a pensioner in the household nor being a pensioner was a significant positive determinant of going to a clinic or doctor, whether government-run or private. (These results are available on request.)

Whether pensioners had an effect on choice of health care provider in pooling households was tested for by adding indicators that the household pools income and that the household has a pensioner and pools income. The coefficients for respondents living with pensioners in pooling households were never positive and significant determinants of health care use. (In fact, these respondents were significantly less likely to have been to a day clinic, a government hospital, or a private doctor.) The results for pensioners themselves were small and insignificantly different from zero. These results were robust to analyzing respondents by race and to restricting the sample to respondents who reported themselves to be in poor health.

There may be many reasons why no effects were found of pension income on health services. There were no controls for the quality of health care available, which may be uniformly poor. The three-month window may be too short to adequately capture health care use. Nonetheless, if income is protecting health status through the health care system, more work will be needed to find that out. The results of this analysis provide no evidence that pension income has a significant effect on doctor and clinic visits or that the protective effect of income on the health status of people living with pensioners comes through health services.

THROUGH WATER AND SANITATION. Another way in which money may influence health is through better sanitation. The pension may be used to upgrade household facilities, and some of the improvements may have health consequences. The presence of a pensioner in the household is positively and significantly correlated with a flush toilet in the dwelling and positively correlated with an indicator that the household has an on-site source of water (Case 2001). Roughly 40 percent of the Black and Coloured households in the sample have a flush toilet; 90 percent have water on site. It may take time to obtain a flush toilet or water on site. These water-related variables are both significantly correlated with the number of years the pensioner should have been receiving a pension (based on age), together with an indicator that a pensioner is present in the household.

THROUGH NUTRITION. A third way in which money may influence health is through its effect on nutrition. Table 6 presents regression results on whether a knowledgeable household member reported that an adult in the household had skipped a meal or had the size of a meal reduced in the last year because of insufficient money to buy food. Roughly 45 percent of all Black and Coloured households reported an adult skipping a meal. The presence of one pensioner in the household on average reduced the probability of an adult skipping a meal by 20 percent. In pooling households the presence of two pensioners reduced the probability by 40 percent.

Skipping meals is associated with poorer health. In a pooling household in which an adult was reported to have skipped a meal, the health status of respondents was

Table 6. Pension Income, Income Pooling, and Hunger

	Household income pooled		Household income not pooled		All Households	
Indicator: household has a state pensioner	−.234 (.073)		−.164 (.151)		−.224 (.052)	
One person in household is eligible for a state pension		−.166 (.079)		−.195 (.143)		−.183 (.045)
Two or more people in household are eligible for a state pension		−.397 (.079)		.491 (.239)		−.354 (.087)
F-test: Joint significance of the number of pensioner variables		15.28 (.0000)		4.05 (.0428)		14.82 (.0001)
Household size	.024 (.013)	.024 (.013)	.030 (.042)	.026 (.040)	.023 (.014)	.023 (.014)
Household does not pool income					−.010 (.077)	−.018 (.076)
Number of observations	187	187	33	33	220	220

Note: The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era. Dependent variable is equal to 1 if a knowledgeable household member answered that "in the last 12 months [an adult] in the household ever cut the size of a meal or skipped meals because there wasn't enough money for food." This variable is a household-level variable, and the regressions include one observation per household. Numbers in parentheses are standard errors. All regressions are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race) and the clustering of observations (by enumerator area).
Source: The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru].

0.14 point worse (Case 2001). In nonpooling households in which an adult was reported to have skipped a meal, there was no significant effect on the health status of respondents. Perhaps in pooling households when one adult misses a meal, it is more likely that they all do, leading to the significant effect for pooling households. (Alternatively, in nonpooling households, perhaps it is less likely that the "knowledgeable" household member knows whether meals were skipped, leading to noise in this variable.) Adding the information on meal skipping reduces the pensioner coefficients in pooling households by roughly 5 percent (Case 2001). With enough information on how pension income is spent, the effect of pension income on health may be attributable to its component parts.

THROUGH MENTAL HEALTH. Finally, money may influence health through its effect on psychosocial stress. Some of the biological pathways through which stress erodes health are now well understood (Sapolsky 1994), and the field is moving apace. The ways in which income may be useful in reducing stress are infinite. When a woman has enough money to feed her children, or a household head knows where the money to pay school fees will come from, stress is reduced.

In the Langeberg Survey all adults were asked a battery of questions about depression, which is inextricably linked to stress. (As noted by Sapolsky [1994, p. 197], "it is impossible to understand either the biology or the psychology of major depressions without recognizing the critical role played in the disease by stress.") Respondents in the Langeberg Survey were asked how often in the past week they were miserable, depressed, sad, cried a lot, did not feel like eating, felt everything was an effort, slept restlessly, could not get going. For each behavior a response of "almost all the time" was coded as 1 and other answers as 0.

The responses for these behaviors were summed to create a depression index. (The mean index was 0.54, with 10 percent of the sample reporting two or more depression-related behaviors.) To test whether depression was correlated with income, the depression index was regressed on the presence of pensioners in the household (table 7). For households pooling income the presence of pensioners has a significant effect on reported depression. The greater was the number of pensioners, the greater this effect. (The results in table 7 are very similar if ordered probits are estimated in place of ordinary least squares regressions.) The presence of pensioners in nonpooling households is insignificant. Though the coefficient is larger, so is the standard error. When pooling and nonpooling households are combined, the presence of pensioners again has a large and significant effect on the depression index.

An alternative explanation for the benefits associated with the presence of a pensioner is simply that having older people in the household results in less depression for all household members. To test for this, controls were included in these regressions for household size, number of children, and number of members ages 55 and older. For pensioner households there was no significant effect of having older members unless those members were pension eligible.

Conclusion

The Langeberg Survey provides evidence that income has a causal effect on health status, an effect that works at least in part through sanitation and living standards, in part through nutritional status, and in part through the reduction of psychosocial stress.

The results demonstrate that information collected at the household level is not adequate to assess the causal effects of income on the well-being of individual household members. In the case of South Africa (and as seems likely to be true elsewhere), such data need to be augmented with additional information on household sharing rules, if accurate readings are to be taken.

The results also demonstrate that governments interested in improving health status may find cash benefits to be one of the most effective policy tools in their kits. Children dying of infectious diseases brought on by HIV/AIDS are often dying of the very same diseases that are killing children whose immune systems have been weakened by malnutrition. While adequate protocols have not yet been developed for treating children with HIV/AIDS, we know how to treat malnutrition—very simply, by providing income to the children's primary caregivers.

Table 7. Pension Income, Income Pooling, and Depression, Black and Coloured Adult Respondents

	Household income pooled		Household income not pooled		All households	
Someone in household is eligible for a state pension	−.531 (.353)		−.054 (.364)		−.499 (.267)	
One person in household is eligible for a state pension		−.498 (.282)		−.066 (.367)		−.492 (.237)
Two or more people in household are eligible for a state pension		−1.01 (.568)		−1.44 (1.14)		−.932 (.533)
Respondent is eligible for a state pension	−.316 (.413)	−.165 (.458)	−.517 (.467)	−.385 (.505)	−.334 (.296)	−.198 (.356)
Number of members aged 55+	.150 (.159)	.203 (.186)	−.525 (.290)	−.494 (.297)	.063 (.178)	.107 (.206)
Respondent is female	.263 (.128)	.261 (.132)	.405 (.395)	.392 (.396)	.286 (.082)	.286 (.084)
Household does not pool income					.049 (.307)	.034 (.322)
Number of observations	432	432	96	96	528	528

Note: The analysis uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era. Numbers in parentheses are standard errors. The dependent variable is the depression index: sum of the responses to the question that "most of the time" the respondent was miserable, depressed, sad, cried a lot, did not feel like eating, felt everything was an effort, slept restlessly, could not get going. For each behavior a response of "almost all the time" was coded as 1 and other answers as 0. All regressions are weighted based on the 1996 South African census, taking into account the stratification of the sample (by race) and the clustering of observations (by enumerator area). Included in all regressions but not shown in the table are respondent's age, an indicator that respondent is Coloured, age interacted with an indicator that the respondent is Coloured, household size, and the number of members ages 0–17.
Source: The Langeberg Survey 1999 [www.uct.ac.za/depts/saldru].

Notes

1. This article uses the apartheid classification of "Black" and "Coloured," in part because the labels carry with them information about the long-term deprivation faced during the apartheid era.

2. Surveys run in the United States to test the reliability of self-reported chronic conditions show a high degree of reliability. See Brownson and others (1994), for example, for reports on hypertension and diabetes.

3. In the National Health Interview Survey, total household income is recorded by income category (for 27 categories). Incomes have been assigned to each category using data from the 1986–95 March Current Population Surveys. See Case, Lubotsky, and Paxson (2001) for details.

4. These findings are broadly consistent with those of Duflo (2000), who finds the South African pension protective of children'ís heights, using data collected in 1993 (before the pension was fully implemented).

References

Acheson, Donald (Chairman). 1998. *Independent Inquiry Into Inequalities in Health Report.* London: The Stationery Office.

Adler, Nancy E., and Joan M. Ostrove. 1999. "Socioeconomic Status and Health: What We Know and What We Don't." In Nancy E. Adler, Michael Marmot, Bruce S. McEwen, and Judith Stewart, eds., *Socioeconomic Status and Health in Industrial Nations*. New York: The New York Academy of Sciences.

Adler, Nancy E., Thomas Boyce, Margaret A. Chesney, Sheldon Cohen, Susan Folkman, Robert L. Kahn, and S. Leonard Syme. 1994. "Socioeconomic Status and Health, the Challenge of the Gradient." *American Psychologist* 49(1): 15–24.

Alderman, Harold, Pierre-Andre Chiappori, Lawrence Haddad, John Hoddinott, and Ravi Kanbur. 1995. "Unitary Versus Collective Models of the Household: Is it Time to Shift the Burden of Proof?" *World Bank Research Observer* 10(1): 1–19.

Brownson, Ross C., Jennetter Jackson-Thompson, Joan C. Wilkerson, and Fatemeh Kiani. 1994. "Reliability of Information on Chronic Disease Risk Factors Collected in the Missouri Behavioral Risk Factor Surveillance System." *Epidemiology* 5(5): 545-49.

Case, Anne. 2001. *Does Money Protect Health Status? Evidence from South African Pensions.* NBER Working Paper 8495. Cambridge, Mass.: National Bureau of Economic Research.

Case, Anne, and Angus Deaton. 1998. "Large Cash Transfers to the Elderly in South Africa." *Economic Journal* 108(450): 1330–61.

———.1999."School Inputs and Educational Outcomes in South Africa." *Quarterly Journal of Economics* 114(3): 1047–84.

Case, Anne, and Francis Wilson. 2001. "Health and Well-Being in South Africa: Evidence From the Langeberg Survey." Princeton University, Research Program in Development Studies, Princeton, N.J.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2001. *Economic Status and Health in Childhood: The Origins of the Gradient.* NBER Working Paper 8344. Cambridge, Mass.: National Bureau of Economic Research.

Duflo, Esther. 2000. *Grandmothers and Granddaughters: Old Age Pension and Intra-Household Allocation in South Africa.* NBER Working Paper 8061. Cambridge, Mass.: National Bureau of Economic Research.

Easterlin, Richard A. 1998. "How Beneficent Is the Market? A Look at the Modern History of Mortality." University of Southern California, Department of Economics, Los Angeles.

———. 1999. "How Beneficent Is the Market? A Look at the Modern History of Mortality." *European Review of Economic History* 3(3): 257–94.

Fan, Jianqing. 1992. "A Design-Adaptive Nonparametric Regression." *Journal of the American Statistical Association* 87(420): 998–1004.

Filmer, Deon, Jeffrey S. Hammer, and Lant H. Pritchett. 2000. "Weak Links in the Chain: A Diagonosis of Health Policy in Poor Countries." *World Bank Research Observer* 15(2): 199–224.

Fogel, Robert W. 1994. "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy," *American Economic Review* 84(3): 369–95.

Gwatkin, Davidson R. 2000. "Health Inequalities and the Health of the Poor: What Do We Know? What Can We Do?" *Bulletin of the World Health Organization* 78(1): 3–18.

Idler, Ellen L., and Stanislav V. Kasl. 1995. "Self-Ratings of Health: Do They Also Predict Change in Functional Ability?" *Journal of Gerontology: Social Sciences* 508(6): S344–53.

Kahn, Kathleen, Stephen M. Tollman, Michel Garenne, and John S. S. Gear. 1999. "Who Dies From What? Determining Cause of Death in South Africa's Rural North-East." *Tropical Medicine and International Health* 4(6): 433–41.

Marmot, Michael. 1999. "Epidemiology of Socioeconomic Status and Health: Are Determinants Within Countries the Same as Between Countries?" In Nancy E. Adler, Michael Marmot, Bruce S. McEwen, and Judith Stewart, eds., *Socioeconomic Status and Health in Industrial Nations.* New York: The New York Academy of Sciences.

McKeown, Thomas. 1976. *The Modern Rise of Population.* New York: Academic Press.

Preston, Samuel H. 1975. "The Changing Relation Between Mortality and Level of Economic Development." *Population Studies* 29(2): 231–48.

———. 1980. "Causes and Consequences of Mortality Declines in Less Developed Countries in the Twentieth Century." In R. A. Easterlin, ed., *Population and Economic Change in Developing Countries.* Chicago: University of Chicago Press.

Sapolsky, Robert M. 1994. *Why Zebras Don't Get Ulcers: A Guide to Stress, Stress-Related Diseases, and Coping.* New York: W. H. Freeman and Company.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relationship Between Health and Economic Status." *Journal of Economic Perspectives* 13(2): 145–66.

Wagstaff, Adam. 2000. "Socioeconomic Inequalities in Child Mortality: Comparisons Across Nine Developing Countries." *Bulletin of the World Health Organization* 78(1): 19–29.

World Health Organization. 2000. *World Health Report 2000: Health Systems—Improving Performance.* Geneva.

Comment on "Health, Income, and Economic Development" by Anne Case

Sudhir Anand

I have only a few remarks on Anne Case's article, which I found interesting and on the whole convincing. They relate to the subjective measure she uses for ill health.

Adults in the survey were asked to place themselves in one of five categories of health: excellent, good, average, poor, or very poor. I have some very recent experience with a similar question. Just three days ago (on 30 April 2001) I was completing the England and Wales decennial census questionnaire, which asked about one's self-assessed health, categorizing it as good, fairly good, or poor. I immediately placed myself in the middle or "fairly good" category of health.

But after a long discussion on the subject with my family, they convinced me that I actually belonged to the "good" category and not to the "fairly good" category. (They thought I was displaying a streak of hypochondria.) The natural tendency when there is some doubt about exactly what a question means is to place oneself in the middle category. (The summary statistics of average health in Case's article are not inconsistent with this possibility.) At any rate, I would expect some noise in self-reported health status measures.

There are other measures of individual health available in the literature that are not based on subjective assessments of health. They are objective in that they depend on the measured loss of functioning associated with morbidity or ill health. An example is the World Bank/World Health Organization measure of disability used in the construction of disability-adjusted life years (DALYs) in *World Development Report 1993: Investing in Health*. Other methods of measuring (ill) health may be found in McDowell and Newell (1996).

A second point concerns the particular cardinalization that Case adopts for the ordinal rankings of the self-reported health of individuals in the Langeberg Survey. She uses a linear scale of 1, 2, 3, 4, 5, with equal distance between the categories. It would be reassuring to see that her results are robust to alternative cardinalizations

of the rank ordering of individuals' health status. Perhaps a quadratic scale might be tried—for example, 1, 4, 9, 16, 25—or some other monotonic transformations of the linear scale.

Reference

McDowell, Ian, and Claire Newell. 1996. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd edition. Oxford: Oxford University Press.

This Page Intentionally Left Blank

World Inequality and the Rise in Longevity

Tomas Philipson and Rodrigo R. Soares

The Human Development Index of the United Nations Development Programme was devised to capture aspects of human development not reflected in income alone, such as, for example, health and education. A full income measure of human development similarly may take into account such nonmonetary aspects of development but nevertheless assesses their value in monetary terms. This article discusses the conceptual and empirical differences between these two measures of development, focusing on the implications of the increase in longevity observed worldwide in the last few decades for their effectiveness in measuring inequality.
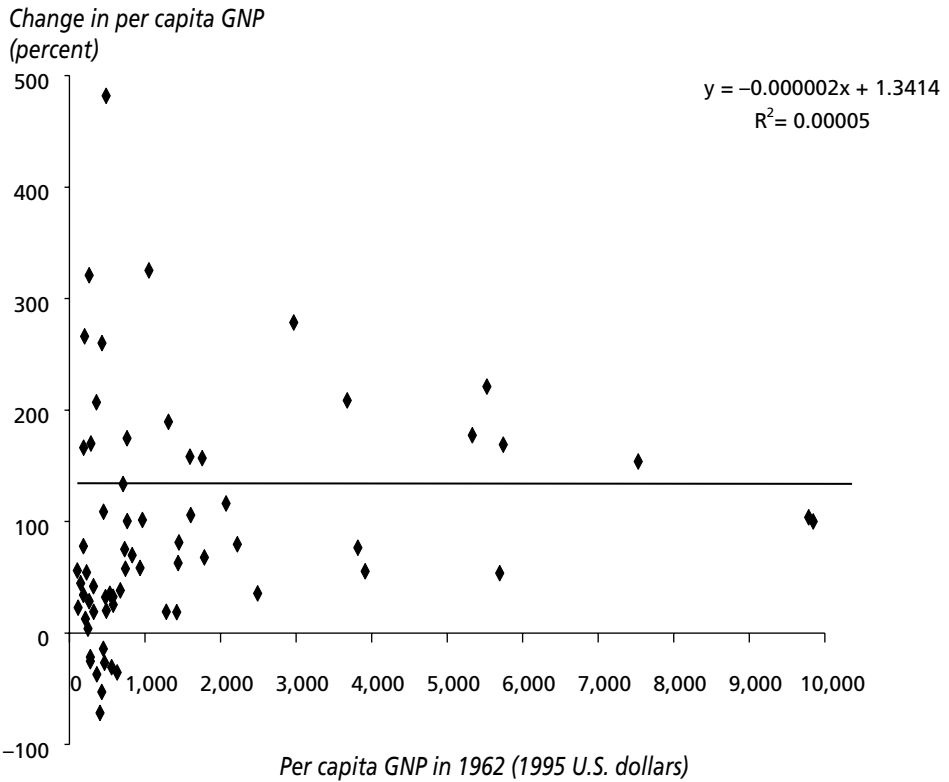
The lack of income convergence among countries has led many people to worry about the impact that globalization of markets will have on world inequality, as some developed countries get richer while developing countries fall behind. Many Organisation for Economic Co-operation and Development (OECD) countries have experienced significant growth in per capita income in the past few decades, for example, while many African countries have not. Growing cross-country inequality, and even the absence of long-run economic growth for some poor countries, raise the question of whether development will inevitably reach all societies or remain restricted to a relatively small group of rich countries.

The lack of income convergence across countries has been extensively documented (see Barro and Sala-i-Martin [1995] and the references therein). Figure 1 replicates the well-known finding that there is no absolute income convergence across countries. It plots the growth rate of per capita GNP in 1962–97 against the initial level of GNP for a cross-section of 83 economies (see appendix for definitions of variables, the data set, and the sample economies). The figure shows that there is no significant covariance between initial income levels and future growth rates.

Figure 1. Regression to the Mean for Per Capita Income, 1962–97

Change in per capita GNP (percent)

y = −0.000002x + 1.3414
R^2 = 0.00005

Per capita GNP in 1962 (1995 U.S. dollars)

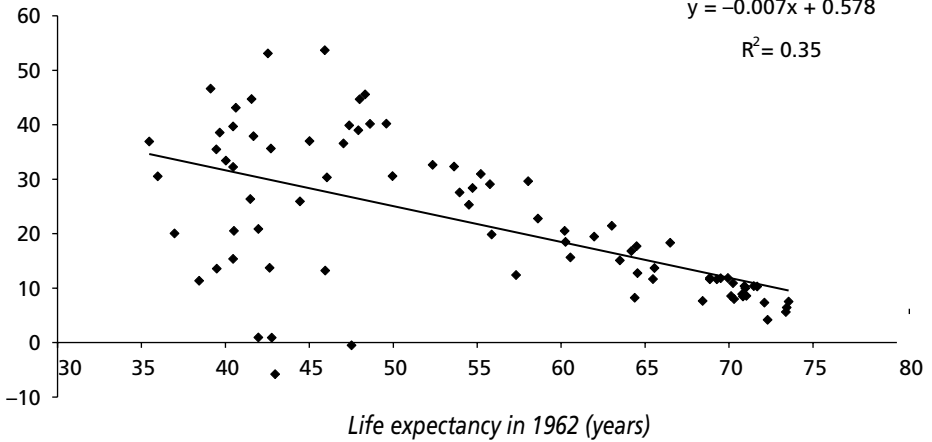*Source:* Authors' calculations based on data from the World Bank (1999).

Overall, poor countries do not tend to grow faster than rich countries and thus seem not be catching up in income.

Of course, material gain is only one of many aspects of life that enhance economic welfare. Another important dimension is how long one lives to enjoy the goods and services available for consumption—the length of the "flow" of well-being represented by yearly income. Although there was a lack of convergence in income in 1962–97, the same was not true for life expectancy. Countries starting at low levels tended to gain more in life expectancy than those starting at high levels (figure 2). The regression line in figure 2 suggests that 10 more years of life expectancy in 1962 meant growth rates for longevity that were 7 percentage points lower in the next 35 years.

Because countries in which life expectancy is shorter are also relatively poorer countries, that means that poor countries are catching up to rich ones in the dimension of longevity. In other words, changes in income have not always been accompanied by similar changes in all the relevant dimensions of development. Thus looking only at income to draw conclusions about changes in overall economic welfare across nations may be misleading. Perhaps due to the globalization of health

Figure 2. Regression to the Mean for Life Expectancy, 1962–97

Change in life expectancy
(percent)

$y = -0.007x + 0.578$

$R^2 = 0.35$

Life expectancy in 1962 (years)

care or medical knowledge, the poor countries may be catching up to the rich and world inequality may be falling.

Measuring the value of the observed gains in longevity in income units allows one to more readily address the change in overall economic welfare across countries implied by these growth rates in income and longevity. Becker, Philipson, and Soares (2001) conducted such an analysis on the international level, extending the national estimates provided by, for example, Cutler and Richardson (1997), Nordhaus (1999), and Murphy and Topel (1999).

Here we attempt to compare such an economic measure of cross-country inequality with a well-known alternative measure of well-being: the Human Development Index (HDI) of the United Nations Development Programme (UNDP). This index incorporates two other dimensions of development besides income—education and longevity—as measures of human welfare.[1] But the HDI has some limitations, stemming mainly from the arbitrary way in which it is constructed. In contrast to the HDI, an economic approach to evaluating welfare differences across countries would attempt to measure overall welfare in monetary units, based on the value the population attaches to the gains in the different dimensions of development. If the dimensions of human development measured by the HDI are indeed important, individuals would be willing to trade income for them, allowing one to measure their value in income-equivalent units.

In this article we examine the qualitative differences between the HDI and an index based on income-equivalent compensation. We consider a general framework in which such an economic index can be computed to incorporate different aspects of human development. Then we consider the quantitative differences between the HDI and this

economic index, by comparing the results obtained from the two indexes for a sample of economies in the 35 years between 1962 and 1997. Our main finding is that the HDI performs well in generating a ranking of countries because the different characteristics that it incorporates—income, health, and education—are highly correlated at a point in time. Nevertheless, changes in the relative position of the countries are not fully captured, since changes in health and changes on income have been uncorrelated.

An Analytical Examination of Measures of Development

Any index can be generally approximated as a linear combination of its components. Thus when the components are highly correlated, the rankings obtained from different indexes will be very similar. Nevertheless, the arbitrariness of an index like the HDI limits its applicability as an instrument for cost-benefit analysis of development programs. Indexes based on the economic concept of income-equivalent compensation do not suffer from this limitation.

The Human Development Index: Goals and Methodology

The Human Development Index, constructed annually since 1990 by the UNDP, summarizes "achievements in basic human development in one simple composite index and produces a ranking of countries" (UNDP 2000, p. 147). Because income gives an incomplete picture of living conditions, the HDI tries to provide a more comprehensive measure of well-being by including education and life expectancy, along with income, as dimensions of human development. The main goal of the HDI is described in the UNDP's *Human Development Report 2000* (p. 148) as follows:

> The HDI value for a country shows the distance that it has to travel to reach the maximum possible value of 1—or its shortfall—and also allows intercountry comparisons. A challenge for every country is to find ways to reduce its shortfall.

The maximum value of 1 results from the fact that the HDI is a relative index. It is a simple arithmetic average of a life expectancy index, an education index, and a per capita income index. These three indexes are constructed, for each country i, according to the following formulas:

Life expectancy index$_i$ = (life exp$_i$ – life exp$_{min}$)/(life exp$_{max}$ – life exp$_{min}$)

Income index$_i$ = [ln(income$_i$) – ln(income$_{min}$)]/[ln(income$_{max}$) – ln(income$_{min}$)]

Education index$_i$ = (2/3)[(adult lit$_i$ – adult lit$_{min}$)/(adult lit$_{max}$ – adult lit$_{min}$)]
  + (1/3)[(enrollment$_i$ – enrollment$_{min}$)/(enrollment$_{max}$ – enrollment$_{min}$)]

$\Rightarrow$ HDI$_i$ = (life expectancy index$_i$ + income index$_i$ + education index$_i$)/3

The subscript *i* refers to the values of country *i*, and the subscripts *min* and *max* refer to minimum and maximum comparison levels. In 2000 the maximum values were taken to be 85 years for life expectancy, $40,000 for income (per capita GDP), and 100 percent for adult literacy and enrollment rates. The minimum values were taken to be 25 years for life expectancy, $100 for income, and 0 percent for adult literacy and enrollment rates (UNDP 2000, p. 269).

The appeal of a single-dimensional measure of welfare is clear. It is generally impossible to evaluate whether one individual is better off than another unless all the dimensions of the individuals' experiences can be aggregated into a single one, on the basis of which comparisons can be made. A serious cost-benefit analysis of development programs requires an outcome measure precisely like this, one that can in addition be directly compared with the costs incurred in implementing the program.

The main motivation behind an index based on human capital seems to be a desire to use it as an outcome measure, to assess whether development programs have their intended effects. The HDI tries to capture this idea by incorporating longevity and education into a measure of development. But as we will see, it has several theoretical drawbacks.

The Human Development Index and Income-Equivalent Compensation Measures

The HDI has often been used to obtain a picture of countries' state of need and development over time. A large literature has criticized the index for its potential problems (see, for example, Kelley 1991; Srinivasan 1994; Acharya and Wall 1994; Gormely 1995; Paul 1996; Noorbakhsh 1998; and Sagar and Najam 1998). In this section we show that under certain conditions the HDI produces rankings of countries similar to those obtained from economic measures of income compensation. Unfortunately, it is under these conditions that per capita income alone is a sufficiently good indicator of overall development. In other words, these are precisely the conditions in which the HDI and income rankings coincide.

The HDI can be generally described as follows: Given an environment characterized by the vector of characteristic $x = (x_1,...,x_K)$, such as per capita GDP, literacy, and longevity, let $\underline{c} = (\underline{c}_1,...,\underline{c}_K)$ be a set of minimum comparison levels and $\overline{c} = (\overline{c}_1,...,\overline{c}_K)$ be a set of maximum comparison levels. The HDI can then be written as the following index *H:*

$$H = \frac{1}{K} \sum_{k=1}^{K} \frac{x_k - \underline{c}_k}{\overline{c}_k - \underline{c}_k}.$$

This expression can be rewritten as the linear combination

$$H = \alpha_0 + \sum_{k=1}^{K} \alpha_k x_k$$

where
$$\alpha_0 = -\frac{1}{K}\sum_{k=1}^{K}\frac{\underline{c}_k}{\overline{c}_k - \underline{c}_k} \text{ and } \alpha_k = \frac{1}{K(\overline{c}_k - \underline{c}_k)}.$$

Although the different aspects of human development may appear to be weighted equally through $1/K$, the comparison vectors $\underline{c}$ and $\overline{c}$ strongly affect the final weighting of the different characteristics, as the expression for α_k makes clear. The larger the difference between the minimum and maximum comparison levels for a given characteristic, the smaller the final weight attached to that characteristic.

There are several important distinctions between the HDI and an economic measure of welfare differences among countries. First, and perhaps most important, the HDI does not reflect the tradeoffs faced by the population whose welfare is being evaluated, but arbitrary weights assigned by the third party doing the evaluation. An economic index attempts to reflect the tradeoffs faced by the people involved, by looking at how their behavior reveals these tradeoffs.

Second, a quantitative interpretation of an index like the HDI is difficult. We can perhaps say that a country with an HDI of 0.2 is much poorer than one with an HDI of 0.8. But the economic meaning of "much poorer" is difficult to understand, since the quantitative value of the index does not have any interpretation that lends itself to cost-benefit calculations. More specifically, the HDI does not produce any number that can be directly used in a cost-benefit analysis of development interventions, a property that, in principle, any useful development index should have. For example, an economic index of development can give monetary values to the welfare gains brought about by development interventions and to the welfare differences across countries—differences that can be compared to the monetary costs of the program. Even though the HDI has been proposed to capture aspects of development that cannot be captured by income alone, these aspects can still be converted into income units. An economic index does that by simply measuring the values of different attributes in income alone.

Finally, the weights used in the HDI do not depend on the levels of the variables being analyzed and are not country specific. As the literature on the value of life has shown, the well-being generated by a given longevity depends on the consumption throughout life, and vice versa. This means that the relative values of income and longevity depend on the levels of these same variables and on the way in which they interact. The HDI assumes fixed weights for the subindexes and thus ignores these interactions.

Another issue that arises in constructing a human capital index of development is the relevance of the dimensions incorporated in determining the final ranking of countries. When the goods are endogenously chosen and their consumption rises with income—that is, they are normal goods—their inclusion does not contribute much in the ranking of countries. Richer countries consume more of normal goods, so comparing income and comparing an index of income and a bundle of normal goods will yield similar orderings. In the extreme case where the utility function is homothetic in all goods, so that budget shares remain constant across income levels, the rankings will be exactly the same, and income alone will suffice to classify countries by welfare level.

Moreover, in the specific case of the HDI, the fact that education is an input into the production of goods and health makes its inclusion in the welfare index a complicated issue. At the extreme, if education has no consumption value but only investment value, it should be excluded from the index to avoid double counting. By this, we do not mean that people do not derive some direct utility from education. The point is that as long as decisions relating to education are based in part on financial returns to investments in human capital, education cannot be seen purely as a consumption good, as it implicitly is in the HDI. This inevitably implies some degree of double counting, through the investment itself (the education measure) and through its outcome (income). In principle, economic analysis can deal with this problem by disentangling the values of education as consumption good and investment from individual market behavior. The parameters obtained from this type of analysis can then be used to model preferences in such a way that education receives the consumption weight that people actually attach to it.

To compare HDI to an index based on the economic concept of income-equivalent compensation, consider a distribution $F(x_1,..., x_K)$ of characteristics across countries. E represents the income-equivalent compensation (EV) defined by the income necessary to make an individual indifferent between facing his own and the maximum comparison environment:

$$U\left(x_1 - E, x_2,..., x_K\right) = U\left(\bar{c}_1,...,\bar{c}_K\right) = U\left(\bar{c}\right).$$

where $U(x)$ is an indirect utility function over income and other aspects of development.

We define E as a negative number, such that countries with smaller values of E (greater absolute values) have lower welfare, and those with greater values of E (closer to zero) have higher welfare.

The intuition for this index should be clear. The absolute value of E ($|E|$) indicates the additional amount of good 1 that would make an individual indifferent between the bundles of goods x and $\bar{c}$.

For a given $\bar{c}$, Taylor expanding the indirect utility function around the point $x = \bar{c}$, we get its linear approximation:

$$U\left(x_1 - E, x_2,..., x_k\right) \approx U\left(\bar{c}\right) + \sum_{k=1}^{K} \frac{\partial U\left(\bar{c}\right)}{\partial x_k}\left(x_k - \bar{c}_k\right) - \frac{\partial U\left(\bar{c}\right)}{\partial x_1} E.$$

The condition defining EV, $U\left(x_1 - E, x_2,..., x_k\right) = U\left(\bar{c}_1,...,\bar{c}_k\right) = U\left(\bar{c}\right)$, can then be expressed as

$$E \approx \sum_{k=1}^{K} MRS_k\left(x_k - \bar{c}_k\right),$$

where $MRS_k = \dfrac{\partial U\left(\bar{c}\right)/\partial x_k}{\partial U\left(\bar{c}\right)/\partial x_1}$

is the marginal rate of substitution between the *k*th good and the first good, evaluated at point $\bar{c}$.

Therefore, as was the case for the HDI, the EV may be approximated by a linear combination, given by

$$E \approx \beta_0 + \sum_{k=1}^{K} \beta_k x_k$$

where $\beta_0 = -\sum_{k=1}^{K} MRS_k \, \bar{c}_k$ and $\beta_k = MRS_k$.

The higher the marginal rate of substitution between a given characteristic and income, the higher its weight in *E*. So, in principle, the EV index can attribute weights to the different dimensions of development that are determined precisely by the relative value attached to these dimensions.

The EV index and the HDI thus coincide only under very restrictive conditions on the utility function *U*. In other words, if one were to interpret the HDI as a measure of willingness to pay, it would make sense only under a very restrictive set of preferences, probably not those that would be estimated from observed demand behavior in the population of interest.

The EV analysis offers a natural way of comparing the welfare of a country with some reference point, by translating the differences in the dimensions of welfare into only one dimension, income. In addition, an economic index that tries to measure the welfare distance between a given country and the maximum level currently attainable should not be affected by the minimum level currently observed. Introducing the lowest comparison levels in the HDI arbitrarily changes the weights of the variables.

Despite these differences, there are conditions under which the two indexes generate similar rankings. To see this, consider the case of only two attributes (x_1, x_2), which are distributed across countries according to $F(x_1, x_2)$. How similar the indices are may be represented by the covariance between the statistics *H* and *E* which is given by

$$Cov(H, E) = \alpha_1 \beta_1 \sigma_1^2 + \alpha_2 \beta_2 \sigma_2^2 + (\alpha_1 \beta_2 + \alpha_2 \beta_1) \sigma_{12}$$

where σ_1^2, σ_2^2, and σ_{12} are the variance of attribute 1, the variance of attribute 2, and the covariance between attributes 1 and 2, respectively.

This implies that the more the two characteristics covary, the more similar the rankings obtained from the two indexes will be. The weighting of the two goods does not matter so much when there is strong covariance between the goods. For income and longevity, if the richest countries are also those in which people live longest, it does not matter whether we rank the countries by income, by longevity, or by some intermediate weighted average. All three produce similar rankings. In particular, income rankings of countries tend to be preserved whenever the added goods are correlated with income; that is, goods are normal.

There are important economic incentives to expect different forms of human capital and income to be highly positively correlated. Education and income would be positively correlated when greater investments in education raise income (Becker 1962). Longevity and income would be positively correlated when larger incomes generate larger investments in health human capital (Grossman 1972). Indeed, as mentioned, when education is simply an input into the production of human capital and will therefore generate higher market and nonmarket income, measuring education as a component of overall well-being double counts its value.

This illustrates that adding a dimension to a development index will not change rankings much when the added good is a normal good, that is, a good consumed in greater quantity in richer countries. Because this implies a positive covariance between income and the good, adding the good will not lead to a substantially different ranking of countries from that generated by income alone. More precisely, because endogenous human capital investments in health and education have a two-way positive relationship with income, adding human capital dimensions to an income measure is unlikely to generate substantially different rankings. But the HDI or any other measure of economic well-being that goes beyond measuring income is likely to include goods that have a strong positive correlation with income. Those creating a development index will often argue that the goods consumed in more developed (richer) countries are precisely what should be included in such an index.

To sum up, the problem in using the HDI is that it reflects *income-equivalent measures only when adding extra goods does not matter*. In other words, the HDI mimics economic measures only when a simple income ranking would be enough.

An Empirical Examination of the Two Indexes: Levels and Changes

In this section we compare the performance of a country ranking based on the HDI for the 35 years between 1962 and 1997 with that of a ranking based on the EV concept and a ranking based on income alone. In considering aspects of human capital, we draw on the methodology discussed in detail in Becker, Philipson, and Soares (2001) for incorporating longevity gains into welfare comparisons. We compare the results obtained for life expectancy and per capita income only, since economic theory offers a straightforward way for incorporating gains in longevity into EV measures of welfare but does not for gains in education. Thus we also apply the UNDP methodology using only life expectancy and per capita income, such that HDI = (life expectancy index + income per capita index)/2.

To compute the EV index, we assume a constant elasticity form for the instantaneous utility function $u[c(t)] = c(t)^\theta$, perfect foresight, and a discount rate equal to the interest rate. This yields an indirect utility, as a function of annual income and life expectancy, given by

$$U\left(y,\tau\right) = y^\theta\left(\frac{1-e^{-r\tau}}{r}\right)$$

where y denotes annual income; τ, longevity; and r, the annual interest rate (equal to the discount rate). The important point here is that the parameter θ gives the relative weight of consumption and life span in terms of utility. The higher θ is, the more important income is relative to life expectancy, and vice versa. The EV index can be immediately obtained from the expression for the indirect utility (see Becker, Philipson, and Soares 2001).

Ideally, the implementation of this methodology would involve country-specific estimation of the preference parameter θ, which can be obtained from labor market data on compensation and job-specific mortality rates (see Murphy and Topel [1999] for further details). But the value of θ has been estimated for only a few countries and for the vast majority the necessary data do not even exist. Nevertheless, the methodology is in principle applicable, and the data will probably become increasingly available. In the preliminary exercise in the next section we assume the same value of θ for all countries and change this value to check the sensitivity of the results to the different specifications.

For both the HDI and the EV index the reference points are based on the values chosen by the UNDP in *Human Development Report 2000*, and interest rates are assumed to be 2.5 percent a year.[2]

To illustrate the difference in weights implicit in the two indexes, we calculate the value of the linear approximation weights discussed in the previous section. For this exercise we calibrate the value of the parameter θ such that the value of the gains in life expectancy in the United States between 1970 and 1990 is consistent with the estimates of Murphy and Topel (1999).[3] With an interest rate of 2.5 percent a year, this gives $\theta = 0.1032$. Using these values for θ and r, we calculate the coefficients associated with per capita income and longevity in the linear expressions for the HDI and the EV index.[4] The HDI in this case assigns a relative weight of 0.91 to the natural logarithm of annual income, and 0.09 to longevity. The EV index assigns a relative weight of 0.97 to the logarithm of annual income, and 0.03 to longevity.[5]

These different relative weights mean that the HDI implicitly assumes preferences that do not correspond to what can be inferred from individual behavior. In this example the HDI weights imply a value of $\theta = 0.0340$, a third of the $\theta = 0.1032$ calibrated using the value of gains in longevity estimated by Murphy and Topel (1999) for the United States.

We now analyze the importance of these methodological differences, given the observed changes in income and life expectancy throughout the world. Since the specific values of the HDI bear little meaning, we focus the analysis on the country rankings obtained from the different indexes rather than on the values calculated for a particular country. The data used in the calculations come from the World Bank's *World Development Indicators 1999*. The variables are life expectancy at birth and real GNP per capita. The sample includes all economies (83) for which data were available for all points at five-year intervals between 1962 and 1997 (see appendix for detailed definitions and the economies in the sample).

We rank the 83 economies in the sample by the HDI, the EV index, and income alone, for each of the eight points at five-year intervals between 1962 and 1997. We

then calculate the correlation between the rankings and the correlation between the changes in the rankings from one interval to the next. Because the EV ranking depends on the parameter θ of the utility function, we try different values, covering the entire range of concave functions for the functional form adopted (θ = 0.1, 0.5, and 0.9).[6]

The correlation between the HDI and EV rankings is extremely high in all cases, no matter what the value of θ is (table 1). The correlation between the HDI and income rankings is also very high, more than 0.95 in all cases. This outcome results from the fact that adding a new dimension to an index has little effect on the ranking by that index as long as the new and existing dimensions are highly correlated. And this is certainly the case for the logarithm of income and life expectancy at birth: the correlation between these two variables exceeds 0.83 for all years in the sample. So rankings of development obtained from the HDI, from an EV index, and from income alone will generally be extremely similar. In figures 3 and 4, which plot the 1997 HDI ranking against the EV ranking (for θ = 0.5) and the income ranking, all points are concentrated around the 45-degree line, indicating the close association between the rankings.

But the changes in these rankings probably have a much smaller correlation, since recent changes in longevity have been largely unrelated to changes in income. The correlation between the changes over time in the income ranking and those in the HDI ranking shows that much of the story about changes in the relative positions of countries might be missed if we look at income only (see table 1). But the changes in rankings by the HDI are also quite different from those in rankings by the EV index. And the smaller θ is, the worse the HDI performs compared with the EV index.[7] The HDI's poorer performance, relative to the EV index, in reflecting changes in welfare levels can be seen from the greater dispersion of points around the 45-degree line in figure 5 as compared with that in figure 3.

Even for rankings of countries at a point in time, the good performance of the HDI results not from some property of the index, but from the fact that income and longevity are highly correlated at a point in time. The results in table 1 illustrate the

Table 1. Correlation Between Rankings and Changes in Rankings for Different Development Measures, 1962–97

	1962	*1967*	*1972*	*1977*	*1982*	*1987*	*1992*	*1997*
				θ = 0.1				
HDI and EV	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97
ΔHDI and ΔEV		0.22	0.23	0.49	0.60	0.56	0.70	0.69
				θ = 0.5				
HDI and EV	0.97	0.97	0.98	0.99	0.99	0.99	0.99	0.99
ΔHDI and ΔEV		0.31	0.32	0.59	0.61	0.63	0.62	0.73
				θ = 0.9				
HDI and EV	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99
ΔHDI and ΔEV		0.33	0.46	0.71	0.71	0.69	0.59	0.73
HDI and GNP	0.96	0.97	0.97	0.98	0.98	0.98	0.97	0.98
ΔHDI and ΔGNP		0.50	0.57	0.61	0.74	0.62	0.46	0.50

Source: Authors' calculations based on data from the World Bank (1999).

Figure 3. Comparison of Rankings by the HDI and the EV Index, 1997

EV index ranking

HDI ranking
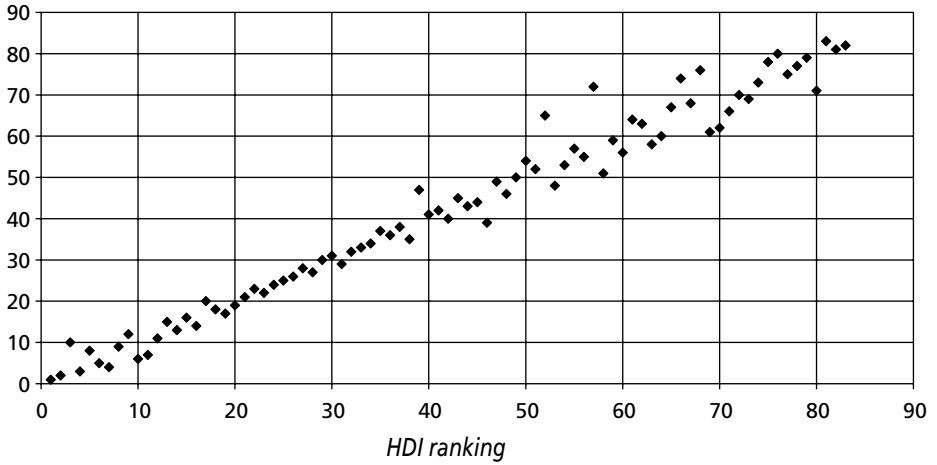
Source: Authors' calculations based on data from the World Bank (1999).

Figure 4. Comparison of Rankings by the HDI and Income, 1997

Income ranking

HDI ranking
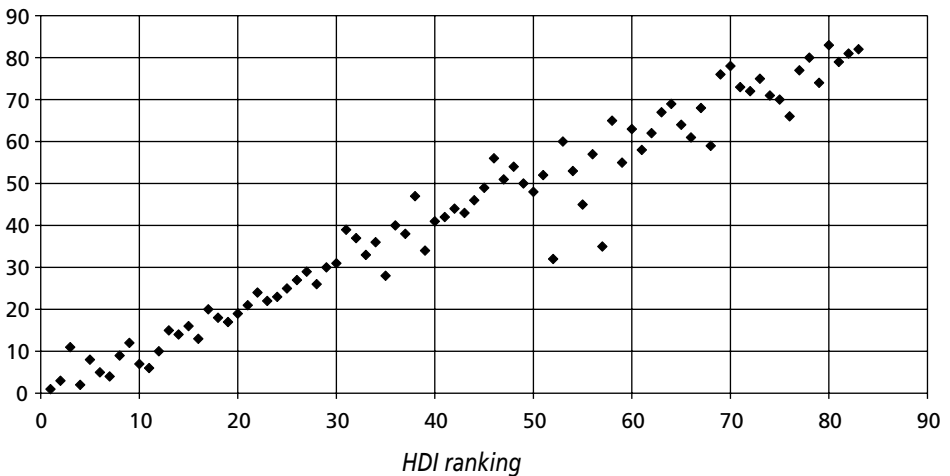
Source: Authors' calculations based on data from the World Bank (1999).

theoretical point discussed before, that the HDI will perform well only when an income index alone would be enough. As long as the dimensions being analyzed are highly correlated, it does not matter how different indexes are constructed; all of them will generate very similar rankings. Once the dimensions become unrelated,

Figure 5. Comparison of Changes in Rankings by the HDI and the EV Index, 1992–97

Change in EV index ranking

Change in HDI ranking
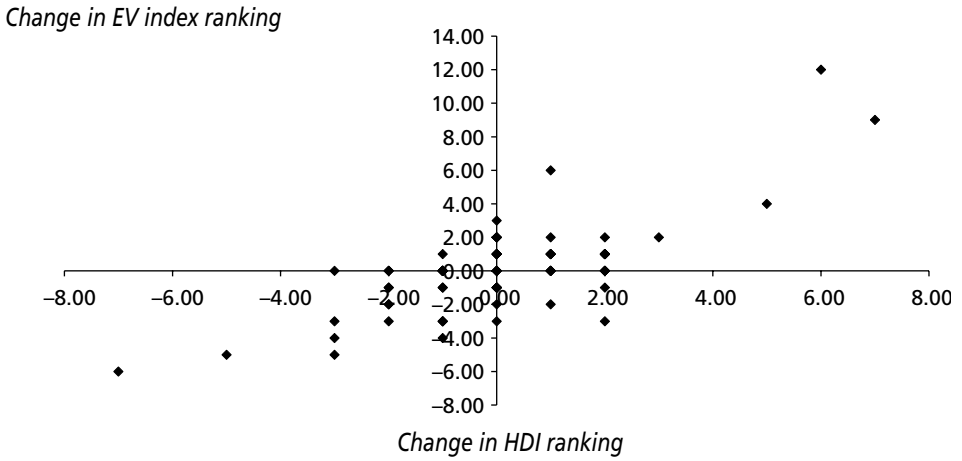
and thus important, the HDI gives results quite different from those obtained with an economic index, such as that based on the EV concept.

Conclusion

The Human Development Index was devised to capture aspects of human development not reflected in income alone. An income-equivalent compensation measure of human development similarly may take into account such nonmonetary aspects of development but nevertheless assesses their value in monetary terms.

Drawing on recent methodology for assessing the value of longevity gains at the national level, we have compared the HDI with income-equivalent compensation measures at the international level, incorporating the longevity gains observed worldwide in 1962–97, with a view to the application of these measures in cost-benefit analysis of development programs. We have argued that the HDI might be difficult to use in quantitatively assessing progress in development, since it does not generate a number that can be directly compared with the costs of investing in development projects. But it can accurately rank countries by development level because the aspects of human development it adds to income are highly correlated with income, so that weighting those aspects becomes less important. Unfortunately, this is precisely the situation in which income alone should give an accurate picture of development.

Many aspects of human development other than monetary income are important. The effects of development programs on these nonmonetary aspects need to be expressed in comparable units, so that progress in the different dimensions of human development can be compared. Measuring such aspects in terms of the forgone income at which they are valued appears to provide a consistent approach, and one

better suited than the Human Development Index for assessing progress in human development.

Appendix

Definition of Variables

- *Income series:* GNP per capita (constant 1995 U.S. dollars). From the World Bank's *World Development Indicators 1999.*
- *Life expectancy series:* Life expectancy at birth, total (years). From the World Bank's *World Development Indicators 1999.*

Economies in the Sample

Algeria, Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Botswana, Brazil, Burkina Faso, Burundi, Cameroon, Canada, Central African Republic, Chad, Chile, Democratic Republic of Congo, Republic of Congo, Costa Rica, Côte d'Ivoire, Denmark, Dominican Republic, Ecuador, Arab Republic of Egypt, El Salvador, Fiji, Finland, France, Gabon, Ghana, Greece, Guatemala, Guyana, Honduras, Hong Kong (China), Indonesia, Ireland, Israel, Italy, Jamaica, Kenya, Republic of Korea, Lesotho, Madagascar, Malawi, Malaysia, Malta, Mauritania, Mauritius, Mexico, Morocco, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Papua New Guinea, Paraguay, Philippines, Portugal, Rwanda, Saudi Arabia, Singapore, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Thailand, Togo, Trinidad and Tobago, Tunisia, United Kingdom, United States, Uruguay, Zambia

Notes

1. Another attempt to incorporate dimensions other than income into welfare analysis is that by Atkinson and Bourguignon (1982), who devise a methodology for analyzing inequality in the context of social welfare functions, with multidimensioned distributions of attributes.

2. The maximum values are 85 years for life expectancy and $40,000 for per capita income; the minimum values are 25 years for life expectancy and $100 for per capita income (UNDP 2000, p. 269). None of the results is particularly sensitive to the choice of interest rate.

3. The estimates in Murphy and Topel (1999) imply that the gains in life expectancy between 1970 and 1990 had a per capita value of $135,621.

4. Since the HDI uses the natural logarithm of income as the characteristic, we relabel the utility function such that the characteristic is also taken to be the logarithm of income. Weights are calculated relative to this variable for both indexes.

5. The apparently small relative weights associated with longevity in both indexes are due to the metric used on the income measure, that is, the natural logarithm. Relatively small changes in the natural logarithm of per capita income are associated with huge changes in the level of per capita income. For example, if income rises from $1,000 to $2,000, the logarithm changes by only 0.7 point, from 6.9 to 7.6. Because longevity is measured in years and can range from such values as 50 to 70, the weights tend to compensate for these differences in

metric. Also because of this, the difference of 0.06 point between the weights assigned to longevity in the HDI and the EV index are actually associated with considerable differences in implied preferences.

6. $\theta = 0.1$ corresponds to the value of life calculated in Murphy and Topel (1999).

7. The difference between this result and what might be expected based on the linear approximation weights probably stems from the high nonlinearity of the utility function and thus the poor performance of the linear approximation when dealing with changes of the magnitude observed in the data. For the rankings discussed here calculations were based on the actual formulas, not their linear approximations.

References

Acharya, Arnab, and Howard J. Wall. 1994. "An Evaluation of the United Nations' Human Development Index." *Journal of Economic and Social Measurement* 20(2): 51–65.

Atkinson, A. B., and François Bourguignon. 1982. "The Comparison of Multi-Dimensioned Distributions of Economic Status." *Review of Economic Studies* 49(2): 183–201.

Barro, Robert J., and Xavier Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw-Hill.

Becker, Gary S. 1962. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy* 70(5): 9–49.

Becker, Gary S., Tomas Philipson, and Rodrigo R. Soares. 2001. "Growth and Mortality in Less Developed Nations." University of Chicago, Department of Economics.

Cutler, David, and Elizabeth Richardson. 1997. "Measuring the Health of the U.S. Population." Brookings Papers on Economic Activity: Microeconomics Issue (1997): 217–71.

Gormely, Patrick J. 1995. "The Human Development Index in 1994: Impact of Income on Country Rank." *Journal of Economic and Social Measurement* 21(4): 253–67.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80(2): 223–55.

Kelley, Allen C. 1991. "The Human Development Index: Handle With Care." *Population and Development Review* 17(2): 315–24.

Murphy, Kevin M., and Robert Topel. 1999. "The Economic Value of Medical Research." University of Chicago, Graduate School of Business.

Noorbakhsh, Farhad. 1998. "A Modified Human Development Index." *World Development* 26(3): 517–28.

Nordhaus, William D. 1999. "The Health of Nations: The Contribution of Improved Health to Living Standards." Yale University, Department of Economics, New Haven, Conn.

Paul, Satya. 1996. "A Modified Human Development Index and International Comparison." *Applied Economics Letters* 3(10): 677–82.

Preston, Samuel. H. 1975. "The Changing Relation Between Mortality and Level of Economic Development." *Population Studies* 29(2): 231–48.

Sagar, Ambuj D., and Adil Najam. 1998. "The Human Development Index: A Critical Review." *Ecological Economics* 25(3): 249–64.

Srinivasan, T. N. 1994. "Human Development: A New Paradigm or Reinvention of the Wheel?" *American Economic Review Papers and Proceedings* 84(2): 238–43.

UNDP (United Nations Development Programme). 2000. *Human Development Report 2000*. New York: Oxford University Press. [http://www.undp.org].

World Bank. 1999. *World Development Indicators 1999*. Washington, D.C.

Comment on "World Inequality and the Rise in Longevity," by Tomas Philipson and Rodrigo R. Soares, and "Health, Income, and Economic Development," by Anne Case

S. R. Osmani

The articles by Anne Case and by Tomas Philipson and Rodrigo R. Soares are both concerned with the relationship between income and health, but in different ways. Philipson and Soares deal with the relationship of "welfare equivalence"—the amount of income that is equivalent to health in terms of welfare. Case addresses a causal relationship—the causal impact of income on health. Philipson and Soares seek to use the income equivalent of health to replace multidimensional measures of development, which combine statistics on income and health, with a unidimensional measure based on income alone. Case seeks to isolate the causal effect of income on health from the reverse effect of health on income, while recognizing that both these effects may underlie the observed positive association between income and health.

Both articles are thought provoking, and there is a good deal to learn from each. However, rather than highlighting the strengths of the articles, a job the authors have done well, I have chosen to focus my comments on the areas that I believe deserve further reflection.

Comparing Measures of Welfare

Philipson and Soares take issue with the United Nations Development Programme's Human Development Index (HDI), a multidimensional index of progress in well-being that combines information on per capita income, health, and education. As is well known, the HDI is based on the premise that society values not just income but also nonincome dimensions of well-being, such as health and education. For Philipson and Soares, however, recognizing this multidimensionality of values does not necessarily require adopting a multidimensional index such as the HDI. Instead, they offer what they call the "income-equivalent compensation" measure, which is constructed by collapsing nonincome dimensions into income through the use of

welfare equivalence. The resulting unidimensional index is, they say, superior to the HDI in a number of ways.

The criticism of the HDI is nothing new. There are indeed good reasons to criticize it, and even its architects recognize many of its limitations. It is unclear, however, that the alternative measure proposed by Philipson and Soares overcomes the limitations that motivate their search for an alternative.

Philipson and Soares argue that an index like the HDI cannot be sensibly used in a cost-benefit analysis of development intervention, while their alternative measure can be. The first point to note here is that the HDI has been recognized by its architects all along as a crude measure of welfare—serving the useful purpose of drawing attention to the nonincome dimensions of well-being but operating at the same level of "vulgarity" as GNP when it comes to dealing with the subtleties of welfare comparison (UNDP 1999). To expect such a crude measure to bear the burden of a sophisticated cost-benefit analysis would be too much to ask in any case.

Nevertheless, if the HDI were forced into service in cost-benefit analysis, it would not fail for the reason that Philipson and Soares suggest. They attribute its failure to its alleged inability to produce an economically meaningful number in the same way that the income-equivalent compensation measure does. In claiming that their alternative measure is economically meaningful, Philipson and Soares presumably mean that their measure, by converting the value of health into equivalent income, expresses everything in terms of the common numeraire income, which is an economically meaningful concept. They contrast this with the values of the HDI, such as 0.2 or 0.8, whose economic meaning is hard to understand.

But if this is the problem with the HDI, it can be easily remedied—for the HDI too can be converted into an income equivalent. The weights used in combining the three components of the HDI imply certain rates of tradeoff between these components. The rates show, for example, the income equivalent to one unit of health in terms of well-being. By using these implicit rates of equivalence, one can easily convert the HDI into a wholly income-based measure. This transformation will not, of course, make any substantive difference to welfare comparison; the point is that one can obtain an income equivalent of the HDI if that is what one is looking for.

Questions may be raised, however, about the appropriateness of the weights used in the HDI. Do they really reflect society's tradeoffs between different components of well-being? This is where Philipson and Soares's second, and perhaps the most important, objection to the HDI comes in. They note that the weights used in the HDI are arbitrary, while proposing that the weights used in their alternative measure reflect the true preferences of the people whose welfare is being considered.

The point about the HDI weights being arbitrary is entirely valid. But it is instructive to ask why they are so. In principle, these weights are supposed to reflect society's tradeoffs between different components of well-being, so there is no arbitrariness in principle. In practice, though, it is exceedingly hard to ascertain the size of these tradeoffs. And we know why this is so from the theories of social choice and of preference revelation. For practical purposes, therefore, the analyst or policymaker must make a pragmatic judgment, and that's where the arbitrariness comes in.

But this problem is not unique to the HDI; it is endemic in any practical exercise in social welfare comparison. Standard cost-benefit analyses must confront the problem all the time. Some crucial parameters of cost-benefit analysis—such as the social rate of discount, the shadow price of savings, or the distributional weights to be attached to the welfare gains of the poor—are often based on the pragmatic judgment of the analyst or policymaker. The arbitrariness inherent in this process can be reduced by carrying out sensitivity analysis—using alternative values of the parameters—but it cannot be eliminated.

Even the alternative measure proposed by Philipson and Soares is not immune to this problem. In the practical application of their measure, they impose the value of the parameter θ, which reflects the relative weights between income and life expectancy, rather than deriving it from people's revealed preferences (as they claim earlier in the article). In the best tradition of cost-benefit analysis, however, they do carry out a sensitivity analysis, by using alternative values of the parameter. But nothing prevents an analyst from carrying out a similar sensitivity analysis with the HDI if it were used for cost-benefit analysis or even when it is used for ranking countries by human development.

Thus neither the ability to yield an "economically meaningful" measure based on income nor the issue of arbitrariness of weights really distinguishes the proposed measure from an HDI-type index. The proposed measure does differ from the HDI in one important sense, but this difference constitutes a problem rather than an improvement.

In deriving the income equivalent of health, as proxied by longevity, Philipson and Soares consider the additional consumption made possible by additional years of life. Thus the value of health is derived solely from the fact that health is a form of human capital that, like any other capital, yields a stream of consumption benefits over time. Their approach is thus firmly rooted in the human capital literature.

And therein lies the problem. Health and education are valued by society for two distinct reasons. They are instrumentally valuable in their role as human capital. But they are also intrinsically valuable—that is, valuable as ends in themselves—regardless of how much additional consumption of goods and services they make possible. It is the desire to capture this intrinsic value that motivates the HDI. If the instrumental value were the primary concern, there would be no justification for having a composite index that incorporates both income and capital, for this would constitute double counting (as Philipson and Soares rightly note in the context of education).[1] In any case, there would be no ground for privileging health and education over other forms of capital, such as roads and factories. It is the recognition that, unlike roads and factories, health and education are valuable in themselves that motivates a composite index like the HDI.

Conceptually, therefore, the HDI and the proposed measure seem to belong to two different categories. The proposed measure is concerned with the instrumental value of health, while the HDI is concerned with its intrinsic value (without, of course, denying the existence or the importance of its instrumental value). Philipson and Soares thus seem to be making a "category mistake" in comparing their measure with the HDI.

The two indexes would be comparable only if the instrumental value coincided with the intrinsic value at the margin. In that case it would be possible to argue that, like the HDI, the proposed measure too incorporates health for its intrinsic worth but looks at its instrumental value for the practical purpose of calculating its income equivalent. But the conditions under which this equality would hold are far too stringent to be met in practice. A sufficient condition is the existence of a grand optimum in which resources are allocated in a way that maximizes the value of a social welfare function that includes both income and health as its arguments. But when one notes that private provision of health care (and education) is rife with pervasive market failures and that public provision is beset with the problem of eliciting the true preferences of individuals, it becomes obvious that such a grand optimum is nothing more than a grand utopia. In reality, therefore, the instrumental and intrinsic values of health (and education) would almost certainly diverge. As a result, an index of the type proposed by Philipson and Soares would remain noncomparable to the HDI-type index.

Relating Health to Income

I now turn briefly to the causal relationship between income and health explored by Case. To isolate the effect of income on health, Case looks for a source of income that is not determined by health status, and finds it in the old age pension given by the state. Her results show that households with pension-eligible members enjoy better health status than those without such members. Case takes this as evidence of the protective effect of income on health, uncontaminated by the reverse effect of health on income.

Case then goes further, exploring the mechanisms through which income might have affected health—perhaps potentially the most useful part of the exercise for policy. She identifies four possible channels through which income might affect health and finds that at least three of them have been at work for the sample studied. But the evidence seems less convincing than it might have been. Case takes the association between pension income and a variable representing a particular channel as evidence that that channel is working. But it is not enough to show that income has an effect on that channel; it is also necessary to show that the channel has an effect on health. Case has investigated this association for only one channel, however, considerably weakening her causal analysis.

There is a more serious problem with the causal analysis. Case interprets the observed association between pension and health status as the causal effect of income on health. This interpretation seems somewhat problematic, however, when one recalls that she has also found a positive association between self-reported health status and whether the respondents suffer from a number of chronic diseases. Note that the association is with the existence rather than the severity of diseases. The observed association between pension and health must therefore be interpreted as an association between pension and the incidence of chronic diseases. The implication is that households with pensioners suffer from a lower incidence of chronic diseases.

But can this association be interpreted as evidence of a causal impact of pension income on health? It seems highly improbable that household members suffering from chronic diseases would suddenly cease to do so when someone in the household begins to receive a pension. Some improvement might occur over a long period, but only if pension income continues to be received for long enough, and then only for reversible diseases. But to capture this effect, the explanatory variable ought to be the length of time during which a pension has been received, not just the binary information on whether someone in the household receives a pension or not.

Interestingly, an alternative interpretation is possible, but it is based precisely on the reverse causation, which Case sought to avoid. When the members of a household are habitually healthy—that is, no one suffers from chronic diseases—one of them is likely to survive long enough to be eligible for an old age pension. But in a household in which most suffer from bad health, none might survive long enough to receive a pension. This may explain the observed association between pension and health. In this case causation runs from health to (pension) income, in the sense that whether a household will have a pension-eligible member or not depends on its members' health status. Case had hoped to avoid reverse causation by choosing pension as the explanatory variable on the ground that pensionable age is not determined by health. The problem is that while pensionable age does not depend on health, whether someone survives to that age certainly does.

One final observation: While exploring the channels through which income might protect health, Case has come up with the apparently surprising result that better access to health care was not among these channels. This is certainly counterintuitive, since one would think that richer people would have better access to health care facilities. Does this finding mean that higher personal income cannot increase people's access to health care? Case sensibly refrains from drawing any such conclusion. Instead, she observes that in view of the poor quality of health care facilities, people might wisely have decided not to spend their extra income on them.

This observation carries an important lesson. The health policy literature often debates whether health is better achieved through higher private income or greater public provision of health care facilities. Indeed, Case refers to the choice between these two routes as the motivation for her article. But her results show that posing the problem as a choice of this kind may be misleading. They suggest a strong complementarity between private income and public provision of health care facilities. When existing facilities are poor, higher private income may not yield much benefit through the route of access to health care, as Case has found. Where facilities are better, income may well turn out to have a big impact. But in both cases the coefficient of the income variable does not by itself indicate anything about the efficacy of income, independent of the availability of health care facilities.

Case is conscious of this. That's why, on finding a nonsignificant coefficient, she is wise enough not to conclude that income does not matter. It is to be hoped that future authors, on finding the opposite result, would be equally wise, refraining from asserting that income is all that matters or even that it matters more than the provision of health care facilities.

Note

1. Strictly speaking, double counting would occur only if a measure of permanent income were used in addition to capital. When actual income for a specific period is used, there may be a supplemental role for capital to reflect lifetime consumption possibilities. This is indeed how Philipson and Soares's measure avoids double counting even though it focuses exclusively on the human capital aspect of health.

Reference

UNDP (United Nations Development Programme). 1999. *Human Development Report 1999*. New York: Oxford University Press.

Comment on "World Inequality and the Rise in Longevity," by Tomas Philipson and Rodrigo R. Soares

Sudhir Anand

Tomas Philipson and Rodrigo Soares have a twofold objective: to formulate a full income measure of human development that takes account of both income and longevity but is "less arbitrary" than the Human Development Index (HDI), and to reflect in this full income measure "the tradeoffs faced by the people involved, by looking at how their behavior reveals the tradeoffs" (p. 6). Philipson and Soares claim that "the HDI does not reflect the tradeoffs faced by the population whose welfare is being evaluated, but rather arbitrary weights assigned by a third party doing the evaluation" (p. 6).

Without wishing to defend the HDI, which is admittedly both arbitrary and crude, I would like to suggest that Philipson and Soares's approach does not appear to be any less arbitrary. Moreover, the authors' approach is not based on people's revealed behavior relating to their tradeoffs between income and longevity.

As a third party themselves, Philipson and Soares specify for the representative individual in society an instantaneous utility function of income that is of constant elasticity form. The individual is assumed to have perfect foresight, to live for exactly *T* years (the life expectancy at birth in the society), and to discount future utility (rather than income or consumption) at the rate *r*, taken to be 2.5 percent a year. That future utility should be discounted in this context (of complete certainty) is highly questionable and cannot in my opinion be justified, but this issue is not central to the points I wish to make.

The present value of the constant instantaneous utility stream *u(y)* for *T* years, discounted at the rate *r*, is what Philipson and Soares call the indirect utility function:

$$U\left(y, T\right) = u\left(y\right)\left(\frac{1 - e^{-rT}}{r}\right)$$

Sudhir Anand is professor of economics, University of Oxford, and fellow of St. Catherine's College, Oxford. This comment refers to the version of the article by Tomas Philipson and Rodrigo Soares (entitled "Human Capital, Longevity, and Economic Growth: A Quantitative Assessment of Full Income Measures") that was delivered at the conference, available on the Web at http://econ.worldbank.org/abcde/.

which is a function of annual income y and life expectancy T (notation as in the conference version of the article).

For simplicity (and correctness), let us assume that $r = 0$, so that the indirect utility $U(y, T)$ is simply $u(y).T$, the sum total of utility over the individual's life span T. With constant elasticity θ for the instantaneous utility function, the welfare of the representative individual is simply $U(y, T) = y^\theta.T$. Philipson and Soares consider values of θ between 0 and 1 ($\theta = 0.1, 0.5,$ and 0.9) in table 1, "covering the entire range of concave functions for the functional form adopted" (p. 13). For these values of θ, the welfare function $y^\theta.T$ is a quasi-concave (but convex) function of the variables (y, T). As Philipson and Soares state (p. 9), "the parameter θ gives the relative weight of consumption and life span in terms of utility." From this welfare function they construct the equivalent variation (EV) index.

Philipson and Soares's approach is certainly one way of incorporating longevity into a welfare index, and it may even be a reasonable way of doing so. But it is not the only way, and it too is open to the charge of being arbitrary. There is nothing in their article to suggest that either the specification or the parameter values of their welfare function—and hence the EV index—are based on the observed or revealed behavior of people, as they claim. Theirs is one specification of the utility function $u(y)$, and they simply consider alternative values of the parameter θ for the exercises they conduct.

Of course, it is possible to imagine other specifications of a welfare index that directly incorporates income y and life expectancy T as its arguments. Atkinson and Bourguignon (1982), for example, do just that, taking welfare to be a concave function of (y, T). Despite all its problems, the HDI too is a concave function of (y, T), being a positive affine transformation of the concave functions $\log y$ and T. Hence the HDI is actually quasi-linear in T, which confers it some advantage in EV analysis.

It is instructive to explicitly consider the marginal rate of substitution (MRS) between income y and longevity T in these two formulations, and to compare the resulting rates. The HDI is essentially additive in $\log y$ and T, while Philipson and Soares's indirect utility function is essentially multiplicative in y and T. The coefficient on $\log y$ in the HDI is 1/6, and the coefficient on T in the HDI is 1/60.[1] Hence,

$$MRS_{y,T} = -\frac{dy}{dT}\bigg|_{HDI\ constant} = \frac{1}{10}y.$$

In other words, with the HDI kept constant (along an "indifference curve"), an extra year of life requires the sacrifice of one-tenth of annual income.

In Philipson and Soares's welfare function, the marginal rate of substitution between income and longevity is given by the equation:

$$MRS_{y,T} = -\frac{dy}{dT}\bigg|_{U\ constant} = \frac{1}{\theta T}y.$$

Here the tradeoff, or the income sacrifice required for an extra year of life, depends on the values of θ and T. For two values of the parameter θ that Philipson

and Soares consider ($\theta = 0.1$ and $\theta = 0.5$) and for two levels of life expectancy T (40 and 70 years), $MRS_{y,T}$ has the following values:

$MRS_{y,T}$ for different values of θ and T in Philipson and Soares's welfare function

T θ	40	70
0.1	$\frac{1}{4}y$	$\frac{1}{7}y$
0.5	$\frac{1}{20}y$	$\frac{1}{35}y$

Of course, other things equal, the higher θ is, the more important income is relative to longevity in the indirect utility function $y^\theta.T$, so that the individual is willing to sacrifice less income for an additional year of life.

Thus we see that the $MRS_{y,T}$ in the HDI of $(1/10)y$ is encompassed by the range of marginal rates of substitution in Philipson and Soares's welfare function for different values of θ and T that they consider. Hence the HDI's marginal rate of substitution is in no sense out of bounds; it is completely within the range of the marginal rates of substitution implicitly assumed by Philipson and Soares.

Now let's consider Philipson and Soares's empirical findings. First, the regression to the mean in life expectancy in figure 2 is not surprising. Since life expectancy has an upper bound, we would expect that countries with a high life expectancy in 1962—for example, around 70 years—simply would not have the room for gains in life expectancy as large as those in countries with a low life expectancy in 1962—for example, around 40 years. Hence with generally rising incomes and improved public health measures, we would expect convergence in longevity across countries despite the lack of convergence in per capita income.

Second, we come to the rankings of the 83 countries according to the HDI, Philipson and Soares's EV index, and income (GNP) alone. The rank correlations (I assume that they are Spearman) between the HDI and the EV index are extremely high for all three values of θ (0.1, 0.5, and 0.9), for each year—0.97 or more in every case.

The rank correlations between the HDI and GNP per capita (apparently measured in constant 1995 U.S. dollars, not dollars adjusted for purchasing parity) are also very high. But Philipson and Soares do not report the rank correlation coefficients between their EV index and income (GNP per capita) or between changes in the EV index and changes in GNP—which they also should do.

The correlations between changes in HDI rankings of countries and changes in EV rankings (over five-year periods) are lower. However, this does not necessarily testify to the superiority of the EV index to the HDI in measuring change in performance.

Finally, I do not understand the relevance of section 4 ("General versus Partial Equilibrium Valuation of Health Improvements") to the rest of Philipson and

Soares's article. Nonetheless, their empirical findings do seem to support the robustness of country rankings by the HDI in relation to their own full income measure of human development.

Note

1. The coefficient on log y in the HDI is $1/(\log_e 40{,}000 - \log_e 100)$, which is equal to $1/5.99$. See UNDP (2000, p. 269).

References

Atkinson, Anthony B., and François Bourguignon. 1982. "The Comparison of Multi-Dimensioned Distributions of Economic Status." *Review of Economic Studies* 49(2): 183–201.

UNDP (United Nations Development Programme). 2000. *Human Development Report 2000.* New York: Oxford University Press.