2004

A B C D E

Annual World Bank Conference on Development Economics

Accelerating Development

Edited by

François Bourguignon and Boris Pleskovic

Annual World Bank Conference on
Development Economics
2004

Accelerating Development

Edited by
**François Bourguignon and
Boris Pleskovic**

A copublication of the World Bank and Oxford University Press

Contents

Participation, Inclusion, and Results

Scaling Up and Evaluation

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. Most 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. Most participants' affiliations identified in this volume are as of the time of the conference, May 21–23, 2003.

François Bourguignon and Boris Pleskovic edited this volume. Developmental and copyediting services were provided by Alice Faintich of The Word Doctor. The book was designed by Naylor Design, Inc. Book production and dissemination were coordinated by the World Bank Office of the Publisher.

Introduction

FRANÇOIS BOURGUIGNON AND BORIS PLESKOVIC

The Annual World Bank Conference on Development Economics seeks to expand the flow of ideas among scholars and practitioners of development policy from academia, government, and the private sector around the world. By fostering a better understanding of development and the problems developing countries face, the conference aims to enhance policymaking at the World Bank and at its partner institutions. It also provides a forum for exposition by academics and practitioners as they seek to identify and elaborate on new ideas and issues pertinent to development.

The 15th conference, held in Bangalore, India, on May 21–23, 2003, was the first such conference to be held in a developing country. The change in locale reflects the growing importance of research done in developing countries and the desire to bring such conferences closer to participants in the developing world. The theme of the conference was accelerating development, which was divided into four topics: fostering entrepreneurship, innovation, and growth; challenges of development in lagging regions; participation, inclusion, and results; and scaling up and evaluation.

The conference opened with an address by S. M. Krishna of the government of Karnataka and a keynote address by Azim Hasham Premji of Bangalore's Wipro Corporation. These were followed by four papers on fostering entrepreneurship, innovation, and growth and the challenges of development in lagging regions. The second day began with keynote addresses by Nicholas Stern, who at that time was with the World Bank, and Rakesh Mohan of the Reserve Bank of India, followed by four papers on participation, inclusion, and results and on scaling up and evaluation. Finally, the third day was devoted to specific Indian development issues on government initiatives and civil society initiatives and public and private partnerships. The rest of this introduction summarizes the keynote addresses and the papers discussed during the first two days.

François Bourguignon is senior vice president, Development Economics, and chief economist at the World Bank. Boris Pleskovic is research manager, Development Economics, at the World Bank.

Keynote Speeches: Challenges and Opportunities in Indian Development

In his opening address, **S. M. Krishna** discusses all four of the conference's topics. He notes that the acceleration of development should ensure that there is participation and inclusion, that no region lags behind other regions, and that no section of society is left behind. This implies that development should encompass equity and justice.

Krishna points to the strides Karnataka has taken in relation to economic development and fiscal reforms and Bangalore's emergence as India's knowledge capital. He underscores Karnataka's achievements in public service delivery. The intent has been to reach as many people as possible and to provide service using innovative delivery mechanisms that harness available technology. Karnataka achieved a growth rate of 7.1 percent during the period of its Ninth Plan, compared with 5.5 percent nationwide. To downsize the bureaucracy and redefine its role, Karnataka set up the independent Administrative Reforms Commission. Krishna believes that a public-private partnership provides the best delivery system in the context of moving toward a market-oriented economy.

In his keynote address, **Azim Hasham Premji** shares his beliefs about and experiences with building human capital at Wipro and the Azim Premji Foundation. He argues that all the economies that prospered in the 20th century did so because of their ability to tap human potential. Economic theories of growth have captured this idea in the last two decades. Endogenous growth theory, often referred to as the *new growth theory*, developed by Nobel Laureate Robert Lucas and Paul Romer in late 1980s, clearly established that growth cannot come from the serendipities of technological innovations, but is driven by systematic social choices. These choices primarily include investment in technological change and in activities such as education and health care that enhance human capital. If we are to accelerate economic development, we have to look carefully at these social choices and ensure that the developing societies in particular make the right decisions.

From his experience at Wipro, Premji believes that accelerating development along the lines he recommends calls for meticulous planning, willpower, commitment at the highest levels of government, and consistent pursuit of a people-oriented approach to development. Wipro's organizational beliefs evolved in the early 1970s, when speaking about beliefs and values was unfashionable. Its management articulated them to the company's rank-and-file and practiced them almost fanatically. The corporation soon found that the talent it recruited immensely enjoyed working in a value-based organization that stood up for, and even made sacrifices for, its values.

Both Wipro and the Azim Premji Foundation have a single-minded focus on significantly enhancing the quality of learning in India's elementary schools. They are working in partnership with the government, and their experience has been extremely encouraging. Most business people understand that their material investments will depreciate over time and will eventually have to be written off. Only one kind of investment will keep appreciating irrespective of interest rates and the overall economic climate: investment in human capital.

Nicholas Stern opens his keynote address by expressing his great pleasure at being back in Bangalore and at the World Bank's holding its Annual Bank Conference on Development Economics in India. Stern presents the lessons for development in a changing world and opportunities for India, noting the importance of looking ahead in a world in which ideas, resource flows, and partnerships are changing rapidly. While greater integration and change do have their risks, they also offer tremendous opportunities that collaboration between high-income and developing countries can do a great deal to enhance. The new international commitment to the Millennium Development Goals (MDGs) offers just such an opportunity.

Stern emphasizes the six lessons derived from five decades of development: the objectives of development, the complementarity between the market and the state, the role of economic growth as the most powerful force for reducing income poverty, the central role of institutions and governance in the growth process, the importance of empowerment, and the relationship between country ownership and development effectiveness. He also underscores the need to strengthen aid flows to developing countries. He believes that world leaders' embrace of the MDGs in New York in 2000 and the reaffirmation of their commitment in Monterrey points to an understanding that it is within the international community's collective power to achieve the MDGs if it has the will to do so. However, meeting the MDGs will require much stronger resource transfers, both in terms of quantity and quality. The potential returns to a successful Doha round for enhancing world trade and expanding market access to developing country exports are enormous.

Stern then turns to how the lessons, changes, and opportunities he has highlighted affect India's opportunities. He notes three areas where he believes that the need for action is pressing and could yield large rewards: strengthening public finances, especially by focusing public expenditures on development-enhancing investments; pushing forward with trade reform, both through advocacy in the international arena and unilaterally; and improving the investment climate. A final issue that cuts across all three areas of action is narrowing the widening gap between more prosperous and lagging regions, sectors, and communities. Addressing these issues will go a long way toward unfettering India's tremendous entrepreneurial talent and capacity for innovation. India's strengths lead Stern to be cautiously optimistic about its future, but reform must move ahead strongly if development is to accelerate. India is only beginning to explore some of the most difficult reform issues and still has a long way to go.

In his keynote address, **Rakesh Mohan** presents the emerging challenges of infrastructure development in India. Infrastructure contributes significantly to economic development, both by increasing productivity and by providing amenities that enhance the quality of life. Elasticities of output with respect to various stocks of infrastructure indicate that in India, as elsewhere, the transport and communication sectors play a dominant role in explaining the variations in gross domestic product and its subsectors. The index of industrial production tracked movements in the composite index of infrastructure industries closely during the 1980s and the 1990s. He argues that infrastructure must become even more of a priority than it has been if India is to achieve the kind of economic growth rates that it aspires to.

Mohan examines the macroeconomic environment that has led to a slowdown in infrastructure development in recent years. He also addresses the lack of adequate reform in the pricing and regulatory environment, which has inhibited both public and private sector investment in India. He argues that the public sector financing constraint is an objective reality by assessing recent developments in roads, telecommunications, ports, power, railways, and urban infrastructure. The lesson is twofold. First, all constraints to private sector investment must be loosened so that it can at least partially compensate for the lower than desirable level of public investment. Second, public sector levels of infrastructure investment must be raised because some infrastructure services can be viewed as public goods and others exhibit partial public good characteristics. However, raising investment levels in public sector infrastructure appropriately will be infeasible without fiscal improvements, particularly through both tax and nontax revenue increases.

Fostering Entrepreneurship, Innovation, and Growth

T. N. Srinivasan argues that innovation has a significant effect on economic growth by increasing the stock of knowledge. The process of innovation is largely endogenous, influenced by incentives, institutions, and the political economy. He analyzes the contribution of innovation to growth by drawing on some recent models of endogenous growth. Increasing openness by reducing trade barriers and encouraging foreign investment plays an important role in spurring innovation, particularly when it occurs through learning by doing. In some models of trade between the industrialized North, which innovates, and the developing South, which imitates the North's innovation, trade accelerates the rate of innovation in the North because of the threat of imitation by the South. He also explores the issues of access to capital or finance, of intellectual property protection, and of the chances of entrepreneurs' ideas being stolen or appropriated if they do not implement them themselves.

Srinivasan finds that the role of venture capital in financing start-up enterprises is important, both because it relieves financial constraints, and because the venture capitalist carefully screens the entrepreneur's ideas and monitors the enterprise after providing financing. He then turns to the phenomenal success of India's software industry and the contribution of India's education system and public policies—particularly foreign trade and investment policies—to the success of this industry. He concludes by examining the roles of patent protection and the strength of intellectual property regimes on increasing the rate of innovation and by stressing the importance of openness to foreign trade and foreign direct investment to developing a positive climate for innovation, entrepreneurship, and growth.

Robin Burgess and **Anthony J. Venables** argue that the process of structural change is central to increasing growth and raising living standards. The focus on structural change as the driving force of economic development has led most low-income countries to pursue inward-looking policies, frequently hand-in-hand with

state-led industrialization. However, their experiences are mixed, ranging from failed expectations in Sub-Saharan Africa to the dramatic success of the Asian economies. Yet despite these experiences, we know little about the likely pattern of structural change in relation to future development.

Burgess and Venables explore some of the issues surrounding the development of new activities in low-income countries. They argue that this process is frequently "lumpy," manifesting itself in the rapid growth of particular regions or sectors. In addition, spatial inequalities tend to increase during periods of rapid economic development. At the sectoral level, many rapidly growing cities, regions, and countries have export specializations in a narrow range of activities.

Recognition of these facts requires a reorientation of the analytical frameworks and empirical approaches that investigators use to examine growth. Burgess and Venables find it useful to divide the factors that are important in determining a location's growth performance into two groups that they term *first advantage* and *second advantage* (terminology adapted from economic geographers' notions of first-nature and second-nature geographies). First advantage refers to the conditions that provide the environment in which new activities can be profitably developed. They include most of the factors on which traditional theory has focused, such as access to inputs (labor and capital), access to markets, provision of basic infrastructure, and the institutional environment. Second advantage factors increase returns to scale and can lead to cumulative causation processes. They may be acquired by learning, through technological spillovers, and by the development of thick markets of suppliers and local skills. The distinguishing criterion is that second advantage factors incorporate some element of increasing returns to scale. Furthermore, these increasing returns are often external to the firm and thus associated with market failure. It is increasing returns that underlie the "lumpiness" of development.

The authors draw out how first and second advantages interact to shape the pattern of development and argue that development policy needs to take account of the different natures of these advantages. On the research side, empirical work also needs to capture this micro-level heterogeneity and diversity. They use recent empirical work on India to illustrate the importance of sectoral and spatial heterogeneity and to demonstrate the role of micro-based empirics.

Challenges of Development in Lagging Regions

Partha Dasgupta presents mechanisms that seem to be responsible for the persistence of acute poverty in Sub-Saharan Africa and large parts of the Indian subcontinent. Both the character of human metabolic pathways and weaknesses in interactions between humanity and nature play significant roles. The latter seems an obvious inference, because the rural poor in the world's poorest regions either practice subsistence agriculture or live around it. The fact that declines in mortality in recent decades have not been matched by fertility declines seems to be allied to this problem, and therefore also needs to be explained.

Toward that end, externalities in human-nature interactions are the natural starting point, reflecting as they do institutional failure, including communities' failure to come to grips with local resource allocation problems. But institutional failure includes government and market failures as well. Nonconvexities in the underlying ecological and economic processes imply that small initial differences in the performance of institutions across space, arising, say, from small differences in the costs of monitoring each another's actions, can lead to large differences in economic consequences over time. Nonconvexities also mean that small differences in the local ecology can lead to growing differences in the economic prospects facing societies.

As the pathways Dasgupta presents involve thresholds (the basis of several of the positive feedback mechanisms), the required resources need to come in lumps: small amounts at a time are pretty much useless. The currently rich countries, which once too were poor, managed to mobilize such resources in earlier times. A number of countries that were poor until recently also managed to do so. But there are many others that simply cannot afford such resources.

The author notes that systems characterized by positive feedback often possess multiple basins of attraction. Thus the countries now caught in a poverty trap possibly once faced good economic prospects, but are where they are because of ill-judged policies and bad institutions. Getting those economies out of the trap may require external help even if the institutions were to improve and the economic policies now chosen were sound. By outlining the theory of poverty traps, he provides an intellectual reconciliation between those who insist that the poorest countries of the world are where they are now because of their own fault and those who plead that these countries require external help if they are to lift themselves out of poverty.

Justin Yifu Lin and **Mingxing Liu** suggest that developing countries' adoption of an inappropriate development strategy can largely explain their poor growth performance and many institutional distortions after World War II. Motivated by nation building, most developing countries, including the socialist countries, adopted a comparative advantage defying (CAD) strategy to accelerate the growth of their capital-intensive, advanced sectors. Many firms in these prioritized sectors were not viable in open, competitive markets because of the violation of the economies' comparative advantages. To implement a CAD strategy, governments in the developing countries adopted a series of distortions in input and output markets to subsidize and/or protect nonviable firms, resulting in rent-seeking, soft budget constraints, macroeconomic instability, and income disparities. Economic stagnation, or even sudden collapse, becomes unavoidable under such circumstances, prompting the developing countries, voluntarily or involuntarily, to adopt market-oriented reforms.

Lin and Liu argue that to achieve dynamic growth, when the reforms begin or the transition starts, the government should liberalize entry into the labor-intensive sectors, which were previously repressed, and create conditions to address the viability of those firms in the CAD strategy's priority sectors.

Participation, Inclusion, and Results

Jean-Philippe Platteau presents the advantages and disadvantages of community-based development (CBD). The most important advantage usually associated with CBD programs is the informational gains arising from the proximity of local decisionmaking bodies to the target populations. However, one serious potential difficulty lies in such programs' vulnerability to capture by local elites. The author examines these two aspects of CBD programs critically in light of economic theory and principles and reviews the available empirical evidence. On the one hand, the informational advantages of the CBD approach are not always as decisive as they might seem. On the other hand, the elite capture problem is not more serious in a decentralized than in a centralized approach. Political economy models lead to indeterminate conclusions and have to be content with identifying factors that are more or less conducive to effective decentralized development.

Platteau then proceeds by discussing the possibility of mitigating the elite capture problem through various reputation mechanisms, both multilateral and bilateral. He illustrates the problems raised by multilateral mechanisms with a special focus on aid that foreign donors disburse to local communities or groups. He then turns his attention to a bilateral mechanism for disciplining local leaders that relies on a sequential disbursement procedure supported by fraud detection technology.

One of the conclusions reached on the basis of such analysis is that too quick and too massive a rush to CBD may prove self-defeating in the sense that the share of aid resources actually reaching the poor will be low if donor agencies are impatient to achieve results. Furthermore, stiff competition among foreign donor agencies engaged in CBD is likely to yield the same perverse result, not only because competition tends to make the reallocation of aid funds more costly in the event of detected project failure (in the sense of elite capture), but also because of the presence of careless agencies that, for various reasons, do not implement the sort of sequential disbursement mechanism envisaged.

Benno J. Ndulu discusses the increasingly important area of inquiry that explores the linkages among aid, macroeconomic policy, politics, and the micro concerns of governance and project design in the African context, emphasizing the need to understand the behavior of governments as aid intermediaries. He argues that one of the main reasons for elusive growth and development in Sub-Saharan Africa is the absence or weakness of citizen voice to engender restraint against predatory behavior and to promote accountability. While markets create managerial discipline and induce efficacy through the exercise of private choice, governments are principally disciplined through the exercise of voice to enforce representative public choice. In the absence of effective domestic mechanisms for restraint, weak autocratic regimes in Sub-Saharan Africa sacrifice overall growth for the sake of transfers to themselves and the narrow elites they serve.

Ndulu starts by reviewing the vast research on aid effectiveness, focusing on how the design and delivery of aid has tended to blunt the development of domestic accountability systems in recipient countries. He then lays out a conceptual

framework that links inclusiveness and accountability to the quality of policy choice and results through the exercise of domestic restraints against predatory behavior and the promotion of accountable behavior among those working in the public sector. Economy-wide and project-level empirical evidence reviewed in the paper broadly supports the association among inclusiveness, accountability, and results. Ndulu concludes by emphasizing that the next big push for the development agenda in Sub-Saharan African is improved governance and presents the key challenges to reforming aid principles and practice to support the reform of governance.

Scaling Up and Evaluation

Orazio P. Attanasio, Costas Meghir, and **Miguel Székely** present evidence from the evaluation of PROGRESA (from the Spanish acronym for Program for Health, Education, and Nutrition), a large welfare program in Mexico, by using randomized experiments and structural models for scaling up. The authors discuss the issues involved in scaling up at length. In particular, they discuss the relative merits of (a) nonparametric evaluation strategies that rely on (possibly experimental) exogenous variation to estimate the impact effects and (b) more structural approaches. The difference between the two approaches is particularly relevant in relation to extrapolation and scaling up. In principle, one could consider two types of extrapolation and scaling up. First, one might want to predict the effects of a program that is different from the one that was evaluated or the effect of changing some aspects of the program evaluated. Second, one might want to predict the effect of exporting an existing program from a population where its effects were evaluated to a different population.

Attanasio, Meghir, and Székely focus on the latter problem. After extensively considering the conceptual and technical issues involved in this type of exercise, they apply the ideas they discuss to the results from the evaluation of PROGRESA, for which a randomized evaluation sample is available that has been studied extensively. In particular, they divide the seven Mexican states in which the evaluation was carried out into two groups and review the extent to which results in one group can be extrapolated to the other. The advantage of such a strategy is that one can compare the extrapolation results (based on a structural model) with the actual ex post evaluation that can be carried out either by a simple comparison of means or by structural methods.

Esther Duflo discusses the key role that impact evaluations should play in scaling up. Credible impact evaluations are needed to ensure that the most effective programs are scaled up at the national or international levels. Scaling up is possible only if one can demonstrate that programs that have been successful on a small scale would work in other contexts. Therefore the very objective of scaling up implies that learning from experience is possible.

Because programs that have been shown to be successful can be replicated in other countries while unsuccessful programs can be abandoned, impact evaluations are

international public goods, and thus the international agencies should have a key role in promoting and financing them. In doing so they would achieve three important objectives, namely: improve the rates of return to the programs they support, improve the rates of return to the programs other policymakers support by providing evidence for the basis on which programs can be selected, and build long-term support for international aid and development by means of the ability to credibly signal what programs work and what programs do not work.

Duflo makes the case that the best way to estimate the impact of a broad class of development programs is through a randomized impact evaluation. She first discusses the methodology of randomized evaluation using several concrete examples drawn mostly from India. She then reviews the shortcoming of current evaluation practices before addressing the most frequent criticisms against randomized evaluation. Finally, she makes some suggestions about the role that international organizations could play in coordinating the accumulation of international knowledge on development projects: defining priorities for evaluation, funding and conducting randomized evaluation through an independent unit, working with partners, and creating a database of evaluation results.

* * *

Opening Address

S. M. KRISHNA

Dear Sri Jaswant Singhji, finance minister, government of India; Nicholas Stern, senior vice president and chief economist, the World Bank; Suman Bery, National Council for Applied Economics Research; distinguished invitees; and ladies and gentlemen: I extend a warm welcome to all of you on the occasion of the World Bank's Annual Bank Conference on Development Economics. I learn from the World Bank country director that the conference has been held annually in Washington, D.C., since 1989 and that, for the past four years, a similar conference has been held each year in Europe. It is perhaps in some measure a recognition of the contribution of developing countries, and among them, the contribution made by Indian economists, to the discourse on development that this conference is being held in a developing country for the first time. It also perhaps reflects the rapid strides taken by Karnataka in economic development and fiscal reforms and Bangalore's emergence as the knowledge capital of India.

I am happy to note that the theme of this conference is accelerating development. The main topics, fostering entrepreneurship, innovation, and growth; challenges of development in lagging regions; participation, inclusion, and results; and scaling up and evaluation, will, I am sure, do justice to the theme. I note that the focus has shifted from "developing" to "accelerating." This encourages me to add that acceleration should ensure that there is participation and inclusion, that no region is found lagging, that no section of society is left behind. It then follows that development should encompass equity and justice. Given that the paradigm shift in macroeconomic policy; the rapid developments in technology; and the compulsions of a world increasingly defined by liberalization, privatization, and globalization have exposed us to new opportunities and unprecedented threats, I draw the attention of this erudite gathering to the challenge of development with equity and justice.

I am happy to note that the achievements of Karnataka in public service delivery will be discussed in the concluding session. In this area, our theme has been to reach the public in the widest manner possible, and our motto has been to provide service through delivery mechanisms that harness available technology and innovative potential. Karnataka achieved a growth rate of 7.1 percent in the Ninth Plan, compared with 5.5 percent achieved at the all-India level.

I constituted the independent Administrative Reforms Commission for downsizing the bureaucracy and to redefine its role. We believe that a public-private partnership provides the best delivery system in moving toward a market-oriented economy. We were perhaps the first state in the country to enact the Right to Information Act. We have also set in motion the process of all public utilities benchmarking best standards in the delivery of services through citizens' charters. All the 2 crore (20 million) land records of nearly 70 lakh (7 million) land owners in 175 taluks (revenue districts) are now fully computerized under the Bhoomi project. The state government has set up the Bangalore Agenda Task Force to consider ways to upgrade the city's infrastructure system by pooling resources for development by involving the corporate sector, industries, institutions, and the general public. We have also set up a revolving market intervention fund with principal of Rs 2 billion and a medical insurance scheme for farmers called Yashaswini, and the distribution of Kisan credit cards (a short-term credit program for farmers) has been accorded top priority. We have embarked upon a comprehensive economic restructuring and fiscal reform program that, in due course, will enable us to provide services more effectively. Yet the vigor and tempo of reforms will have to be tempered by the constraints resulting from the structural inadequacies and historical inequities we continue to endure. This, I assume, is true of most developing countries.

Building Human Capital for Economic Development

AZIM HASHAM PREMJI

Ladies and gentlemen: It is my pleasure to be here with all of you at this Annual Bank Conference on Development Economics. I thank the World Bank for this opportunity.

My view is that the world in general, and India in particular, is poised at a critical phase of development that has the potential to determine the way things are going to shape up over the next 70 years.

It was Adam Smith, who is regarded as the father of economics, who said that every nation's national product is regulated by two different circumstances: first, by the skill, dexterity, and judgment with which its labor is generally applied; and second, by the ratio of those who are employed in useful labor to those who are not so employed. Today the world is far more developed than in the days of Adam Smith, but the fundamental principles of economic development have remained the same.

The key differentiator between rich and poor countries remains their ability to unleash the power of their populations and to invest it in economic activity as human capital. The difference in the composition of wealth between these groups of nations underscores this point: 75 percent of the national wealth of Western Europe is in the form of its human resources, 23 percent in the form of industrial products, and 2 percent in the form of natural resources. For West Africa, the figures are 60, 19, and 21 percent, respectively.

Clearly all the economies that have prospered in the last century have done so because of their ability to tap human potential. Economic theories of growth have captured this idea adequately in the last two decades. Endogenous growth theory, often referred to as new growth theory, developed by Nobel Laureates Robert Lucas and Paul Romer in late 1980s, clearly established that growth cannot come solely from the serendipities of technological innovations but is driven by systematic social choices. These choices primarily include investment in technological change and

activities to enhance human capital, such as education and health care. Thus if we want to accelerate economic development, we have to look carefully at these social choices and ensure that developing societies in particular make appropriate choices.

Human development is about people, about expanding their choices to lead lives they value. Fundamental to enlarging human choices is building human capabilities: the range of things that people can do or be. Human development is certainly a broader and more important goal than mere economic development. That is why we measure development using such tools as the Human Development Index instead of simply average national income. This enables us to include important aspects of development, like longevity and knowledge, in the development framework along with the level of economic well-being.

So the all-important question, even more important than economic development itself, is whether human development in a given area is improving or not. While clear progress has taken place in some areas, the frightening issue is the continuing disparity between the industrial and the developing countries. A girl born in Japan today has a 50 percent chance of seeing the 22nd century, while a newborn in Afghanistan has a 25 percent chance of dying before age five. At the same time, the richest 5 percent of the world's people have incomes 114 times those of the poorest 5 percent. In Sub-Saharan Africa, human development has actually regressed in recent years, and the lives of its poorest people are getting worse.

Turning to India, the scenario is surely encouraging. We in India have always believed that our country is rich in culture, values, and people power. The world has now acknowledged us as a storehouse of talent, primarily because of India's contributions in the areas of information technology and leadership talent, particularly in the last two decades. The question is, how do we build talent for our own overall development? How do we create a sense of accountability to deliver commitments to the electorate, to customers, to our children, and to the law-abiding citizens of this country? How do we reach global standards of excellence in quality, productivity, and efficiency in the shortest possible time? How do we disseminate and adapt the best practices of some of our corporate organizations for use in all aspects of public life?

From our experience in Wipro, we believe that achieving these goals needs meticulous planning, willpower, commitment by those at the highest levels, and consistent pursuit of an approach that stresses the development of people. If you consider Wipro to be a reasonably successful organization and ask me the reasons for this success, without doubt my answer will be "our people."

What did we do to develop people in a strategic manner? To start with, as a leader of the organization, I personally made a commitment back in 1969 that we would develop leadership by recruiting people from management campuses. We were among the first Indian companies to do so. The main attribute that we were looking for in the people we recruited was whether their values matched our company's values.

Next, at a time when speaking of beliefs and values was not fashionable, we developed our organization's beliefs in 1972, articulated them to the rank and file in the organization, and most important, practiced them almost fanatically. We soon found that the talent we recruited immensely enjoyed working in a value-based organization that stood up for its values.

We took numerous actions to nurture the young talent in the organization. We gave people huge responsibilities early in life; trusted them fully; allowed them to make mistakes; promoted a professional, open, and informal work environment; and even compromised on short-term results as long as people exhibited the potential to develop. We were the first organization in the country to provide stock options, which we did as early as 1983. Our phenomenal growth in sales, profits, and businesses stimulated our employees to continue to deliver their best and built their pride in working with a high-performing organization. They had to cope with the intellectual challenges that brought out the best in them. We obviously had to match our business growth with the best of people practices, to adopt newer approaches to learning and development, and to promote a globally competitive work environment that increased the likelihood that our staff would choose Wipro over other employers.

My sincere belief is that the development and deployment of the right talent in the public and private sectors can soon catapult India to that long awaited status of being an industrial economy. To this end, both Wipro and the Azim Premji Foundation have a single-minded focus on significantly enhancing the quality of learning in Indian elementary schools. We are working in partnership with the government, and our experience has been highly encouraging. We work with the government of Karnataka and with Andhra Pradesh bureaucrats and politicians who exhibit maturity and an in-depth understanding of the relevant issues in education. The talent that exists in the government is enormous and is on a par with—if not better than—that found in the corporate world. However, we find that the need to develop world-class talent at the grassroots level is enormous. Also needed are administrative reforms that promote professional, merit-based employment practices that develop accountability in relation to end users or customers.

The reason we decided to focus on elementary education was our belief that a sound elementary education with high-quality learning could form a solid foundation for India's development. Whenever the discussion on education begins in policy circles, the focus is often on budgetary provisions, that is, the need to allocate a certain percentage of gross domestic product to education. I am not denying the importance of financial resources for education, but based on our experience to date, we have realized that the key challenge is developing competent teachers and education functionaries who consider learning by the children to be the number one priority in their lives.

This problem is unlikely to be solved merely by allocating more money. What is required is more focused efforts and imagination. If we are to make major headway, we have to stop thinking in terms of better school buildings and start thinking in terms of passionate, highly motivated, and, more important, highly competent teachers and principals. We have to think of the economic scenario that awaits us 50 years hence and the kind of talent we require to meet that scenario successfully. We need to visualize the qualities and attributes of the kinds of people we want education to produce. We need to enrich our curriculum to incorporate those attributes and competencies that the current curriculum lacks. We need to develop methods of assessment and evaluation that can help us identify whether the desired attributes have indeed been developed. In addition to the assessment-oriented evaluation, we also need development-oriented evaluation. We are actively working with the government

to introduce the concept of providing a so-called learning guarantee in the schools. If we are to achieve human development on a continuous basis, what we need is a continuously learning individual, a continuously learning nation, and a continuously learning world.

Most of the world's business people realize that the material investments they make will depreciate over time and will have to be written off. Only one kind of investment will keep appreciating irrespective of the interest rate in the economy and the economic climate: investment in human capital.

In India, we have islands of excellence in institutions and corporate organizations that have achieved world-class standards. The challenge before us is to develop human capital at the grassroots level that is deployable in a globalized economy, to develop people at every level who interact with the general public and deliver services and products to them.

Investments made in education and human development have cascading, long-term benefits. Making such investments is the only way to accelerate economic growth and development for the entire world economy.

Keynote Address

Opportunities for India in a Changing World

NICHOLAS STERN

It is a great pleasure to be back in Bangalore and to be holding the Annual Bank Conference on Development Economics here in India. This is the first time that this annual conference has been held in a developing country. We are here to learn from and share ideas with scholars from around the world, and particularly from South Asia, on the important question of how to accelerate development, the theme of this conference. The World Bank chose this theme in light of the ambitious development goals the international community has agreed on. If they are to be achieved, we must accelerate development. We have much to learn from countries that have accelerated their development and made strong progress in reducing poverty. We can also draw important lessons from the experience of lagging regions in the world and within countries, which is the focus of one of our conference sessions.

Much of my talk will center on the experience of India, which has contributed an extraordinary amount to our understanding of development and has demonstrated the positive impact that reforms can have on accelerating growth and reducing poverty. My choice to focus on lessons from India, however, goes beyond the importance of India to the world and our location in Bangalore—it is also personal. I am fortunate to have had frequent opportunities to visit and learn from India in the past 30 years. This experience and my interactions with my many Indian friends have in large measure shaped my views on development.

Lessons for Development in a Changing World

The world has changed dramatically in the last two decades, transforming how we look at economic development, at interdependence, at opportunity, and at risks. At

the end of the 1980s, the Berlin Wall came down and 400 million people of Central Europe and the former Soviet Union set off on an unprecedented transition from communism and a command economy to, for most, democracy and a market economy. There have been achievements and trauma, but the changes are irreversible, and the European Union is now welcoming countries previously separated from Europe by the Iron Curtain. We have also witnessed the dramatic rise of China on the world economic stage following the reforms initiated by Deng Xiaoping at the end of the 1970s. More quietly, in the early 1990s, India set off on the road to economic change and is now seeing growth and poverty reduction that few would have thought possible 20 years ago. The rising economic power of China and India is already beginning to transform international economic and political relations.

Technology has continued to change rapidly, with the falling costs of communications and transport transforming information flows and interactions and further propelling rapid growth in services trade. This includes the international outsourcing of many accounting and information services and the movement of people. These trends are likely to continue, presenting new opportunities and posing new challenges. Trade integration and growth are likely to intensify, particularly if the World Trade Organization (WTO) Doha round of trade talks launched in 2001 is successful, as I hope it will be. Technological change will remain rapid and will lead to emerging industries that are hard to anticipate or imagine now. In many countries, economic growth is likely to be highest in services, continuing the transformation of production and employment toward the service economy and increasing the role of services in international trade.

Integration, rapid growth, and intensive use of common resources, however, can also bring risks as well as opportunities. Pandemic events have unfolded that have faced few barriers to their spread, including AIDS and financial contagion. Public health may face major threats from new drug resistant bacteria and viruses, while the continuing rise of transborder criminality affects security. Many, not least in India, have associated increased global integration with an assault on local culture and social values from a vulgar, lowest common denominator of loud music and fast foods or from dubious behavioral standards. By contrast, many of us in Paris, London, Washington, and elsewhere have seen our culture and opportunities enriched by the arrival of music, philosophy, and cuisine from the subcontinent and from other parts of the developing world.

Major demographic trends within countries, such as urbanization and the almost universal aging of populations, are likely to have a huge societal impact with a potential for conflict between competing social groups. At the same time, fertility rates are declining in most developing countries, and the dependency ratio (the ratio of children and the elderly to the working population) is projected to fall over the next three to four decades in Africa, Latin America, and South Asia. This opens up a window of opportunity for spending more on physical investment before the ratio begins to rise as the number of old people increases.

To take advantage of and manage these enormous changes, economies and societies will need to have sufficient flexibility to learn and to adjust. Governments, the private sector, and civil society will need to look ahead to identify opportunities and

be ready to learn and adapt. Those who manage this process well are likely to flourish in this changing world; those who do not may find themselves marginalized.

A useful approach to guide strategy in this rapidly changing environment is to draw on the key lessons learned from development experience over the past several decades and distill those lessons that can help guide future policy. In this context, I would like to discuss three areas of change affecting opportunities for developing countries: changes in ideas, in resource flows, and in partnerships. Later I will return to the specifics of India's development strategy.

Changing Ideas: What Lessons Can We Draw from Five Decades of Development?

Before describing the key lessons of recent decades, we should be clear that development cannot sensibly be understood simply in terms of the growth of aggregates, such as income per capita. Development is about fundamental change in economic and social structures. It is about the movement out of agriculture into services and industry, about migration to cities and peri-urban areas, and about transformations in trade and technology. Change—in health and life expectancy, in education and literacy, in population size and structure, in gender relations, and in social relations—is at the heart of the story. The challenge to policy is to understand, release, and guide the forces of change.

Thus, what lessons can we draw from five decades of development? At the risk of some oversimplification, let me express what we have learned in terms of the following six lessons:

- *The objectives of development have changed.* Economists' perspectives on the goals of development have changed substantially in the last 20 years. We now look beyond incomes to health and education. Indeed, we now look beyond these simple, basic elements of human development and see the objectives of development as concerning people's ability to shape their own lives.

- *The state and markets complement each other.* The state is not a substitute for markets but a critical complement to markets. We have learned that markets need government and government needs markets, and that government action is crucial to people's ability to participate in and help shape economic opportunity. These lessons point to the need for an active state that fosters an environment where contracts are honored and markets can function, which requires infrastructure; financial systems; sound administrative, legal, and regulatory structures; and defense. But the role of the state should go way beyond this to fostering health; education; and the social, political, and physical environment, which together determine how people are able to participate in decisions that affect their lives.

- *The most powerful force for the reduction of income poverty is economic growth.* Three points are relevant here. First, the countries that have been most effective in reducing income poverty are those that have grown the fastest, while countries

that have stagnated or fallen back economically have seen the greatest expansion in poverty. Second, the private sector is the strongest, and indeed the driving force, behind sustained economic growth. Third, international trade has been a crucial engine of growth. Trade patterns have changed dramatically since the 1970s. Many countries have shifted away from import-substituting strategies, and developing countries have exhibited a changing pattern of exports. In 1980, only 25 percent of the exports of developing countries were manufactures; by 1998 this had risen to 80 percent, most of which were labor intensive. East Asia (particularly China) and South Asia (particularly India) have been at the forefront of this story.

- *The role played by institutions and governance in the growth process is central.* Development research in the 1990s and, in particular, the recent experience of former socialist countries in transition to market economies, has taught us that institutions lie at the heart of the development process. Institutional failure has been at the core of many environmental problems, and as last year's *World Development Report* (World Bank 2003) argues, building institutions must be central to the solution. Social cohesion is also an important foundation for sound policies and institutions, and its lack can be devastating (for example, the lawlessness and the stealing of state assets that went on in some of the countries of the former Soviet Union in the 1990s). The most severe form of breakdown of the rule of law and governance is civil war. The World Bank's recent research report on conflict (Collier and others 2003) sets out evidence showing the devastating effects of conflict on development. By the end of the typical civil war, average incomes are around 15 percent lower than they would otherwise have been and about 30 percent more people are living in absolute poverty.

- *The empowerment of people is important.* Individuals are empowered if they have the ability to shape the basic elements of their own lives. Thus empowerment requires that people, regardless of gender, be educated and healthy, or, in other words, that investment in their human capital takes place. But empowerment goes beyond human capital. It also means effective participation, which in turn depends on information, accountability, and the quality of local organizations. We have learned that development works better where poor people are empowered. For example, schools function better if the community is involved; infrastructure, electricity, water, and the like work more effectively if consumers' voices are heard; and poor people are more productive and creative in their economic lives if they have clear title to their property. Thus effective participation and social inclusion mean better results.

- *A link exists between country ownership and development effectiveness.* Aid can provide critical support to communities and countries in which a genuine movement for change is present, but the country must be "in the driver's seat" if reform programs are to be sustained and to succeed. Ownership means that the country itself embraces and shapes the program. Reform programs forced on countries from outside, with weak societal commitment, are likely to fail. The ownership of

the development agenda by a country and society is a vital ingredient for its effective implementation.

Changing Resource Flows: The Need to Strengthen Flows to Developing Countries

As I have argued, we have learned a great deal about the drivers of development, and developing countries have increasingly been applying the lessons with visible results in terms of stronger outcomes, as well as improvements in policies and institutions. These lessons are increasingly shaping the work of those who are trying to support developing countries in their attempts to promote development, with the result that aid effectiveness has improved. A number of factors are at work. First is the improvement in policies of the developing countries themselves. For example, in many developing countries, better macroeconomic policies and greater openness have made the domestic economic environment more conducive to growth. Second is a stronger understanding of the role of effective governance in facilitating the provision of physical infrastructure and investing in building the institutional capacity essential to sustained development efforts. Third is the improvement in aid allocation following the end of the Cold War and the consequent decline in aid for political purposes. Aid is increasingly allocated on the basis of where poor people are located and where it can be used effectively. Fourth is the greater emphasis being placed on the quality of a country's overall public expenditure programs and the reduced weight assigned to individual projects' rates of return.

This new emphasis on the policy environment in developing countries as a determinant of aid allocation is based on evidence that the quality of the policy environment and overall expenditure allocations are crucial to the productive use of funds, and, furthermore, that project-specific aid is fungible. The development community has also come to recognize that traditional conditionality, whereby aid is conditional on promises of policy reform, is of limited effectiveness, reflecting the fifth and sixth lessons presented in the previous subsection. This has resulted in greater emphasis on local ownership and jointly monitorable results.

The lessons concerning development and development assistance are increasingly widely shared and helped shape the new partnership and commitment that emerged from the 2002 Monterrey conference, which focused on the need to increase foreign assistance flows and open up market access to developing countries. There is an understanding that it is within our collective power to achieve the Millennium Development Goals (MDGs) embraced by world leaders in New York in 2000 and reaffirmed in Monterrey, *if the international community has the will to do so*.[1] However, much stronger resource transfers, both in terms of quantity and quality, will be required to meet the MDGs. The real level of foreign aid has fallen from about 0.50 percent of rich countries' gross domestic product (GDP) in the early 1960s to only 0.22 percent of their GDP today. The peace dividend from the end of the Cold War has not materialized in the form of higher assistance flows to developing countries.

Announced commitments at and around the time of the Monterrey conference of a further US$16 billion a year by 2006 can help reverse the downward trend. The World Bank estimates that an effective increase of several billion dollars more could come from improved quality, including harmonizing aid to a greater extent, increasing the share of investment, and untying aid from donor restrictions. Taken together, these steps would provide a significant contribution toward meeting the MDGs, but the amount would still be far below the estimated doubling of aid that is needed to achieve the MDGs according to estimates by the World Bank and others. From a historical perspective, however, a doubling of aid is far from inconceivable.

I believe that India has an important role to play in demonstrating the effective use of aid and in supporting the call for additional resource flows from the rich countries. External assistance has a good track record in India. Indian leadership will exert a major influence, as well as have a strong demonstration effect, in relation to mobilizing and effectively using aid to advance social and poverty reduction objectives.

Changing Partnerships: The Challenge of Doha and of Commitment

Finally, changing international partnerships are also affecting the dialogue and the opportunities for development. The Monterrey consensus on measures to achieve the MDGs represented a watershed. It recognized the importance of collaboration by and commitment among industrial and developing countries and the obligations of both in achieving targets set out in the MDGs. A new partnership was understood and embraced at Monterrey that aims to deliver on these goals and commitments.

Developing countries committed themselves to reform their institutions, governance, and policies so as to drive faster growth and development. These reforms have two main thrusts: a better investment climate for growth, productivity, and jobs, and more effective investments in people through the delivery of such core public services as education and health that enable poor people to participate in the decisions that shape their lives. While the policy, institutional, and governance challenges are severe, the developing countries are making strong—and in many cases successful—efforts to deliver on their side of the bargain.

The obligation of rich countries requires easing trade restrictions to increase access for developing country products to their markets, increasing aid flows to countries engaged in fundamental reforms, providing technical support for capacity building, and providing support for enhancing capital flows to developing countries. We have seen that where the policy environment in developing countries is healthy and such countries have access to industrial country markets, aid flows can be highly effective. We must recognize, however, that the value of aid flows is not only, or even primarily, as resource transfers. Aid flows represent less than 1 percent of GDP of the developing countries. Their value lies, in particular, in promoting and supporting the reforms of developing countries that will determine their success in the fight against poverty. Thus *aid must be used to finance the costs of change, not the costs of not changing.*

We must recognize that sometimes in the past, such as in Mobutu's Zaire, aid has performed the latter function rather than the former. Being a good partner on the aid

front means not only increasing aid but also allocating it in a supportive, user-friendly, and timely fashion. The potential returns to a successful Doha round for world trade and expanded market access for developing country exports are enormous. Global gains from the removal of tariffs could reach around US$800 billion per year by 2015, with the lion's share going to developing countries.[2] Removing barriers to trade could bring an additional 300 million people out of poverty by 2015. It is difficult to imagine something that would give a stronger boost to a fragile world economy than a successful outcome of the current WTO discussions.

The developing countries are now looking for movement on liberalizing agricultural trade, which is enormously important to their growth opportunities, as the critical test of the rich countries' commitment. Movement at a sufficient scale and speed has yet to materialize. Visible movement on agriculture, on the Trade-Related Agreement on Intellectual Property Rights, and on medicines will be needed to move the agenda forward. A great responsibility for success lies with rich country leaders.

This agenda for change—in ideas, resources, and partnerships—places developing countries at center stage of the development process. The partnership entails obligations by high-income and developing countries. The opportunities that rich countries allow will help to determine how developing countries respond. The challenge of Doha is a call for leadership for change. As a leading democracy and a major developing country with a track record in making development happen, India has an especially important role to play in making Doha work. India can be a still stronger advocate for delivering on Doha by continuing to remove barriers to trade unilaterally. Indeed, in the 1990s India demonstrated that a substantial boost to growth and poverty reduction can result from trade reform. It is therefore in a special position to provide leadership and speak up for developing countries as a whole.

Opportunities for India

I turn now to how the lessons I have highlighted and the changes and opportunities I have discussed affect India. India has made major contributions to development thinking and action, as demonstrated by the five-year plans that, starting in 1950, have both tracked and shaped world thinking on development. Reading India's five-year plans is not unlike reading a history of the debates about development.

I have been fortunate to have witnessed important parts of this history. In 1974–75, I was based in the Planning Commission with my host institution, the Indian Statistical Institute, working with T. N. Srinivasan, who is here today. At that time the institute housed some of the major contributors to, and critics of, Indian planning. Twenty years ago S. P. Gupta invited me to help organize the 1982 conference, held here in Bangalore, that launched the Sixth Plan.

Enormous changes have taken place in India's and the world's thinking about development since the Second Plan, the so-called Mahalanobis-Nehru plan of 1955–59. Influenced by what was seen as successful planning in the former Soviet Union since the 1920s and in the United Kingdom during World War II, the Second Plan reflected a deep mistrust of markets. The vision was of a command economy led

by heavy industry. This was understandable at the time. The worldwide depression of the 1930s, together with the publication of Keynes' General Theory, showed how markets could fail on a massive, and sometimes catastrophic, scale. The Second Plan responded with state planning and input-output programming, in its early forms, at center stage. The role of the private sector was extremely limited.

The role of markets and the importance of services had gained much greater acceptance by the Sixth Plan (1980–85). Government failure was seen as a serious issue and discussed together with market failure. Planning was less mechanistic and was more sophisticated in its approach toward indirect instruments for guiding the economy. Yet the Sixth Plan still saw the central government playing the major role in the allocation of resources, with public investment being the key focus of the planning document.

By the Tenth Plan (2002–07) the perspective was completely different and the private sector is now at center stage. The plan seeks to tap the dynamism of India's private sector, which has proven its ability to deliver rapid growth (for example, in software and services exports) if it is given the chance. The plan sees the role of the state as complementing markets by concentrating on providing the right environment, that is, one in which people (especially poor people) are enabled to participate.

India's achievements in improving its growth performance and reducing poverty have been impressive in recent years. Those of us who are longtime friends of India got used to hearing about a "Hindu" rate of growth of 3.5 percent total and 1.5 to 2.0 percent per capita. That has changed. For much of the 1980s and 1990s, per capita growth rates more than doubled as a result of more market-friendly and open policies, particularly since the 1991 reforms led by Manmohan Singh. The outcome has been a strong drop in the proportion of people living in absolute poverty: 5 to 10 percentage points over the second half of the 1990s.

Nevertheless, India still confronts major and difficult challenges. The Tenth Plan has set an ambitious growth target of 8 percent per year through 2006–07, well above past growth rates and at a time of slowing growth relative to the good performance of the mid-1990s. Related to this are equally ambitious targets in relation to social indicators. I will focus on three of the major issues confronting India in achieving these goals that are attracting growing attention: public finance, trade liberalization, and the investment climate. International development experience offers clear lessons in each area.[3] Progress on these issues would make a major contribution to accelerating development in lagging regions and rural areas, a major theme of this conference, but alone they will not be sufficient. In addition, special efforts will be needed to include states, sectors, and communities that are being left out of the Indian story of growth and development.

Public Finance

Weakness in public finance runs through many of the problems confronting the Indian economy today. It has contributed to the poor public infrastructure, the weak public health and education systems, the ineffective social protection systems, and an

often inhospitable investment climate. The problems of public finance include both quantitative, macro dimensions, as well as qualitative and more micro dimensions related to public service delivery and the effective use of public funds.

International experience demonstrates the importance of sound public finances, stable macroeconomic management, and efficient targeting of resources for promoting economic growth. Indeed, one of the key lessons of international experience is that fostering markets by ensuring a stable currency, stable interest rates, and healthy and competitive financial markets is a key aspect of the state's role in promoting development.

Over the past six years, India's consolidated government fiscal deficit has been consistently high, and general government debt has increased significantly. India has been able to live with high deficits by relying on domestic financing through the banking system, strong growth performance, and capital controls. External reserves are strong, the current account is in surplus, food stocks are abundant, and inflation and interest rates are at all time lows. But how much longer can this continue without increasing India's economic vulnerability to shocks and undermining its growth performance? This question is currently being actively debated in India.

A look at some economic and social outcomes suggests that weak public finances are already exacting a toll. Of particular concern is the rising revenue deficit, which more than doubled between 1995–96 and 2002–03.[4] General government revenues fell by 2 percent of GDP between the late 1980s and the Ninth Plan period (1997–2002), even as expenditures on salaries, pensions, subsidies, interest payments, administration, and defense increased. Meanwhile, capital expenditure fell by 3 percent of GDP. In such a situation, can we say that scarce public resources are being spent in a way that is conducive to the achievement of national development goals? Development spending on health and education has stagnated, while spending on infrastructure has fallen. Expenditure on salaries is crowding out nonwage operations and maintenance. These trends compromise national growth and human development.[5] The deterioration of fiscal health in the past decade has led to the running down of the capital stock and a reduction in the quality of core public services at the central and state levels. The high fiscal deficit is only one aspect of this problem.

The fall in revenue as a percentage of GDP and the deteriorating composition of expenditure are equally worrying. The shift in public spending away from investment and infrastructure toward recurrent expenditure inhibits private investment. The states have played a major role in this, as growth in establishment costs has outpaced revenues, squeezing out investment and the maintenance of existing infrastructure. The Tenth Plan quite rightly calls for a major fiscal correction.

Improving India's fiscal health will call for a range of reforms, including the following:

- International experience suggests that the framework for fiscal management needs to be strengthened. This is especially important in large countries such as India, where the fiscal problem must be addressed at both the central and state levels. The fiscal responsibility acts several states have passed, including Karnataka, Punjab, and Tamil Nadu, and the central government's recently passed Fiscal

Responsibility and Budget Management Bill, are welcome steps in this direction. In particular, the commitment to eliminate the revenue deficit by 2007–08 under the government's Fiscal Responsibility Act is encouraging.

- Tax reform will be required to broaden the tax base in terms of coverage, exemptions, and improved administration. Bringing services and agriculture into the tax net should be part of the reform. Tax administration is a key challenge, as discussed in depth in the recent Kelkar committee report (Kelkar 2002).

- Expenditure restructuring aimed at improving the quality and effectiveness of public service delivery is called for to address the quality of spending. Public expenditures should be scrutinized for their development impact. For example, is pricing resources such as water and electricity at less than their cost of production a better way of fostering growth and human development than focusing resources on health, education, and public infrastructure? Do food procurement and price controls do more to prevent poverty and malnutrition than well-designed programs targeting poor people? Are salary increases more important than nonsalary operations and maintenance expenditures on health and education? Radical reform of the "subsidy culture" is needed to cut nonmerit subsidies that largely fail to reach their targeted beneficiaries, incur huge opportunity costs, and distort factor prices while hurting the economy and the environment.

- A fundamental challenge to improving the government's ability to deliver public services effectively involves addressing the performance and incentive structure of the civil service. Improving the quality of spending includes finding ways to get teachers to show up for classes and to get health workers to focus on immunizing children rather than working in private clinics. Enhancing public access to information and strengthening accountability mechanisms will help. Empowering community organizations, such as village education committees, and reducing political interference in the civil service are part of the answer. Many states have shown that improved service delivery is possible: Karnataka is one of them, as the case studies to be presented later during this conference will show.

Trade Policy

One of the strong lessons of development is that trade accelerates growth and growth reduces poverty. India is one of many countries to demonstrate this. Trade reform, together with partial dismantling of the "permit raj" (a term coined to denote India's stifling bureaucratic red tape), played a major role in accelerating growth during the 1990s and boosting the ratio of trade to GDP. India doubled its share of exports in GDP in the 1990s while roughly halving tariffs. But a major agenda remains ahead, both for India and internationally, in relation to developing countries' potentially large gains from liberalization. Trade reform is a key part of improving the investment climate and taking advantage of growth opportunities. It increases competition and drives better performance, leading to greater innovation and more productive opportunities. The experiences of China and India provide testimony to the role of

trade in stimulating growth and poverty reduction, notwithstanding the barriers rich countries often erect. Four important ideas apply to the global trade arena of which India is a part, namely:

- The WTO must give priority to a huge market access agenda. Trade restrictions in both Organisation for Economic Co-operation and Development (OECD) and developing countries lower incomes in developing countries. Agriculture is key here: many barriers with tariff peaks, quotas, and nontariff barriers are vexatious and protectionist. OECD agricultural subsidies exceed US$300 billion a year, six times the size of its foreign assistance flows. They are costly to OECD taxpayers and, more importantly, impose a huge burden on farmers and rural households in developing countries. Progress in agriculture will be crucial to moving forward on the entire Doha agenda.

- The WTO is an important mechanism for negotiating the removal of policies that distort trade, but many of the gains from reform can be realized through unilateral action, which does not have to wait for WTO negotiations. India has demonstrated this in the past, and it applies equally to India's future.

- The reciprocity in the WTO negotiating process creates the potential for perverse incentives, that is, holding back on reforms that are good for development. This makes trade reform by OECD countries even more important: it will help the governments of developing countries pursue beneficial domestic reforms.

- The trade agenda extends far beyond the traditional trade policy issues of tariffs and quotas. "Behind-the-border" policies that affect the operation of infrastructure, service industries, and government bureaucracy are major determinants of competitiveness. Difficulties with quality standards and documentation requirements are examples of the need to reduce the soft infrastructure barriers, as well as the hard infrastructure barriers, such as ports and transport systems.

Turning to the trade policy agenda for India, four points should be emphasized, namely:

- While liberalization in the 1990s cut tariffs and achieved a great deal, India stands to benefit significantly from its own further trade liberalization. Its average tariffs of 35 percent remain too high.

- Both past and future liberalization is in danger of being eroded by other forms of protection, such as antidumping action. India has become the second largest user of antidumping in the world after the United States. The use of substitute instruments of protection needs to be disciplined.

- Policies that improve the investment climate are key to reaping the gains from trade. These will require behind-the-border policy changes to improve the power supply; to reduce bureaucratic burdens; and to lower the costs of transport, customs, and port clearance, to give just three examples. These will promote domestic investment and encourage further increases in foreign investment.

- Existing barriers to firm entry and exit must be the focus of particular policy attention. Such barriers reduce competitiveness, entrepreneurship, and the growth of productivity and thereby diminish India's ability to respond to new opportunities.

To this list let me add the importance of securing progress in the General Agreement on Trade in Services, in the temporary movement of service providers, and in cross-border trade in services, all of which are especially important to the emergence of Bangalore and other Indian cities as information technology centers. The movement of people is also important to India as a whole and internationally. International worker remittances to developing countries are about US$80 billion annually, well above the US$50 billion in annual aid flows.

Note that these arguments in favor of trade liberalization pertain equally to reducing internal barriers to trade in India. Reducing the cost of moving goods between states—transit taxes, harassment, and poor infrastructure—can yield large benefits. These costs are particularly detrimental to the lagging states and predominantly rural regions of India, where infrastructure and connectivity to markets are already weak.

Investment Climate

While trade liberalization is important for growth, it is not the whole story. An additional factor is the investment climate, that is, the qualities that determine a country's attractiveness to potential investors. The issue is not primarily attractiveness to foreign investment. While that is important, for large countries such as India, domestic investors will be by far the most important source of the investment that will drive growth and poverty reduction. Off-farm employment plays an increasing and critical role in generating incomes for the majority of poor people who live in rural areas in India.

As concerns the investment climate, two issues are of special importance in India: governance and institutions along with infrastructure. Governance and institutional capacity are critical elements of the investment climate. Bureaucratic harassment and poor service delivery, which discourage investment, are likely to be worse in rural areas than in cities, which tend to have more checks and balances on bureaucratic power. Differences across states underscore this point. A 10-state survey conducted jointly by the Confederation of Indian Industries and the World Bank (World Bank and Confederation of Indian Industries 2002) found that companies in Uttar Pradesh had to contend with visits from bureaucrats twice as often as firms in Maharashtra. Tackling governance will require building up local community organization and accountability mechanisms. It will also require pushing more aggressively to implement the 73rd and 74th constitutional amendments devolving power to local governments, while at the same time working to strengthen community self-help groups.

The importance of public infrastructure and market access—power, water, sanitation, roads, and transport—cannot be emphasized enough. Probably the single

greatest factor preventing farmers in Bihar or Orissa from achieving comparable productivity standards as those in Haryana or Punjab in agriculture and agri-industry processing is investment in rural infrastructure. Rural power, road connectivity, and water control are critical for introducing new technologies and for getting goods to the market. The deterioration in public expenditures, with declining expenditures on road maintenance, irrigation and water systems, and power over the past decade, reinforces barriers to off-farm employment. Again comparing Maharashtra and Uttar Pradesh, 90 percent of firms in Uttar Pradesh must rely on more costly, small-scale, private generators for electricity, whereas in Maharashtra fewer than half do so.

Let me finally say a few words about innovation and the investment climate. One cannot come to Bangalore without marveling at the tremendous success of the high-technology and software industries here. Many have agreed that this reflects the practical difficulties bureaucrats encounter in attempting to regulate new technology and service industries. The Tenth Plan (paragraph 1.41) argues this, noting that the service sectors have performed better than industry because they face fewer bureaucratic impediments.

Fostering innovation will be key to India's efforts to increase its market penetration internationally and expand the share of the manufacturing sector in the economy. India's manufacturing sector relative to GDP is less than half the size of China's and most Southeast Asian countries, and India's penetration of world markets in industrial products stagnated during the 1990s. A recent World Bank study (World Bank 2004) shows that starting a business in India requires much more time than in China, serving as a tax on entrepreneurship, impeding innovation, and raising the cost of bringing ideas to the market.

Lagging States, Sectors, and Communities

A central concern in the Tenth Plan is the growing disparities in growth performance and social outcomes across states. More than half of India's poor people live in just four states: Bihar, Madhya Pradesh, Orissa, and Uttar Pradesh. More than two-thirds of poor people are in rural areas and depend largely on agriculture. The highest incidence of poverty is found among people of scheduled castes and scheduled tribes, who face major social barriers that exclude them from opportunity. Women also face special barriers.

With the concentration of poor people in these areas and sectors, the problems poor people face are also intensified: lower access to, and lower quality of, public services; lack of leverage to obtain fair treatment from public institutions; and greater vulnerability to corruption and abuse (poor people pay proportionately more of their incomes to corrupt officials than do the wealthy).

Accelerating national growth, reducing poverty, and achieving national social targets will not be possible without strengthening the performance of, and access to, services in these lagging states, sectors, and communities. Each of the areas of reform discussed earlier—public finance, trade policy, and the investment climate—are

highly relevant to improving the development of lagging regions. The central government and the states themselves need to make special efforts to improve the allocation and use of public resources, to improve public service delivery, to lower chronic absenteeism among service providers, and to strengthen the investment climate for accelerated growth. These are highly complex problems for which no simple answers are available, but consider some ideas based on the foregoing analysis.

- Accelerating growth in the lagging regions is largely a question of boosting agricultural productivity and off-farm job opportunities. The investment climate in these states needs to be promoted—in particular to strengthen governance and institutions—in order to reduce the obstacles to investment. Removing or reducing central government food grain subsidies and retargeting these resources toward productivity-enhancing infrastructure, irrigation, and technology upgrading could help in this process. While some states that benefit most from these subsidies would lose, the lagging regions have benefited the least.

- Strengthening the physical infrastructure and market connectivity of lagging regions will require more resources and increased efficiency in investment practices. Greater flexibility in the design and use of centrally supported investment programs is one measure that could improve investment effectiveness, combined with technical support. Reducing the distinction between plan and nonplan expenditures to coordinate investment with necessary recurrent expenditures could provide a major boost to the productivity of public investment, particularly in lagging regions, where the productivity gap is the greatest.

- Improving public service delivery requires strengthening public accountability mechanisms by empowering local community organizations, especially in lagging communities and sectors. Where institutional and governance constraints inhibit systemic reform, mobilizing local community groups—village education committees, women's self help groups, local irrigation management societies—can have an enormous impact on improving service delivery.

- Paying attention to the role of local governments can also be beneficial. With a well-designed fiscal framework and greater autonomy over resources balanced with greater accountability, local governments have the potential to play an extremely important role in improving public service delivery and effectively supporting local, particularly rural, development efforts.

I have spent a large portion of this talk focusing on the experience of India for a number of reasons. India's tremendous wealth of experience and intellectual contribution to development thinking make it an extraordinary source of lessons on accelerating development that can be of great use to the rest of the world. India has experienced the costs of excessive rigidity and has, particularly in the last decade, demonstrated that adjusting to international change and channeling it to productive purpose can raise growth and reduce poverty. India is also home to around one-third of the world's poor people. Meeting the global challenges of the MDGs will hinge in no small measure on its continued success in reform.

Conclusion: Development in a Changing World

I began this talk by noting the importance of looking ahead in a world in which ideas, resource flows, and partnerships are changing rapidly. While greater integration and change involves risks, it also presents tremendous opportunities—opportunities that can be greatly enhanced by collaboration between high-income and developing countries. The new international commitment to the MDGs offers just such an opportunity.

For their part, rich countries must deliver on their pledges to open their markets wider to developing country exports and to increase aid to developing countries that are genuinely pursuing reforms. Some movement has taken place in both areas, but if the Doha process is to move forward and the MDGs are to be achieved, stronger progress on market access, particularly in agriculture, and on more and better aid, is urgent. Many developing countries are proceeding with reform, working to improve their investment climates and to invest in and empower poor people. While they have far to go, many developing countries are delivering on the commitment to improve their policies, governance, and institutions.

India has enormous opportunities in this changing world. It enjoys huge advantages in its institutions of governance, democracy, and a professional civil service, and the continuation of its commitment to reform through different governments is encouraging. Nevertheless, all societies have vested interests that resist initiatives that would benefit the majority. Strong leadership and vision will be needed if the reforms needed to accelerate development are to be taken forward.

The need for action is pressing in three areas and can yield large rewards: strengthening public finances, especially by focusing public expenditures on investments that will enhance development; pushing forward with trade reform, both through advocacy in the international arena and unilaterally; and improving the investment climate. A final issue that cuts across all three areas of action is narrowing the widening gap between more prosperous areas and lagging regions, sectors, and communities. Addressing these issues will go a long way toward unfettering India's tremendous entrepreneurial talent and capacity for innovation.

India's strengths lead me to be cautiously optimistic about its future, but reform must move ahead strongly for development to accelerate. India is only beginning to explore some of the most difficult reform issues and still has a long way to go. I am optimistic because the barriers are big, and therefore the returns to dismantling them are high, and because India has demonstrated its ability to move forward. I am cautious because pushing forward with reform always requires overcoming short-sightedness and the special pleading by many vested interests that gain from existing structures.

In an environment in which the rich countries are not doing enough to support the partnership for development, India is well placed to take a position of leadership in shaping the discussion about international aid flows and their effectiveness, and about trade reform and the obligations of both rich and developing countries in the partnership. India has the authority of its great intellectual contributions, its demonstration of the returns to reform, its size, its democracy, and its international respect. I am confident that the coming years will see India's role in the world strengthen still further for the benefit of us all.

Notes

1. The eight MDGs, which are to be reached by 2015, comprise the following:
   - Eradicate extreme poverty: halve the proportion of people living on less than US$1 a day, halve the proportion of people who suffer from hunger.
   - Achieve universal primary education: ensure that boys and girls alike complete primary education.
   - Promote gender equality and empower women: eliminate gender disparity at all levels of education.
   - Reduce infant mortality: reduce by two-thirds the under-five mortality rate.
   - Improve maternal health: reduce by three-quarters the maternal mortality rate.
   - Combat HIV/AIDS, malaria, and other diseases and reverse the spread of HIV/AIDS.
   - Ensure environmental sustainability: integrate sustainable development into country policies and reverse the loss of environmental resources, halve the proportion of people without access to potable water, significantly improve the lives of at least 100 million slum dwellers.
   - Develop a global partnership for development: raise official development assistance, expand market access, encourage debt sustainability.

2. The calculation assumes endogenous productivity growth from tariff reductions but excludes liberalization of services and gains from reducing nontariff barriers, thereby underestimating the total gains resulting from full trade liberalization.

3. I have also discussed these topics during other recent lectures in India, for example, at the National Institute of Public Finance and Policy in January 2002 and at the National Council of Applied Economic Research in November 2002. See http://www.econ.world bank.org/staff/nstern.

4. The revenue deficit is defined as total government revenues less total expenditures on recurrent (noninvestment) items in the budget.

5. According to Saxena (2000):

   There are doctors, but there are no medicines; there are engineers, but there are no funds for construction and maintenance; there are teachers but there are not enough school buildings. While the government starts to resemble an employment agency, investment vital to maintaining the public capital stock and providing quality public services is not taking place.

References

Collier, Paul, V. L. Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy.* World Bank Policy Research Report. New York: Oxford University Press.

Kelkar, V. J. 2002. *Report of the Task Force on Direct-Indirect Taxes.* New Delhi: Akalank Publications.

Saxena, N. C. 2000. "Poverty Alleviation through Innovative Policies and Better Governance." Presentation at the India States' Reform Forum 2000, November 23–25, New Delhi. http://www.statesforum.org/reinv-govt.htm.

World Bank. 2003. *World Development Report: Sustainable Development in a Dynamic World—Transforming Institutions, Growth, and Quality of Life.* New York: Oxford University Press.

_____. 2004. *Doing Business in 2004: Understanding Regulation.* New York: Oxford University Press.

World Bank and Confederation of Indian Industries. 2002. *Improving India's Investment Climate.* Washington, D.C.

Keynote Address
Infrastructure Development in India: Emerging Challenges

RAKESH MOHAN

Infrastructure contributes significantly to economic development both by increasing productivity and by providing amenities that enhance the quality of life. The impact of infrastructure on economic growth is well documented internationally. In the Indian context too, elasticities of output with respect to various stocks of infrastructure indicate that the transport and communication sectors play a dominant role in explaining variations in gross domestic product (GDP) and its components. In addition, the index of industrial production closely tracks movements in the composite index of infrastructure industries during the 1980s and the 1990s.

In India, infrastructure development came to be explicitly recognized as part of the state's responsibilities during the regime of Emperor Sher Shah (1540–45), when a multitude of irrigation projects and roads were completed, including the famous Grand Trunk Road linking the eastern and western extremities of the country. Another major turning point in the history of infrastructure development was the initiation of investment in Indian Railways in 1853 during the British Raj. The implementation of the National Highway Development Project, coupled with Prime Minister's Gram Sadak Yojana (Prime Minister's Rural Roads Plan), will perhaps mark a similar milestone in this century.

The National Planning Committee (in 1938) and the Bombay Plan (in 1944) in pre-independence India also emphasized the state's role in putting an infrastructure network in place. They provided the necessary building blocks for action on infrastructure under subsequent plans in postindependence India, when infrastructure largely remained a state responsibility. The major policy shift on infrastructure occurred in the 1990s when, along with other wide-ranging economic reforms, the government opened the sector up to private and foreign participation because of its large financing needs.

Macroeconomic Projections

The government of India constituted the Expert Group on the Commercialisation of Infrastructure Projects in October 1994 to consider issues related to the commercialization of infrastructure projects, including institutional arrangements, legal frameworks, and financial arrangements that would facilitate the free flow of resources to infrastructure. The group submitted its report, which came to be known as the *India Infrastructure Report*, in June 1996.

The group estimated infrastructure requirements up to 2005–06 based on annual GDP growth projections of 7.5 percent during 1996–97 to 2000–01 and 8.5 percent during 2001–02 to 2005–06. Clearly the high growth of 7.3 percent in 1994–95 and its continuation in 1995–96 provided the basis for these projections. The rising trend in savings and investment also promoted an optimistic outlook for the future. The ratio of gross domestic savings to GDP had shown a steady improvement since 1993–94, attaining its peak of 25.1 percent in 1995–96. In tandem, the ratio of gross domestic capital formation to GDP also attained its peak of 26.8 percent in 1995–96. Set against such a backdrop, the Expert Group believed that the accelerating GDP would require enhanced and accelerating investment in infrastructure to sustain its momentum. In line with the East Asian experience, the group felt that sustained high growth would require an increase in the overall investment rate from the prevailing 25 percent of GDP to about 29 percent by 2000–01 and 31.5 percent by 2005–06. The achievement of 7.5 to 8.5 percent annual GDP growth at such investment levels would also require the incremental capital output ratio to decline to about 3.5, a level achieved by only the most efficient economies. The rate of industrial growth would also have to accelerate to 10 to 12 percent per year during the following 10 years.

In line with such macroeconomic projections, infrastructure investment was projected to increase from 5.5 percent of GDP in 1995–96 to 7.0 percent by 2000–01 and 8.0 percent by 2005–06. This implied that the annual level of investment was to increase from Rs 600 billion in 1995–96 to about Rs 1,100 billion by 2000–01 and Rs 1,800 billion by 2005–06 (at constant 1995–96 prices). This also implied total infrastructure investment requirements of about Rs 4,000 billion to Rs 4,500 billion between 1996 and 2001, rising to Rs 7,500 billion between 2001–02 and 2005–06. Assuming that 30 to 35 percent of total external capital inflows would go toward financing infrastructure, the group expected 15 percent of the total capital requirements for infrastructure to be externally financed, with the rest financed from domestic resources. The public sector, the engine of infrastructure investments in the past, would continue to play a major role in this sector, even as its share was expected to decline from more than 80 percent in 1995–96 to 55 percent by the end of 2005–06. The share of private sector investment in infrastructure was projected to rise from about 1 percent of GDP in 1995–96 to 2.5 percent by 2000–01 and 3.5 percent by 2005–06. This meant that in absolute terms, this investment would have to rise from about Rs 120 billion in 1995–96 to Rs 380 billion in 2000–01 and Rs 800 billion in 2005–06.

The subsequent reality has, however, turned out to be different. While the high growth momentum had continued in 1996–97, GDP growth slumped to 4.8 percent

in 1997–98. Although some recovery took place, with 6 percent plus growth during the two subsequent years, it has since remained far below the 7 percent plus attained during the high-growth phase of 1994–95 to 1996–97. GDP growth declined to 4.4 percent in 2002–03, the lowest since 1992–93. On the whole, the period from 1997–98 to 2002–03 was marked by decelerated growth of 5.3 percent, much lower than the average growth of 7.5 percent attained during the high-growth phase of 1994–95 to 1996–97 (table 1).

TABLE 1.
Selected Macroeconomic Indicators, 1992–2003

Item	1992–93 to 1993–94	1994–95 to 1996–97	1997–98 to 2002–03[a]
Growth Indicators (percent per year)			
GDP	5.5	7.5	5.3
Gross value added			
Agriculture and allied activities	5.0	4.6	1.0
Industry	4.6	9.6	4.8
Manufacturing	6.3	12.2	4.2
Services	6.5	8.3	8.0
Other Indicators (percentage of GDP)			
Gross domestic capital formation			
Gross domestic capital formation/GDP	23.3	25.8	23.9[a]
Gross capital formation/GDP	14.1	16.1	15.8[a]
Private sector			
Household sector gross capital formation/GDP	8.1	7.9	9.7[a]
Private corporate sector gross capital formation/GDP	6.0	8.2	6.1[a]
Public sector gross capital formation/GDP	8.4	7.8	6.6[a]
Gross domestic savings			
Gross domestic savings/GDP	22.2	24.4	23.1[a]
Private sector			
Gross domestic savings/GDP	21.0	22.6	24.1[a]
Household sector gross domestic savings/GDP	18.0	18.3	20.2[a]
Private corporate sector gross domestic savings/GDP	3.1	4.3	3.9[a]
Public sector gross domestic savings/GDP	1.1	1.8	−1.1[a]
Public finances			
Central government taxes/GDP	9.4	9.3	8.7
Central government interest payments/GDP	4.2	4.3	4.6
Combined revenue deficit/GDP	3.8	3.5	6.1
Combined government expenditure/GDP	27.1	25.9	26.9
Combined interest payments/GDP	4.9	5.1	5.6

Note: Gross domestic capital formation, gross capital formation, gross domestic savings, taxes, and interest payments are calculated as a percentage of GDP at current market prices. Tax and interest payments pertain to the central government.

a. Figures relating to investment and saving are averages over 1997–98 to 2001–02.

Sources: Central Statistical Organisation and Reserve Bank of India data.

At the sectoral level, the continuing slowdown was reflected in the steep decline in industrial growth from 9.6 percent during the high-growth phase to a mere 4.8 percent during 1997–98 to 2002–03. The industrial slowdown was led by a severe slackening in manufacturing growth from 12.2 percent to a meager 4.2 percent during the same period. Simultaneously, agricultural growth dipped from 4.6 percent to a modest 1.0 percent, in part because of less rain than normal during the monsoons.

The most important negative development to have taken place was a marked deterioration in public sector savings, which turned negative and fell from 1.3 percent in 1997–98 to −2.5 percent in 2001–02, as against 1.8 percent of GDP during the high-growth phase and 1.5 percent at the beginning of the 1990s.

The deteriorating fiscal health of both the central and state governments caused by high and increasing interest payments coupled with a fall in the tax to GDP ratio has contributed to the sharp decline in public sector savings. The gross tax to GDP ratio for the central government declined from 10.1 percent at the beginning of the 1990s to 9.3 percent in 1994–97 and further to 8.7 percent in 1997–2003. At the same time, the interest payments to GDP ratio for the central government increased from 4.3 to 4.6 percent. The combined revenue deficit of the central and state governments, which had declined from 3.8 percent of GDP in 1992–94 to 3.5 percent in 1994–97, increased substantially to 6.1 percent in 1997–2003. Thus more than 60 percent of the borrowing to finance the fiscal deficit is going into current account expenditures with little or no expected returns. With the declining or stagnant tax to GDP ratio coupled with a deteriorating quality of government expenditure, a substantial improvement in the fiscal deficit seems unlikely. Interest payments can be expected to increase continuously, along with revenue and fiscal deficits. Such a fiscal situation leaves few resources for investment purposes. Consequently, contrary to projections made for a stable and constant level of public sector investment in infrastructure, actual public investment has been falling since 1995–96.

The fall in public sector savings also reflects the operations of public sector enterprises in the infrastructure sectors. Levying and collecting appropriate user charges for the delivery of infrastructure services is crucial for the financial health of public sector enterprises engaged in these services. Despite increasing understanding of this issue, progress in this area has been minimal during the last decade.

The authorities have made little headway in implementing the necessary pricing reforms in the railways, state road transport corporations, state electricity boards (SEBs), and urban water tariffs in terms of levying appropriate user charges. In addition, the fare freight ratio (earnings per passenger kilometer compared with earnings per tonne kilometer) continues to be one of the lowest in the world: 0.47 in 1950–51, 0.30 in 1990–91, and 0.31 in 2001–02. A scrutiny of the financial performance of state road transport corporations shows that only 87 percent of costs could be recovered through revenue receipts in 1999–2000, leading to losses of around Rs 19.5 billion. The gap between the cost of supplying electricity and the average tariff per kilowatt-hour worsened from Rs 0.23 in 1992–93 to about Rs 1.10 in 2001–02, with the result that revenues dropped from 82 to 69 percent of costs. Similarly, even though the average cost per cubic meter of water in urban areas is estimated to be about Rs 15, average

water charges are about a tenth of this amount. The political resistance to the rise in local telephone tariffs to economic levels is also indicative of the widespread reluctance to pay user charges. The poor performance of public sector savings has to be seen in light of these developments, which have clearly constrained the public sector's ability to sustain public investments in infrastructure at the required levels.

Private corporate savings, which improved from 3.1 percent of GDP during 1992–94 to 4.3 percent during 1994–97, slipped to 3.9 percent in 1997–2003. This was largely an outcome of lower corporate profits, which could have resulted from the corporate restructuring necessitated by the impact of increased competition caused by the overall opening up of the economy. As a result, the overall gross domestic savings to GDP ratio fell from 24.4 percent during 1994–97 to 23.1 percent during 1997–02, despite a simultaneous improvement in the ratio of household savings to GDP from 18.3 to 20.2 percent.

As a consequence, the projected increase in investment has not taken place at the macro level or with reference to infrastructure. Gross domestic capital formation had improved to 25.8 percent of GDP in 1994–97 from 23.3 percent during the first phase of reforms (1992–94). It subsequently declined to 23.9 percent because of a decline in both private sector gross capital formation and public sector gross capital formation to 15.8 percent and 6.6 percent, respectively, and as a percentage of GDP from 16.1 percent and 7.8 percent, respectively, during 1994–97. Because of the corporate sector's poor performance, private sector gross capital formation fell from 8.2 to 6.1 percent from 1994–97 to 1997–2003, notwithstanding improved performance by the household sector in the latter period.

Thus the acceleration in overall economic growth that had been envisaged in 1995–96 has clearly not taken place. Instead, significant deceleration has occurred. The trend in infrastructure investment has been correspondingly lower than that projected. The public sector has been unable to keep up the level of investment that was required and envisaged because of fiscal deterioration. Similarly, the optimism about increasing private participation in infrastructure has also been belied. The enthusiasm related to private sector investment in infrastructure that was widespread in the mid-1990s has probably proved to be unrealistic worldwide, in part because of unrealistic expectations by prospective private sector investors in relation to the potential returns to equity, and also because of an underestimation of the risks inherent to infrastructure. The 1997 East Asian financial crisis also contributed to a general meltdown in the hitherto burgeoning growth in foreign direct investment in developing countries, particularly in infrastructure. Even though India was not significantly affected by the East Asian crisis, the accompanying world economic slowdown probably affected domestic developments as well, particularly through the effects of low product prices on corporate profits.

In addition to the macro developments that have led to the slowdown in infrastructure development, the lack of adequate reform in relation to the pricing and regulatory environment has also inhibited both public and private sector investment in India. Thus revisiting the framework for private investment in infrastructure is as important as the macroeconomic corrections required.

Performance of Infrastructure

As noted earlier, the Expert Group's expectations in relation to infrastructure investment have not been fulfilled because of the macroeconomic slowdown that occurred after they had made their projections. As table 2 shows, as a percentage of GDP, total actual investment in infrastructure was much lower than projected during 1995–2002. For individual sectors, projections were consistently higher than actual performance for electricity, gas, and water supply and railways but were close to projections, and even better in a few years, in the case of other transport and communications.

During 1995–2000, actual infrastructure investment undertaken by the public sector was lower than anticipated overall and for all sectors except communications (table 3). As for the private sector's performance during the same period, actual investment was lower than what had been projected for railways and communications (table 3). This is surprising, because the data on private investment in communications that are available in the national accounts do not seem to reflect the substantial private investment that has taken place in telecommunications. Mobile telephone services, radio pager services, Internet access, and other value added services that were introduced in the late 1990s were almost entirely provided by the private sector. Private investment exceeded projections for some years in the case of electricity, gas, and water supply and other transport, and total private sector investment exceeded projections during 1995–97.

The shortfall in actual infrastructure investment compared with projected investment, expressed as a proportion of GDP, seems to have been significantly greater for the public sector (about 33 percent) than that for the private sector (about 10 percent). However, the private sector shortfall has also been growing in recent years. The data suggest that an appropriate private sector response has occurred, but that it could have been better if it had not been constrained by inadequate public policy action. For example, had the planned power sector reforms taken place, particularly in the case of payment security, private sector investment in this sector would likely have been significant. Therefore the challenge for inducing greater infrastructure investment in the future involves both enhancing public sector investment and improving the policy framework for private investment.

As part of the plan to double per capita income over 2002–12, the Tenth Five-Year Plan set an ambitious target of 8 percent annual GDP growth during 2002–07. The Tenth Plan identified the lack of adequate infrastructure as one of the binding constraints facing the growth process and advocated increased public investment in infrastructure so that the level of infrastructure is commensurate with the demands of the economy. The plan assumed that public sector savings would increase based on expectations that the profitability of SEBs and other public sector enterprises would improve, which would then facilitate increased public investment.

The Tenth Plan observes that the ongoing disinvestment and privatization process has constrained the ability and willingness of the central government's public sector units to undertake investment activities. In view of the lack of sufficient alternative

TABLE 2.
Projected Investment in Infrastructure and Actual Performance, 1995–2002
(percentage of GDP)

Years	Electricity, gas, and water supply		Railways		Other transport		Communications		Total infrastructure	
	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual
1995–96	2.9	2.0	0.7	0.5	1.2	1.8	0.7	0.7	5.5	5.0
1996–97	3.0	2.2	0.7	0.4	1.4	1.7	0.7	0.6	5.8	4.9
1997–98	3.1	2.1	0.7	0.4	1.6	1.1	0.7	0.6	6.1	4.2
1998–99	3.1	2.2	0.7	0.3	1.8	1.2	0.7	0.6	6.4	4.3
1999–2000	3.2	1.9	0.7	0.3	2.0	1.3	0.8	0.7	6.7	4.2
2000–01	3.3	1.8	0.8	0.3	2.1	1.4	0.8	1.0	7.0	4.6
2001–02	3.4	1.7	0.8	0.3	2.3	0.9	0.8	0.8	7.2	3.8

Sources: Central Statistical Organisation (various issues); Expert Group on the Commercialisation of Infrastructure Projects (1996); Ministry of Finance (various issues).

TABLE 3.
Projected Investment in Infrastructure and Actual Performance by the Public and Private Sectors, 1995–2000
(percentage of GDP)

Years	Electricity, gas, and water supply		Railways		Other transport		Communications		Total infrastructure	
	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual
Public Sector										
1995–96	2.8	1.8	0.7	0.5	0.3	0.3	0.6	0.7	4.4	3.3
1996–97	2.8	1.7	0.6	0.4	0.3	0.2	0.6	0.6	4.4	3.0
1997–98	2.8	1.7	0.6	0.4	0.4	0.2	0.5	0.6	4.4	2.9
1998–99	2.8	1.7	0.6	0.3	0.5	0.2	0.5	0.6	4.4	2.9
1999–2000	2.8	1.6	0.6	0.3	0.5	0.2	0.5	0.7	4.5	2.8
Private sector										
1995–96	0.1	0.2	0.0	0.0	0.9	1.5	0.1	0.0	1.1	1.7
1996–97	0.2	0.4	0.0	0.0	1.0	1.4	0.1	0.0	1.4	1.9
1997–98	0.3	0.3	0.1	0.0	1.2	0.9	0.2	0.0	1.7	1.3
1998–99	0.3	0.5	0.1	0.0	1.4	0.9	0.2	0.0	2.0	1.4
1999–2000	0.4	0.3	0.1	0.0	1.5	1.1	0.2	0.0	2.2	1.4

Sources: Central Statistical Organisation (various issues); Expert Group on the Commercialisation of Infrastructure Projects (1996); Ministry of Finance (various issues).

institutions through which public investment can be made, especially in the infrastructure sectors, the central government may use the institutional capacity of state governments for infrastructure investment. The Tenth Plan envisages not only the use of financial resources, but also of food stocks (which could be used to pay wages in kind in rural infrastructure projects), for augmenting investment in rural infrastructure. The Tenth Plan cautions that unless the private sector's access to equity and long-term debt is enhanced substantially, the likelihood of adequate private investment in infrastructure is remote.

In contrast to actual investment performance in infrastructure, the policy sphere has witnessed active and continuous attempts to bring about reforms in the infrastructure sector. The government is slowly moving away from its traditional role as a provider of services to one of facilitator by ensuring that infrastructure services are actually delivered and are provided in a desirable manner. While liberalizing rules and procedures, the government is attempting to create an environment conducive for private participation, including foreign investment, in the infrastructure sector.

The power sector was the first infrastructure sector to be opened up to private participation, yet this sector has encountered the greatest difficulties, and private investment has been substantially short of expectations. The basic strategy was to invite private participation in power generation by means of independent power producers (IPPs) that sell power to the SEBs. The reluctance of state governments to tackle the issues of power theft and inadequate tariffs that have led to the bankruptcy of SEBs has, however, prevented financial closure of the proposed new private sector projects to generate power. State governments have attempted to bypass these problems using such stratagems as escrow arrangements and central government guarantees, which have not worked. In addition, the protracted acrimonious negotiations over the Dabhol power project in Maharashtra have highlighted the political risks facing IPPs, which are sufficient to deter any new private investment in generation.

Notwithstanding the problems in the power sector, there have been some success stories, especially in roads, ports, and telecommunications. The quality of highways will improve substantially in the coming years given the government's rapid progress in implementing the National Highway Development Project. Ports have achieved some efficiency gains through the privatization of port services and berths. The telecommunications sector has seen perhaps the most significant development: greater clarity in regulatory and policy environments has accelerated activities and expanded coverage. Several private operators are already in the market raising funds through bond financing. Tariffs in the telecommunications sector have fallen rapidly as a result of deregulation, competition, and technology. In no sector of the economy have prices come down so fast as in the telecommunications sector.

Clearly, the most urgent need is to renew the focus on the lagging sectors, namely, power, railways, and urban infrastructure. Here the guiding principle has to be the introduction of adequate user charges, either direct or through a cess, which have been central to the success of the better-performing sectors.

Sectoral Review

This section presents a detailed sectoral review of roads, telecommunications, ports, power, railways, and urban infrastructure along with an agenda for future policy initiatives.

Roads

The National Highway Development Project consists of two major components: the Golden Quadrilateral and the North-South and East-West projects. The Golden Quadrilateral project will connect the four major metropolitan cities—New Delhi, Mumbai (Bombay), Chennai (Madras), and Kolkata (Calcutta)—with four- to six-lane highways with a total length of about 5,850 kilometers. The North-South and East-West projects will connect the northern-most point of the country to the southern-most and the east to the west with a total length of about 7,300 kilometers.

The National Highway Development Project was started in 1998 and is expected to cost Rs 540 billion. Private sector investment amounts to only Rs 40 billion (7.4 percent of the total). Over the course of the project, institutions such as the World Bank, the Asian Development Bank, and the Japanese Bank for International Cooperation are expected to finance about Rs 200 billion, and another Rs 200 billion of investment would be financed from a cess levied on gasoline and diesel sales. An extrabudgetary infusion of funds in the form of market borrowing for the first phase of the National Highway Development Project is expected to amount to about Rs 100 billion, of which the National Highway Authority of India had already borrowed about Rs 85 billion as of November 2002. To date, the National Highway Authority of India has also secured loans amounting to Rs 80 billion from the World Bank, the Asian Development Bank, and the Japanese Bank for International Cooperation for the first phase of the National Highway Development Project.

This project is progressing well. Approximately 1,300 kilometers of the Golden Quadrilateral and 560 kilometers of the North-South and East-West corridors have been "four-laned" by means of a combination of widening existing roads and constructing new roads. The Golden Quadrilateral component is expected to be completed by December 2004 and the North-South and East-West projects are expected to be completed between 2007 and 2009.

Notwithstanding the efforts made over the years at the state and central levels, about 35 to 40 percent of villages have yet to be connected by all-weather roads. According to information provided by state governments, as of January 2004, 262,000 villages were still unconnected. In addition to 50 percent of the cess on diesel earmarked for the Prime Minister's Rural Roads Plan, additional funds for rural roads are to be made available from the Rs 0.50 increase in the cess on diesel in the Union budget for 2003–04. Even these funds may be inadequate if all villages with more than 500 residents are to be connected within five to seven years. Multiple agencies are involved in implementing road projects at the district and state levels under the Prime Minister's Rural Roads Plan, and coordination among these agencies needs to be

enhanced to improve efficiency. Even though the Central Road Fund is meant to be used for the development of all roads, ranging from national highways to state highways and rural roads, the funds available do not seem to be sufficient to develop state and rural roads.

The current condition and stage of development of state highways and major district roads vary widely from state to state. The status of major district roads is particularly worrisome, primarily because the funds available for developing this secondary system are inadequate. National highways are provided with reasonable development funds at the central level, while rural roads receive the lion's share at the state level. As a result, secondary roads are neglected. To deal with this issue in the future, the fund flow from the Central Road Fund for state roads could be enhanced. Several state governments, such as those of Gujarat, Madhya Pradesh, and Maharashtra, have initiated steps to involve the private sector in the development of state roads. Andhra Pradesh, Maharashtra, and Tamil Nadu have set up road development corporations that have issued bonds for financing road projects.

In all countries, the public sector has undertaken the bulk of investment in roads. Private sector investment has been confined to limited access toll highways. Even in this segment, private investment has been marginal; however, a burst of optimism occurred with respect to private investment in highways in the early 1990s, with large private projects being undertaken by such countries as China, Indonesia, and Mexico. Unfortunately, the financial difficulties both Indonesia and Mexico encountered in such road projects have discouraged further private investment in roads.

The risks involved in road investment are typically high and involve government action at almost every step. Limited access toll highways tend to encompass large-scale programs of land acquisition that are subject to unforeseen delays and often to protracted litigation by those affected. This indeed seems to be the case in Bihar, Orissa, and Uttar Pradesh in relation to the implementation of the National Highway Development Project. Rehabilitation and environmental concerns add further to construction risks and tend to cause unexpected delays. Following construction, traffic risks are typically high. Hence private investors generally perceive road investment as a high-risk activity and therefore demand relatively high ex ante returns to equity. As new highways have to be designed in such a way that their capacity is adequate for some date in the future, say 10 years, overcapacity is built in for the initial years. The returns from economically and rationally priced tolls are therefore unlikely to produce profits in the initial years of a highway's operation, with later profits compensating for losses during the initial years. These characteristics of roads deter private investment.

In view of these considerations, considerable discussion has taken place on how best to induce the private sector to invest in highways, and the government will have to undertake considerable risk mitigation to deal with private investors' concerns. There has also been the added concern in India about the level at which tolls can be levied given the widespread reluctance to pay user charges. Clearly any tolls that provide adequate returns to private investment in highways would be too high in the Indian context. Hence the inescapable conclusion is that leveraging private investment would involve the provision of some level of subsidy from the government.

The initial discussions on this issue had indicated that the best way to provide such a subsidy would be to fix the toll at a level considered feasible and then to award a highway project on the basis of the bid from a prospective investor that asks for the least amount of subsidy. The level of subsidy demanded by prospective investors would obviously vary inversely with the level of traffic, and public resources would be best used by leveraging the maximum level of private investment. Such a system would have the added advantage that heavily traveled routes would presumably be completed first, thereby leading to economic efficiency in overall resource allocation. The subsidy could, of course, be given ex ante as a lump sum to finance construction or as an annuity to finance the servicing of capital.

With the adequate availability of funds from the fuel cess and the enthusiasm of multilateral agencies to fund these projects, such a scheme for leveraging private investment in roads has not found favor, and an opportunity to leverage substantial private funds for road construction has been lost. Instead, a fully funded annuity scheme has been put into effect. Under this method, the National Highway Authority of India pays road developers an annuity to cover their full costs during the concession period following certification by an independent consultant in relation to the quality of services provided to road users. Given the much lower risks involved in this arrangement (for example, investors do not have to face any traffic risk), the result has been a relatively healthy response from the private sector. The only risk involved is miscalculation of the size of the annuity. To date, eight annuity projects have been awarded, accounting for about 476 kilometers of the National Highway Development Project.

In the 2003–04 budget speech, the finance minister announced a funding mechanism to attract Rs 600 billion to the infrastructure sector through public-private partnerships. Under the new scheme, the government will provide a subsidy in the form of an annuity flow to cover the shortfall between anticipated revenues and loan repayment obligations for build-operate-transfer projects in the road sector. The shortfall may arise from various types of risk involved in the project. Armed with government guarantees that it will cover any revenue shortfalls, the executing companies will find it easier to secure resources from financial institutions. This is another opportunity to leverage private sector investment in highways through the kind of process described earlier. However, some important issues about the authority that will be responsible for and capable of determining the level of annual subsidies and prudent and reasonable loan repayment obligations on the part of developers have yet to be agreed on. The subsidy is likely to have an adverse fiscal impact, and in addition, there is an urgent need for a proper unbundling of risks and their assignment to those participants best able to manage them to minimize the cost of risk management.

The 2002 Control of National Highways (Land and Traffic) Bill was passed to prevent unauthorized occupation of national highway land, to control access points to national highways, and to control traffic on them. As the road sector is facing considerable constraints in terms of financing (especially because of massive expansion, maintenance, and upgrading requirements), private sector participation in road

building activities has to be encouraged through well-structured operation and maintenance contracts and a combination of construction and maintenance contracts. The future plans of the National Highway Authority of India include an ambitious toll program that is going to present a huge challenge given user resistance, traffic leakages, and lower-than-expected traffic as witnessed in the case of a few projects.

Telecommunications

The substantial progress made in telecommunications since the early 1990s is a success story. The number of telephone lines grew by 25 to 30 percent each year throughout the 1990s, starting with only about 5 million in 1991. The *India Infrastructure Report* (Expert Group on the Commercialisation of Infrastructure Projects 1996) provided demand projections ranging from about 30 million to 45 million, with 35 million to 36 million being the mean projection for 2002–03. The report projected the demand for cellular mobile telephones to be between 3.5 million and 11 million by 2003. Actual outcomes were in line with the projections, with the number of basic service lines reaching 41 million and of cellular mobile connections reaching 12.6 million by the end of March 2003 (table 4).

Private sector investment required for achieving this kind of growth was estimated to be about Rs 500 billion between 1997 and 2002 (at 1995–96 prices). The estimates

TABLE 4.
Telecommunications Subscriber Base, 1988–2003
(millions)

Year[a]	Basic lines	Cellular mobile telephones	Village public telephones	Public call offices
1988	3.8	n.a.	n.a.	—
1989	4.2	n.a.	n.a.	—
1990	4.6	n.a.	n.a.	—
1991	5.1	n.a.	n.a.	—
1992	5.8	n.a.	n.a.	—
1993	6.8	n.a.	n.a.	—
1994	8.0	n.a.	n.a.	—
1995	9.8	n.a.	n.a.	—
1996	12.0	n.a.	n.a.	—
1997	14.5	0.34	—	—
1998	17.8	0.88	—	—
1999	21.6	1.20	0.341	0.52
2000	26.5	1.88	0.375	0.65
2001	32.4	3.60	0.409	0.86
2002	39.0	6.40	0.468	1.07
2003	41.0	12.60	0.507[b]	1.37[b]

a. As of March in every year.
b. As of December 2002.
Source: Telecommunications Regulatory Authority of India, *Annual Report* (various years).

of actual private sector investment during this period range from Rs 380 billion (in current prices) to Rs 450 billion, with about a quarter coming from foreign direct investment. Thus despite frequent policy changes, regulatory problems, and tariff restructuring, the growth in telecommunications must be viewed as a success. This experience shows that the Indian private sector is willing to take risks and invest in infrastructure provided it can envisage a certain degree of payment security and profitability.

Reform in the telecommunications sector began in 1992–93 with the opening of value added services to the private sector. Subsequently, after extensive deliberation within and outside the government, in 1994 the national telecommunications policy (referred to as the NTP 1994) opened basic telecommunications services to competition and initiated cellular mobile telephone services. The intent was that private initiative would complement public sector efforts to raise additional resources through increased internal generation of funds and the adoption of innovative approaches, such as leasing, deferred payment, build-operate-transfer projects, and the like. The NTP 1994 also envisaged the provision of a public telephone for every 500 people in urban areas and at least one public telephone in every village. (As of December 2002, there were 1.37 million public telephones in the country and about 500,000 village public telephones).

The method employed for encouraging the private sector to become involved in both basic and cellular services was, as in many other countries, the auctioning of license fees. The auction process elicited excessively high bids, even from bidders that lacked substantive telecommunications experience, or even any other business experience. Once the licenses had been awarded and operations had begun, inevitable complaints arose about the license fees being too high and uneconomic. As various developments had taken place in the telecommunications sector in the intervening years and new issues had arisen, the government announced a new telecommunications policy (referred to as the NTP 1999). The issues that had arisen during this period related to the perception that the original license fees had been excessive, the lack of adequate competition (only two operators were permitted to operate in each area), the continuing changes in technology, and the emergence of India as a significant player in the information technology industry.

In July 1999, the NTP 1999 offered a package whereby existing cellular and basic service providers could move from a fixed license fee to revenue sharing, and Mahanagar Telephone Nigam, Limited (MTNL), the government-owned telephone company operating in New Delhi and Mumbai, was permitted as a third cellular services operator to promote competition. The government opened up national long distance services to private operators without any restriction on the number of operators and with moderate entry fees. International long distance services were then opened up in 2001, also with no limit on the number of operators and moderate entry fees. Both are subject to license fees being paid as revenue sharing.

Thus significant competition was introduced in the telecommunications market starting in 2000–01. The consequence has been dramatic: cellular mobile tariffs have fallen by about 75 percent since 1999, and long distance tariffs, both domestic and international, fell by 60 percent between 2000 and the end of 2002.

Corresponding organizational changes also took place during 2000–01. The two government departments involved in providing telecommunications services were corporatized, namely, the Department of Telecommunications Services and the Department of Telecommunications Operations. A new public sector company, Bharat Sanchar Nigam, Limited (BSNL), was allocated all the service provision functions of these two departments with effect from October 2000. A fourth cellular operator was also permitted in all areas. With the introduction of effective competition in the cellular mobile services sector, the Telecommunications Regulatory Authority of India made cellular mobile tariffs free from regulation while reserving the right to intervene in the case of any malpractice, such as predatory tariffs.

The success of the telecommunications policy is reflected not only in the growth of services and the substantial reduction in tariffs, but the objective of providing rural telephones is also making good progress.

Even though the outcome of reforms in the telecommunications sector is illustrated by the greater role of the private sector, the higher tele-density, the decreasing gap between supply and demand, and the availability of services at substantially cheaper rates, the new challenges in the telecommunications sector concern the ongoing convergence that is taking place and the appropriate regulatory response to greater competition in the market. Excessive fragmentation has occurred in the provision of services on the basis of type of service, geographical coverage, and even technology (for example, cellular mobile services versus wireless local loop limited mobile services). This is constraining the convergence of different services, broadcasting, multimedia, and telecommunications through a single national service provider. As a result of this fragmentation, none of the operators is able to build a technology base that would strengthen their bargaining power in the international market. Nevertheless, even though consolidation is indeed taking place through mergers and acquisitions, India is likely to end up with four or five private operators along with the public sector incumbent. A key issue for policy is how to hasten this process through the consolidation of different licenses.

The corporatization of the Department of Telecommunications into BSNL and the privatization of Videsh Sanchar Nigam, Limited, the previously government-owned international long distance monopoly operator, have been bold steps on the part of the government. Now both BSNL and MTNL have to be strengthened as corporate entities. With about 40 million lines between them, they would probably rank among the top 6 to 10 telecommunications entities in the world. Government equity in MTNL is now about 55 percent, while BSNL remains 100 percent government owned. Two opposite kinds of problems are arising from their status as government-owned incumbents. On the one hand, with open entry in almost all service segments, the privileges given to incumbents (such as nonpayment of license fees) result in unfair competition with private sector operators. On the other hand, the incumbents' public sector status constrains them from competing freely with the new operators. The recent intervention by the government on restraining BSNL and MTNL from adjusting their tariffs is a case in point. A key issue, therefore, relates to the further

restructuring of BSNL and MTNL so that they can compete effectively, while also providing fairness and transparency in their operations with respect to the private sector.

Some have argued that BSNL deserves preferential treatment from the government, because it has to bear the social obligations of providing unprofitable telephone services in rural and other remote areas. Private operators' failure to comply with their licensing requirements related to the provision of rural services has reinforced this view. BSNL has argued not only that it should not pay license fees, but that it should also receive additional subsidies from the government. Whereas the argument for subsidies for social obligations is justified, this must be done in a transparent fashion through the universal service obligations fund as proposed in the NTP 1999. The universal service obligations fund should be funded from earmarked resources from the license fee (for example, 5 percent of all license fee revenues from all operators), including BSNL. Fairness between public and private sector operators would then be established and perceived as such. Furthermore, as and when BSNL is disinvested and privatized, such a system could continue to operate without difficulty.

Efficient, credible, and authoritative functioning of the regulator is essential for the functioning of the telecommunications market. The overall functioning of the Telecommunications Regulatory Authority of India must be judged an unusual success: it has been effective in introducing competition, in reducing tariffs, and in rapidly expanding services. However, its functioning has been subject to a number of controversies, and recent events suggest that it needs to be strengthened considerably. The key issues are as follows:

- The government must eschew the temptation to interfere in decisionmaking by the Telecommunications Regulatory Authority of India, in particular, as a consequence of any corporate or political lobbying. This erodes the regulator's credibility and authority.

- The Telecommunications Regulatory Authority of India must be strengthened professionally. This will be difficult unless it is made financially independent and its compensation structure is delinked from the government, but made completely transparent.

- The authority's hiring practices should be more transparent and process linked, so that it does not become a sinecure for retiring civil servants. The appointment of respected professionals would also add to its technical strengths and boost its respect in the market.

With the increasing difficulties encountered in tariff restructuring, strengthening the regulator and making it more independent is absolutely essential. The restructuring of basic service rentals and tariffs has been resisted politically, even when there were only about 41 million fixed lines in the country, and clearly the poor did not have telephones. The political sensitivity of telephone tariffs will only increase in the future as tele-density expands. Tariffs will have to be at economic levels if the telecommunications sector is to remain sound and not go the way of the power sector.

Ports

A common characteristic of the rapidly growing East and Southeast Asian countries has been the speed with which their trade grew during their high-growth periods. A higher share of trade in an economy helps to improve efficiency. A country improves its resource allocation by exporting those goods for which it exhibits a competitive advantage and importing those for which it does not. As its comparative advantage changes, so does the composition of its exports and imports. Thus to achieve higher economic growth and efficiency levels, the trade to GDP ratio needs to increase. Improving the efficiency of ports and expanding their capacity is essential for promoting the growth of trade and export competitiveness.

In the wake of 20 percent plus annual growth of exports during 1993–96, the *India Infrastructure Report* (Expert Group on the Commercialisation of Infrastructure Projects 1996) had projected annual average growth in exports of about 15 percent between 1995–96 and 2005–06. Thus exports were projected to increase from about 10 percent of GDP in 1995–96 to about 17 percent in 2002–03 (about US$87 billion) and 18 percent by 2005–06. Correspondingly, imports were expected to increase from some 11 to 12 percent of GDP in 1995–96 to about 18 percent in 2002–03 and 19 percent in 2005–06. Total trade as a proportion of GDP was expected to have reached 35 percent by 2003. The projections for port traffic were based on these trade projections, thus total port traffic was expected to increase from a level of around 215 million tonnes (MT) in 1995–96 to 480 MT by 2002–03 and 650 MT by 2005–06. In fact, the growth in trade was much slower than anticipated, reaching a level of about 22 percent of GDP in 2002–03, with exports accounting for 10 percent and imports for 12 percent. Thus total port traffic in 2002–03 was only 313 MT. It is partly because of this relatively slow growth in trade in goods that ports were not a constraint in the late 1990s.

Port functioning has improved significantly, and there has been no shortage of investment by both the private and public sectors. The amount of cargo handled by the major ports has increased steadily, albeit modestly, rising from 215 MT in 1995–96 to 313 MT in 2002–03 (table 5). The Ninth Five-Year Plan (1997–2002) had only targeted traffic of 289 MT in 2001–02 compared with the anticipated capacity of 344 MT, thus port capacity is not a constraint. During the Ninth Five-Year Plan, 17 private sector or captive port projects totaling 60 MT capacity and involving total investment of about Rs 35 billion were approved and are at various stages of construction.[1] The Tenth Five-Year Plan (2002–07) visualizes 6 percent annual growth in traffic, a traffic load of 415 MT in 2006–07, and a total capacity of 470 MT at the major ports.

The principal indicators of efficiency have improved at the major ports as follows:

- The average waiting time for a berth fell from 1.7 days in 1996–97 to 0.5 days in 2001–02.

- The average turnaround time declined from 7.5 days in 1996–97 to 3.7 days in 2001–02.

TABLE 5.
Port Traffic, 1992–2003

Year	Port traffic (MT)	Growth (%)
1992–93	167	9.0
1993–94	179	7.6
1994–95	197	10.0
1995–96	215	9.2
1996–97	227	5.5
1997–98	254	12.0
1998–99	252	−1.1
1999–2000	272	8.0
2000–01	281	3.4
2001–02	288	2.4
2002–03	313	8.9

Source: Ministry of Statistics and Programme Implementation, *Annual Report* (various years).

- The output per ship berth day increased from 4,497 tonnes in 1996–97 to 6,972 tonnes in 2001–02.

- The productivity of labor in terms of output per gang shift increased from 307 tonnes in 1997–98 to 413 tonnes in 2000–01.

Productivity indicators have, however, varied widely across ports. This suggests that institutional innovations have occurred at the better-performing ports, such as the Jawaharlal Nehru Port Trust in New Mumbai, which other ports could also adopt.

The government has laid down guidelines for private sector participation in leasing assets; constructing and operating terminals, berths, warehouses, tank farms, container freight stations, and captive power plants; and providing pilotage and captive facilities for port-based industries. In December 1998, the government approved 100 percent foreign direct investment in ports through an automatic approval route via the Reserve Bank of India. In addition, guidelines on joint port ventures were issued in September 2000; model bid documents, including license agreement, have been finalized to attract private sector investment and participation; and the norms for private build-operate-transfer projects in major ports were relaxed in 2001–02 in the context of limited private participation in major ports. To date, 41 private sector projects involving capacity addition of about 160 MT and investment of about Rs 108 billion are at various stages of evaluation and implementation, of which 7 projects have already been completed. Private sector participation in major ports has mainly involved licensing the operations of existing container berths or granting build-own-operate-transfer concessions for increasing terminal capacity. This demonstrates that there is no shortage of private sector willingness to invest in the port sector, and higher trade growth is likely to elicit an even stronger private sector response.

Containerization is increasing in relation to international trade. Accordingly, container traffic at major ports has shown impressive growth of more than 10 percent

during 2001–03. At present, about 70 percent of the containers that flow through India are trans-shipped through the ports of Dubai, Salalah (Oman), Singapore, and Colombo (Sri Lanka), leading to increased delays and higher costs. Thus India needs to develop hub ports, one on the east coast and one on the west coast. Proposals already exists for developing an international container trans-shipment terminal at Vallarpadam in Cochin.

Even though the major ports continue to handle about 75 percent of port traffic, state ports are increasingly exhibiting higher traffic growth than the major ports. During the Ninth Plan (1997–2002), when overall traffic growth was 8.5 percent, it grew by 27.9 percent for state ports and 4.9 percent for major ports. In any case, alternatives to the congested major ports are needed, particularly in the context of the development of coastal shipping. The potential for private investment in developing minor and intermediate ports is large, and Gujarat is by far the most active state in attracting private participation. Andhra Pradesh, Orissa, and Tamil Nadu have also invited private participation to develop minor ports by offering long-term leases. The state governments and the Union Ministry of Surface Transport have set up the Maritime States Development Council to formulate an integrated ports policy that includes minor ports.

When the government changed its policy to permit private investment in ports, it had to set up a mechanism for setting tariffs on a transparent and fair basis. With all major ports being in the public sector and structured as port trusts, previously tariffs were proposed by port trusts and approved by the Ministry of Surface Transport. Accordingly, the Tariff Authority for Major Ports was set up in 1997. The authority has generally worked on cost-plus principles but has done so on a transparent and consultative basis. It initiated a system of transparent consultation with all stakeholders before approving or setting any tariffs. Previously each port trust had set tariffs on an ad hoc basis with little rationale or uniformity in approach. The authority has attempted to work on broad cost-plus principles but moderated by a normative approach that allows port trusts a certain return to capital. It also aimed for a generally similar approach across ports, which was then applied to the new private port operators.

With the achievement of some initial rationalization, the Tariff Authority for Major Ports now needs to focus more proactively on promoting competitive tariffs: the users of cargo services should not end up paying for port or labor inefficiencies. Tariff policy should now be used as leverage to prescribe standards of service, thereby contributing to productivity and efficiency, and can also be used to ration port capacity, with high tariffs for congested ports; however, this cannot be taken too far, because the land transportation facilities vary tremendously between ports. Competition between ports should also be encouraged through flexibility in pricing.

With the movement toward the corporatization of ports and the increasing importance of minor ports in total cargo throughput, restricting the jurisdiction of the Tariff Authority for Major Ports to major port trusts, as is currently the case, will progressively reduce its effectiveness. Ironically, as major port trusts are corporatized, they will move out of the authority's jurisdiction, thereby defeating its original purpose

of providing a level tariff playing field for port trust operations and newer private operators. Thus further thought needs to be given to the functioning of the authority. It should have jurisdiction over all ports over some stipulated size, but its pricing strategy needs to be much more flexible to promote competition and reduce prices.

Indian port tariffs are generally higher than those of other ports, and despite the improvements cited, their efficiency also compares poorly with the world's major ports including those in the region, such as the ports of Colombo, Shanghai, and Singapore. If Indian trade is to expand substantially in the future, significant new investment will be required along with the modernization of existing facilities. This calls for significant organizational restructuring of existing port trusts, including an overhaul of traditional, restrictive labor practices. Fortunately, the tariff structure is such that port investment can be extremely lucrative, even as efficiency increases and port tariffs come down.

Constrained by their structure, the port trusts are unable to compete with private operators and minor ports. Even though the port trusts are being encouraged to convert themselves into limited companies, corporatization with continuing government ownership and control may be of little help without the design of appropriate governance structures for the new corporatized entities. The existing port trust structure does promote participation by local governments and other stakeholders through membership in the trusts, thus a good degree of consultation will be needed for any transition. One option would be for the port trusts or their successors to be converted into a landlord port structure, whereby each terminal may be spun off as a separate company and competitively bid, with or without the port authority being a joint venture partner, to boost intra-port competition. These terminal companies could compete and invest across diverse ports so as to mitigate the risks inherent in operating out of a single base. They should be exposed to competition in raising funds from the capital market and to competition for management control. To enhance inter-port competition, port connectivity needs to be improved by means of intermodal coordination between the railways, the Container Corporation of India, the Central Warehousing Corporation, and the National Highway Authority of India. To sustain both inter- and intra-port competition, the key regulatory issues—such as dispute resolution mechanisms, entry of new players, concessions, and mergers—in the ports sector need to be clearly enunciated. In ports where private investment has taken place through container terminals, such as in Chennai and Nhava Sheva, conflicts of interest have arisen with the host port trusts performing the role of landlords, as well as with operating competitors. To deal with these issues, perhaps the Tariff Authority for Major Ports could be turned into an overall port regulator rather than being restricted to determining and approving tariffs.

Power

The power sector was the first infrastructure sector to be opened up to private sector investment. When economic reforms were initiated in 1991 and significantly higher growth was expected as a result, policymakers realized that public sector resources would be inadequate for the kind of investment required in the power sector.

Internationally, and particularly in East and Southeast Asia, governments had hit on the idea of using IPPs to generate and sell power to existing grids. Consequently, large-scale plans were drawn up to attract investment in IPPs from both domestic and international investors.

Installed capacity at the beginning of the Eighth Five-Year Plan (1992–97) was about 69,075 megawatts (MW). This increased by about 16,500 MW during the Eighth Plan, compared with the planned addition of about 30,500 MW (table 6). The *India Infrastructure Report* (Expert Group on the Commercialisation of Infrastructure Projects 1996) had projected capacity addition of about 33,000 MW between 1996 and 2001 and 50,000 MW between 2001 and 2006, along with significant improvements in efficiency to improve the utilization of existing capacity. Correspondingly, the Ninth Five-Year Plan projected the addition of some 40,000 MW between 1997 and 2002, and the Tenth Plan projects the addition of another 41,000 MW between 2002 and 2007 (table 6). The actual achievement was an addition of about 19,000 MW during 1997–2002. Given this poor outcome, had it not been for the slowdown in both industrial and agricultural growth during 1997–2002 and the significant improvement in the plant load factor from 57 percent in 1992–93 to almost 70 percent in 2001–2002, India would have faced a severe power shortage.

In view of the poor financial condition of SEBs caused by both inefficiency and low agricultural and domestic tariffs, the original idea was to provide payment security to private investors through government guarantees and escrow arrangements that gave them payment priority. The expectation was that these would be temporary measures, because the SEBs would improve their functioning and become able to pay IPPs adequately. These arrangements were made in the context of generous assumptions about returns to equity embedded in the power purchase agreements. Consequently, the response from both international and domestic investors was enthusiastic. However, the program was replete with errors on all sides, with the outcome being that expectations in relation to private sector investment in generation were scaled down considerably for the Tenth Five-Year Plan (table 6).

The reluctance of state governments to tackle the basic issues of power theft and inadequate tariffs has led to the bankruptcy of SEBs, making it impossible for viable private investment to take place in the power sector. It is now well recognized that the financial difficulties of the SEBs lie at the heart of the power sector's problems. The financial position of all the SEBs deteriorated rapidly during the 1990s. Except for the SEBs in Himachal Pradesh and Maharashtra, all [SEBs] recorded losses between 1992–93 and 2001–02 ranging from Rs 40 million to Rs 37 billion. According to Planning Commission estimates, during 2001–02 alone, commercial losses amounted to some Rs 240 billion (excluding subsidies). Such huge losses have adversely affected SEBs' ability to supply electricity reliably, leading to default in payments to the public sector units responsible for power generation and transmission.

The root of the problem lies in the gap between user charges and the cost of supply. Despite a graduated increase in user charges, the gap between the cost of supply and the average tariff per unit of electricity produced has actually worsened from a level of Rs 0.23 in 1992–93 to about Rs 1.10 in 2001–02 (table 7). Revenues

Table 6.
Capacity Addition in the Power Sector, Targets Versus Achievements, 1992–2007
(MW)

Plan and Power source	Target				Achievement			
	Central	State	Private	Total	Central	State	Private	Total
Eighth Five-Year Plan (1992–97)								
Hydroelectric	3,260	5,860	162	9,282	1,465	795	168	2,428
Thermal	8,498	9,010	2,646	20,156	6,252	6,040	1,262	13,555
Nuclear	1,100	n.a.	n.a.	1,100	440	n.a.	n.a.	440
Total	12,858	14,870	2,810	30,538	8,157	6,835	1,430	16,423
Ninth Five-Year Plan (1997–2002)								
Hydroelectric	3,455	5,815	550	9,820	540	3,912	86	4,538
Thermal	7,574	4,933	17,038	29,545	3,084	5,538	4,975	13,597
Nuclear	880	n.a.	n.a.	880	880	n.a.	n.a.	880
Total	11,909	10,748	17,588	40,245	4,504	9,450	5,061	19,015
Tenth Five-Year Plan (2002–07)								
Hydroelectric	8,742	4,481	1,170	14,393	n.a.	n.a.	n.a.	n.a.
Thermal	12,790	6,676	5,951	25,417	n.a.	n.a.	n.a.	n.a.
Nuclear	1,300	n.a.	n.a.	1,300	n.a.	n.a.	n.a.	n.a.
Total	22,832	11,157	7,121	41,110	n.a.	n.a.	n.a.	n.a.

n.a. Not applicable.
Source: Planning Commission, plan documents.

TABLE 7.
Cost Recovery, Cost of Supply, and Tariffs in the Power Sector and the Wholesale Price Index, 1992–2002

Year	Average cost per unit (Rs)	Average tariff per unit (Rs)	Cost recovery (%)	Index of average cost per unit (1993–94 = 100)	Growth of average cost per unit (%)	Index of average tariff per unit (1993–94 = 100)	Growth rate of tariff (%)	Average wholesale price index (1993–94 = 100)	Average wholesale price index inflation rate (%)
1992–93	1.28	1.05	82	86	n.a.	92	n.a.	n.a.	n.a.
1993–94	1.49	1.17	78	100	16.3	100	10.7	100	8.4
1994–95	1.63	1.28	78	110	9.6	110	9.7	113	12.5
1995–96	1.80	1.39	77	121	9.9	119	8.6	122	8.1
1996–97	2.16	1.65	77	145	20.0	142	18.9	127	4.6
1997–98	2.40	1.80	75	161	11.2	155	9.1	133	4.4
1998–99	2.63	1.87	71	177	9.8	160	3.6	141	5.9
1999–2000	3.05	2.07	68	205	16.0	177	10.8	145	3.3
2000–01	3.27	2.26	69	220	7.3	194	9.3	156	7.2
2001–02	3.50	2.40	69	235	6.9	206	6.0	161	3.6

n.a. Not applicable

Sources: Finance Ministry, *Economic Survey 2002–03*; Government of India.

dropped from 82 percent of costs in 1992–93 to 69 percent by 2001–02. While some rationalization of tariffs for the subsidized sectors, that is, agriculture and the domestic sector, has taken place, the process is clearly far from complete.

What is interesting about table 7 is that the increases in average tariffs have actually exceeded average inflation as measured by the wholesale price index. Thus tariffs have increased substantially, but costs have risen even faster.

The gap between the cost of supply and the average tariff has been accentuated because of losses in transmission and distribution, which correspond to electricity produced but not paid for. SEBs' transmission and distribution losses rose from 24.8 percent in 1997–98 to 30.9 percent in 1999–2000. Such losses are caused by a variety of factors, such as electricity sold at low voltage, sparsely distributed loads across large rural areas, inadequate investments in distribution, improper billing, and outright theft.

The SEBs' poor financial health and the resultant spillover of financial losses onto the generating utilities are reflected in the subdued growth in power generation, which was 3.1 percent in 2001–02, down from 3.9 percent in the preceding period (table 8). The growth in capacity addition has slowed down since a turnaround in 1997–98 and became negative in 2000–01 and 2001–02. By contrast, the plant load factor, which is a measure of the operational efficiency of thermal power plants, has shown a steady improvement from 57 percent in 1992–93 to about 70 percent in 2001–02 in all the regions except the northeast. The demand-supply gap peaked at 11.5 percent of demand in 1996–97, declined to 5.9 percent in 1998–99, but increased again to 7.5 percent in 2001–02.

The government initiated power sector reforms in 1991 to encourage competition in generation, transmission, and distribution. The Central Electricity Regulatory Commission was set up at the national level in 1997, and state electricity regulatory commissions have now been set up in 22 states, of which 13 have issued tariff orders so far. Private participation in power transmission has been allowed with the passage of the Electricity Laws (Amendment) Act in 1998. The Central Electricity Regulatory Commission established the Indian Electricity Grid Code in January 2000 to ensure grid discipline for individual players in the transmission and distribution sector. The commission's availability-based tariff order of January 2000 is expected to encourage reliability and efficiency in generation.

In 2003 parliament passed the Electricity Bill, originally drafted in 2001, to replace the 1910 Indian Electricity Act, the 1948 Electricity (Supply) Act, and the 1998 Electricity Regulatory Commission Act. The bill recognizes trading in power as a distinct activity and provides a legal framework for enabling reforms and restructuring the power sector. SEBs in nine states have already been unbundled and/or corporatized, and distribution operations in Delhi and Orissa have been privatized. The new distribution companies are expected to emulate the success of private sector distribution in Ahmedabad, Kolkata, and Mumbai and to ensure effective enforcement of user charges.

The Accelerated Power Development Programme, initiated in 2000–01 and subsequently modified as the Accelerated Power Development and Reforms Programme,

TABLE 8.
Physical Performance of the Power Sector, 1992–2002

Year	Generation (kilowatt hours, billions)	Thermal plant load factor (%)	Capacity addition (MW)	Demand-supply gap	
				MW	Percentage of demand
1992–93	301	57.1	3,537	25,442	8.3
1993–94	323	61.0	4,538	23,758	7.3
1994–95	351	60.0	4,598	24,979	7.1
1995–96	380	63.0	2,123	35,676	9.2
1996–97	394	64.4	1,624	47,590	11.5
1997–98	420	64.7	3,226	34,175	8.1
1998–99	448	64.6	4,242	26,349	5.9
1999–2000	480	67.3	4,507	29,836	6.2
2000–01	499	67.7	3,775	39,816	7.8
2001–02	515	69.9	3,115	39,276	7.5

Source: Ministry of Power, *Annual Report* (various years).

with an outlay of Rs 35 billion, has been designed to assist reforms in the distribution sector. It aims to achieve 100 percent metering, introduce energy audits, replace distribution transformers, and develop information technology solutions relating to power flow at critical points to ensure accountability at all levels. The Accelerated Power Development and Reforms Programme also gives the states financial incentives to reduce the gap between revenues and the unit costs of supply. The Ministry of Power has already signed memoranda of understanding with 25 states to undertake reforms in a time-bound manner.

On the issue of outstanding SEB dues to central public sector units, the Ahluwalia Expert Group constituted by the Ministry of Power in 2001 designed a securitization scheme for a one-time settlement of the approximately Rs 420 billion the SEBs owed the central government's public sector units following a write-off of Rs 96 billion. To this end, the Ministry of Power, state governments, and the Reserve Bank of India signed a tripartite agreement in March 2003. In accordance with the agreement, the respective state governments securitized the past dues of SEBs in the form of state government securities issued to the central public sector units. It was agreed that, in the future, defaults in current payments would result in a phased reduction in the supply of power from central power stations.

The one-time settlement of SEB dues needs to be strengthened by pricing power correctly. Once the SEBs' capacity to pay has been enhanced in a durable manner, investment in power generation and transmission with active private participation is expected to take off. The passage of the 2003 Electricity Act has now created an enabling environment for a competitive power sector; however, a great deal of work remains to be done, particularly at the state level. The reform template the Ministry of Power's Expert Committee on State-Specific Reforms recently put forward could be helpful in making further progress.

The last 10 years have seen intensive discussions related to power sector reforms, and the policy framework to implement them is now in place. The key issue is curbing theft and restructuring tariffs. The increase in tariffs that has already taken place is indicative of consumers' willingness to pay provided they have access to a reliable power supply. The focus now has to be on a nationwide campaign to eliminate power theft. Such theft is clearly not being done by the poor, because electricity consumption requires the ownership of items that consume electricity. If the government does not demonstrate its intent to curb the theft of power, increasing power tariffs much further will be difficult. Industrial tariffs are already too high in comparison with both international standards and local costs.

Railways

Indian Railways has been India's prime mode of transport and has the distinction of being the second largest railway system in the world under single management. The railways have historically played an important integrating role in the country's socioeconomic development. Their role in economic development is also important,

because of its innate advantage as a mode of surface transport that is more energy efficient and environment-friendly than other modes of transport.

However, Indian Railways ran into severe financial difficulties in the 1990s, which has hampered its growth and cast severe doubts on its ability to provide competitive transport services in the future. If the growth of the economy accelerates, the supply of all transportation services would have to accelerate correspondingly. With an open economy and an increasingly competitive world environment, transportation will also have to become increasingly competitive in terms of both cost and quality of services. Given that India is a large, continental economy with a large proportion of its activities located inland, the role of Indian Railways in providing such competitive services will be a critical part of the solution to India's infrastructure needs.

As in the case of power, railways have traditionally been seen as an essential public service whose use should not be denied even to those unable to pay. As a result, freight users have been seen as a class of users who could easily pay, and pay more. Consequently, freight services subsidize passenger services and upper-class passengers subsidize other passengers. This was a viable system in the context of a closed economy, because higher freight charges could simply be passed on to consumers. Moreover, as the main freight user was the public sector operating under an administered price system (for oil, steel, coal, food grains), this caused no marketing problems for the railways. This arrangement is no longer sustainable in an open economy.

With the implementation of the Fifth Pay Commission award,[2] the railways' financial difficulties became accentuated in the late 1990s. For the first time in 17 years, in 2000–01 (and 2001–02) Indian Railways was unable to pay its dividend to the government. A commercial enterprise in such a situation would have been declared in default and would have had to undergo restructuring.

Competition has been increasing across all sectors of the economy, and the transportation sector is no exception. After trucking was deregulated in the 1980s, road transportation grew rapidly and has had a significant impact on the railways' market share. Features like greater customer orientation, flexibility, and lower cost of short hauls are increasing the share of roads, even for bulk commodities that have been the traditional stronghold of the railways. The annual growth rate of freight carried by Indian Railways (in net tonne kilometers) averaged 5.33 percent between 1984 and 1991, but dropped to 1.86 percent over 1992–99. The dominance of roads is likely to increase further with the completion of the Golden Quadrilateral component of the National Highway Development Project and the increasing use of pipelines for transporting petroleum and petroleum products.

The loss of market share in the profitable freight business, the lack of flexibility in pricing, the high cost of internally sourced products and services, and the investments made in unprofitable projects have meant that the rate of increase in costs has outstripped the rate of growth in revenues. Investment in unprofitable projects escalated during the 1990s. First, the adoption of the project to implement a standard gauge nationwide involved large investments during this period and has been particularly harmful to the finances of Indian Railways. Second, the temptation to begin myriad new lines for political reasons was much greater during the politically fractured

1990s and continues to this day. Revenue growth has also suffered from the saturation of freight traffic on trunk routes, particularly the Golden Quadrilateral. This is partly due to the large differential in speed between passenger and freight trains, which severely constrains the freight-carrying capacity of trunk routes. Timely investment in introducing new technology and upgrading tracks and signaling could have eased this constraint.

The proportion of expenditure on repairs and maintenance has been declining over the years, and adequate investment has not been made in track renewal and other safety-related areas. Consequently, over the last 10–11 years, the arrears in track renewals have grown from about 3,500 kilometers to some 13,000 kilometers. To arrest the steep decline in its share and to improve the quality of its services, Indian Railways needs to significantly increase investment in infrastructure. However, if the existing trends in cost increases, uneconomical tariff setting, and investments in unprofitable projects continue, Indian Railways will be unable to find the funds for such investments.

In recognition of these problems, in 1999 Indian Railways appointed an Expert Group on Indian Railways, which submitted its report in July 2001. The Expert Group concluded the following:

> If Indian Railways is to survive as an ongoing transportation organisation, it has to modernize and expand its capacity to serve the emerging needs of a growing economy. This will require substantial investment on a regular basis for the foreseeable future. With the prospect of getting substantial free or subsidised resources from the government being unlikely, new investment will have to be financed on a commercial basis. This is the challenge facing Indian Railways (Expert Group on Indian Railways 2001, p. 61).

Indian Railways' ability to accelerate the growth rate of its revenues from freight and passenger traffic is central to the success of any effort to restructure the organization and to finance the necessary investments. With these objectives in mind, the Expert Group recommended a multitrack strategy as follows:

- *Tariff rebalancing.* To correct the imbalance between freight and passenger traffic, both tariffs have to be rebalanced. Any further increase in freight tariffs would lead to a further loss of freight share. Within passenger tariffs, the ratio of the lowest to the highest fares would have to be reduced from the existing high ratio of about 1:14 to about 1:9, as the Railway Freight and Fares Committee had also recommended in 1993. This could be done over a period of about five years.

- *Major investment program.* Increasing revenue will involve a significant increase in both freight and passenger traffic of about 7.0 to 7.5 percent per year. Achieving such growth will require modernization; introduction of high-speed, modern passenger services; commodity-specific freight strategies; and new technology, particularly in signaling and communications.

- *Organizational restructuring and corporatization.* The foregoing cannot be accomplished using a business-as-usual approach. Thus Indian Railways has to be reorganized from its current departmental form of organization to give it a corporate

framework. This would be essential to achieve the kind of commercial and customer orientation needed to achieve the goals outlined. Such major restructuring would be a complex task and would have to be done carefully over a five- to seven-year period. The process could start by recasting accounts in a corporate framework to enable accountability and the commercialization of Indian Railways.

- *Separation of functions.* In line with developments in other infrastructure sectors, separating policy setting, regulatory functions, and operational functions would be desirable. The government should obviously be responsible for policymaking; a regulatory authority could be in charge of regulation, including tariffs; and the corporatized Indian Railways Corporation would be responsible for all commercial operations.

The Expert Group argued that the financial situation of Indian Railways is currently such that anything short of a bold program would not rescue it from its current financial straits. The response by Indian Railways and the government to the Expert Group's report has been less than lukewarm. Nevertheless, some tariff rebalancing has been done. Such rebalancing might continue, particularly because of market pressures. Raising the highest fares is becoming increasingly difficult because of emerging competition from the airlines. In addition, modernization of the trunk rail routes between the four metropolitan cities of New Delhi, Mumbai, Chennai, and Kolkata has begun with assistance from multilateral agencies, but how the project will be fully financed and how these finances will ultimately be serviced is not yet clear. On the organizational side, there has been no response at all. The prognosis for Indian Railways being able to perform its functions efficiently in the future is not positive.

Urban Infrastructure

The attainment of high growth requires well-functioning and efficient cities. India now has 27 cities with more than a million inhabitants each and more than 400 cities with populations greater than 100,000 people. As growth proceeds, urbanization is likely to accelerate, with activities shifting away from agriculture and toward industry and services. Efficient production of industrial goods and delivery of services require agglomeration.

In the other areas of infrastructure, either significant policy changes have already been made to improve performance (telecommunications, roads, ports), or policymakers are aware of the required policy changes, but they are proving difficult to implement (power and railways). In the case of urban infrastructure, however, there is little understanding of what needs to be done and how. Part of the problem is intrinsic. The implementation of urban infrastructure programs lies with city-level agencies or some state agencies, but these are typically short of resources and technical competence. A great deal of policy discussion takes place at the national level, but the central government has few policy instruments for steering urban infrastructure in any particular direction. The central government has initiated a number of centrally sponsored schemes, but the resources allocated to these schemes are minimal.

As in the case of other infrastructure, urban infrastructure investment suffers from a lack of resources resulting from inadequate municipal finances and low user charges. Given the current state of central and state finances, little downward devolution is possible. As concerns the provision of public goods, municipalities have few alternatives to generating local resources through local taxes, especially property taxes, which should be buoyant in the face of rapid urbanization and an incipient housing boom. Because of inefficient and antiquated property tax systems, many cities have relied on octroi taxes, which public finance specialists have always referred to as bad taxes.[3] At the same time, cities keep user charges for water supply, sewage, solid waste disposal, and so on low, citing efforts to help the poor, but thereby starving these services of the investment they require. Correcting this situation needs city leaders who can generate resources and improve service delivery in a transparent manner. If this happens in a few cities, then the demonstration effect should spread to other cities. The legislative provisions for the exercise of such leadership now exist.

The 74th Constitutional Amendment Act of 1992 gave constitutional status to municipal governments, defined municipal responsibilities, and proposed the formation of state finance commissions to lay out the financial relationships between state and local governments. In many states, the devolution of administrative powers relating to planning, financing, and managing municipal services has not yet become effective. In addition, to augment resources for urban infrastructure, since 2001, 100 percent foreign direct investment has been permitted in the development of integrated townships.

Given the declining availability of state finances, urban local bodies are compelled to explore alternative sources of financing for urban infrastructure, such as municipal bonds, and have already done so in, for example, Ahmedabad, Bangalore, Ludhiana, and Vijayawada. However, markets need to be provided with incentives so that retail and institutional investors find that investing in such long-term instruments is attractive. Municipal bonds are among the most potent ways of raising resources for urban local bodies. Systems for obtaining credit ratings for urban local India are now in place, but the use of municipal bonds for raising resources is not yet widespread because of various problems related to rigid land laws and regulations, stamp duty, and the like, which inhibit urban infrastructure investment.

Forward-looking states such as Andhra Pradesh have achieved some success in terms of customer-oriented and transparent delivery mechanisms. Andhra Pradesh has taken the lead in introducing e-governance and has made considerable progress in this area. One of the state's main e-governance projects, e-Seva, provides multiple services to the public through integrated citizen service centers. These services include providing information, allowing people to make payments to government departments and agencies, dealing with tax returns, and providing certificates and licenses. Other states should adopt similar systems.

In the area of water supply, municipalities, state governments, and water boards initially showed considerable interest in attracting the private sector for funding, constructing, operating, and maintaining such facilities as bulk water treatment plants;

however, several projects were subsequently abandoned in Cochin, Hyderabad, and Pune because of various obstacles. For better service in cities, water supply and solid waste collection and disposal systems need to shift away from a government-only approach. Because user charges do not cover the costs of service provision, the colossal subsidy component has compelled various urban local bodies to engage in massive borrowing. There is no regulatory authority for the water sector that could make private projects less risky. Actual private participation in this area has, therefore, not yet succeeded.

The World Bank and the United Nations Development Programme recently conducted careful studies of the cost of urban water in India and the tariffs charged and found that all water consumers receive large subsidies. Whereas the average cost of water is about Rs 15 per cubic meter, the average tariff is about Rs 1.50 per cubic meter. Even industrial and commercial tariffs are significantly below cost. Because better-off customers, whether industrial, commercial, or domestic, typically consume larger volumes of water, they end up receiving larger effective subsidies than the poor, who consume much less water. Yet the prevailing view is that water cannot be priced at rational economic levels because then the poor could not afford it.

Urban local bodies that provide services are theoretically autonomous, but must still adhere to an extensive set of government regulations. The poor financial position of urban local bodies is the main constraint to the growth of urban infrastructure. In addition, considerable political interference affects operations, managerial decision-making, and tariff setting. Tariffs should be based on average incremental costs, including operations and maintenance charges, depreciation charges, and debt payments. The current institutional arrangements do not create the proper structures and incentives for improving the operational efficiency and quality of service. Problems relating to inadequate availability of information on the current financial and physical condition of existing service providers and their assets and the existence of tariffs well below cost recovery levels are the issues that concern international water companies as they attempt to invest in cities in India.

Urban local bodies, particularly small ones, are without the wherewithal to access capital markets to raise resources at competitive rates and lack the institutional capacity to manage the complexities and tasks involved in operating infrastructure services. Their main source of funds has been urban housing and development corporations and insurance companies. The problem is accentuated when urban local bodies are unable to levy sufficient user charges to service their debts and when projects are too small to be financed through debt-equity arrangements. One possible approach is the pooling method for financing multiple small projects that the European Bank for Reconstruction and Development has adopted in East European countries, which can save both transaction and borrowing costs. Recently, 14 urban local bodies in Tamil Nadu pooled a number of water and sanitation projects to float bonds. Currently more than 50 urban local bodies are experimenting with various kinds of arrangements to introduce private sector participation in solid waste management.

Conclusions

Infrastructure investment used to be a staid, regular, uninteresting activity that the public sector was involved in and that most of the population took for granted. The kind of active discussion, experimentation, and innovation that has taken place in the past decade is unprecedented in the area of infrastructure in India and elsewhere, but now is the time for consolidation and actual implementation. If India is to achieve the kind of economic growth rates that it aspires to, infrastructure must become even more of a priority than it has been.

That the public sector financing constraint is a serious one is an objective reality. The lesson is twofold. First, all constraints to private sector investment must be loosened so that the private sector can at least partially compensate for the lower-than-desirable level of public investment; and second, levels of public sector investment in infrastructure must be increased, because some infrastructure services are public goods and others exhibit partial public good characteristics. However, raising public sector investment in infrastructure to appropriate levels will not be feasible without fiscal improvements, particularly through revenue increases in both tax and nontax areas.

Rural Infrastructure

Growth in the agriculture sector has slowed significantly during the last five years (table 1). India will be unable to achieve annual growth of 7 to 8 percent over the next 5 to 10 years if agriculture continues to grow at its long-term trend rate of 2 to 3 percent annually. In the past, growth in agriculture has been equated with growth in food grain production, but with rising incomes, the average diet in both rural and urban areas is becoming increasingly diversified. Consequently, expenditure on food grains is falling as a proportion of total household expenditure on food. Thus higher agricultural growth will have to come from a much more diversified agriculture sector, as other rapidly growing Asian countries have found.

Agricultural diversification and accelerated agricultural growth will be difficult to achieve without much greater investment in rural infrastructure, such as roads, storage facilities, telecommunications, power, and the like. Diversified agriculture will also need much more complex commercial linkages between farms and markets. Thus investment in rural infrastructure will yield high economic returns, but developing the methodology to yield adequate financial returns is difficult. However, states such as Goa, Haryana, Kerala, Punjab, and Tamil Nadu, which invested in rural roads relatively early, have demonstrated how this can be done.

A key challenge for India in the coming years will be investing in and financing rural infrastructure. Given the difficult fiscal situation, innovation will be critical. New approaches to public-private partnerships, participation by local governments, and funds sourced from dedicated levies such as the fuel cess will all have to be explored. Politicians will also need to realize that continuing such policies as the low rural electricity tariffs will hinder the rapid development of rural infrastructure.

Regulation

Increases in private sector investment can take place on both an exclusive and a partnership basis. In sectors such as telecommunications, where service users can pay user charges at economic levels, there is no constraint, in principle, to exclusive private sector investment. The main area of policy concern in such sectors is the removal of regulatory risk. The rapid technological change that has characterized telecommunications over the past decade gave rise to unavoidable regulatory changes, which often trailed technology. New kinds of services became possible, and falling equipment prices often gave new entrants an advantage over incumbents, giving rise to disputes about tariff issues. Even though technological change is continuing to take place in telecommunications, the surge of innovation that took place in the 1990s has probably receded. The situation is similar in the power sector, where information technology made consumer choice and competition possible where previously none was thought feasible. Here too regulators have had to cope with new forms of organizational frameworks that require new kinds of regulatory intervention.

Regulatory risks that relate to predictability and transparency can be removed. The principles of regulation must be well thought out and transparently articulated. They must also be technically sound so that market participants will understand and accept them. If this is done, abrupt and ad hoc changes will be avoided and private investment will flow. India's various regulatory authorities are still in their infancy and have been handicapped by a lack of technical expertise at both the staff and authority levels. Thus many of their operations can be characterized as "learning by doing."

Staff at the authority level has typically consisted of previously retired civil servants, and employees at the staff level are generally temporarily seconded from government departments. Regulatory authorities have exhibited a marked reluctance to hire from the market. Regulatory authorities in other countries typically have a better mix of personnel that includes technical experts hired from the market along with a sprinkling of civil servants; however, with the imposition of public sector compensation structures, attracting appropriate expertise from the private sector is difficult at any level. Thus a key requirement is that regulatory authorities must be financially and administratively autonomous. The remuneration of their staff must be market related, but must also be transparent for reasons of accountability, as the government is ultimately responsible for the authorities' actions.

A good deal of regulatory risk arises from government or political interference in the work of the regulator, though it would be naive to argue that the regulator must be completely immune from political and government pressures. This problem can be reduced by the articulation of conditions and the manner in which the government can give directions to the regulator, which must be done transparently. In Sri Lanka, for example, if government direction results in operators losing revenues, the government must provide funds to make up for the losses.

Tariffs are both the most contentious of issues that regulators handle and the reason why regulators are created. Regulators are often set up specifically to remove

tariff setting from political pressures. Hence it is especially important for the government to devise procedures that allow regulators independence in tariff setting. This will be achieved only if there is better public understanding of the criticality of appropriate tariff setting.

Financing

The second issue related to private sector investment in infrastructure is the availability of sufficient and appropriate financing. Each infrastructure sector has its own characteristics, and financing mechanisms need to correspond to these characteristics. For example, once a power station or toll road has been built and tariffs have been set in a transparent and predictable manner, the cash flows are fairly regular and predictable. They can then be easily securitized. However, the pre-operational risks are extremely high, thus risk mitigation and credit enhancement are necessary to attract resources at reasonable cost. Market instruments need to be designed to meet these requirements, along with specific government interventions, where necessary, to mitigate pre-operational risks.

The 1990s witnessed extensive innovation in structured finance worldwide, aided by the huge expansion in cross-border flows of capital. This was also the period during which the former socialist economies embarked on their transition to market economies, all with large backlogs of infrastructure investment that had to be financed. Similarly, China, Indonesia, and other rapidly growing East Asian countries exhibited a large demand for infrastructure finance. However, the exuberance of international capital flows, particularly toward developing countries, has dampened since the 1990s and needs to be revived. As already noted, the initial expectations of returns to equity were perhaps excessive and were particularly unsuited to infrastructure investment. Thus international arrangements for cross-border financial flows for infrastructure investment need to be reexamined, something that the international financial institutions could do from the perspective of both investors and recipients. In some cases, large investments were made to privatize existing utilities, particularly in Latin America. In other cases, greenfield investments were made in areas such as telecommunications, power, and roads. Serious problems have arisen for investors where large changes in currency adjustments have taken place, as was characteristic of the 1990s. Equity investors have consequently realized lower returns than anticipated, while recipient countries have encountered debt servicing problems. Discussions related to the new financial architecture should pay attention to these issues. Mechanisms need to be developed to provide some protection to both potential investors and recipients, so that large financial flows for infrastructure investments can be rekindled. Mechanisms could be devised to enhance borrowers' creditworthiness and to ensure a minimum return to investment below which some form of investment insurance kicks in (with the minimum being well below the market return level to minimize moral hazard).

Various developments have taken place in India recently that should make it easier for resources to be intermediated toward infrastructure projects. The government

securities market has developed well, hence debt market benchmarks are now available. The market is now quite liquid, and government securities can also be traded on the stock exchanges. Thus the technological infrastructure for the debt market is in place in the stock exchanges. In addition, interest rate derivatives are about to be introduced to aid in risk management, and an act has been passed enabling the securitization of receivables on a widespread basis and strengthening creditors' rights.

The Infrastructure Development Finance Company was founded in 1997 as a joint venture between the government of India, the Reserve Bank of India, domestic financial institutions, and foreign investors like the Asian Development Bank and the International Finance Corporation. Thus institutional financing is available for infrastructure. Indeed, credit disbursed from the banking system for infrastructure increased from 1 percent of overall nonfood credit in March 1998 to 3 percent by the end of March 2003. The Infrastructure Development Finance Company had disbursed Rs 41.74 billion as of March 2003. Thus credit is available for viable infrastructure projects.

The equity market has been generally dormant since the mid-1990s; therefore, raising private equity has not been easy, particularly with the decline of the Unit Trust of India. Accordingly a new initiative, the India Development Fund, which is shortly expected to reach Rs 10 billion, has been set up to help private sector promoters raise equity. Furthermore, private sector insurance companies have begun operations, and the entry of independent pension funds has been announced. This means that a greater variety and volume of institutional investors will now become available for both direct and indirect investment in infrastructure.

A major development that has taken place in recent years is the significant and sustained reduction in inflation, both internationally and in India. When annual inflation was running at 7 to 10 percent, equity investors, especially international investors, expected 20 to 25 percent returns on risk capital. With inflationary expectations having fallen to some 3 to 5 percent, corresponding reductions should take place in the costs of both debt and equity. This should also make formerly unviable projects viable and hence amenable to commercial investment. The consequent reduction in user charges necessary for viability would also come down, leading to greater acceptability.

India also needs to learn more from the many different financing techniques employed in the industrial countries that effectively amount to private financing of public infrastructure or private-public partnerships. For example, in the United States, the federal government has made municipal bonds tax free, thereby enabling local governments to tap the large U.S. capital market to finance local urban and other infrastructure. Because the tax-free status of these bonds recognizes the public good element of urban infrastructure, local governments can raise finances at lower cost, thereby keeping user charges at affordable levels or municipal taxes at acceptable rates. The Pfandbrief system in Germany is another effective example of private financing of public infrastructure. In this case large mortgage banks pool local authority debt into pfandbriefs, which enhances the creditworthiness of entities that may not otherwise be creditworthy. Pfandbriefs have succeeded in maintaining their credit quality for more than a 100 years, through hyperinflation and two world wars, supported by a complex system of government guarantees. India and other

developing countries need to think about how they might make use of such innovative systems to generate greater private financing of infrastructure.

Public Sector Management

Even if all constraints to private sector investment are removed, the public sector will continue to maintain a significant presence in infrastructure. Most roads, except for toll roads, will necessarily remain in the public sector, as will sewerage systems, public lighting, and other similar components of urban infrastructure. Water supply systems, particularly distribution systems, are natural monopolies and will always have large public good elements. Water generation, purification, and treatment can be privatized, but the rest of the system will either remain in public sector hands or be heavily regulated. Similarly, as concerns railways, privatizing the common carrier (that is, the track), is difficult, whereas operating trains and carrying out maintenance can be privatized—although international experience with the private sector's operating the railways is, at best, mixed. In other areas, such as ports and airports, whereas activities such as operating terminals and handling traffic and airline operations are typically in the private sector, the public sector generally owns the actual ports and airports.

As a result, much more attention needs to be paid to public sector management. The public sector component of infrastructure must be commercially oriented and well managed, because infrastructure systems are typically large systems. Capabilities need to be built up to bring in modern management systems, and activities must attract the best managers. As happens in many systems, India's public sector and government management systems have become ossified with excessive rigidities and built-in careerism. Entry is essentially at the basic entry level, with little or no mobility of personnel at higher levels and little infusion of new blood at higher levels; thus inward-looking attitudes have become the norm. What is needed is a system that will bring in outside expertise at all levels in an organized framework; however, labor mobility is hampered by the rigidities in the social security system whereby pensions and other benefits are not portable.

In India, which now has a mix of public and private companies operating in many of the same sectors, great inequalities in compensation have arisen between people performing similar functions. This issue needs to be addressed.

A major issue in public sector management is the tenure of chief executives. In Indian Railways, for example, the chief executive's tenure is seldom longer than one year. This means that chief executives can hardly be expected to have even a medium-term vision, let alone a long-term one, and even if they did, they certainly would have no chance of implementing it. In addition, the system of promotion by seniority is not conducive to commercial operation.

Public-Private Partnerships

Given that most infrastructure sectors exhibit some public good characteristics, public-private partnerships can maximize investment. If the achievable return to an

infrastructure activity through user charges is, say, 70 percent of market returns, a public subsidy of 30 percent would elicit market-based private investment, and the enterprise could then be run on commercial basis. One example of such an activity has already been given in the case of toll roads. Another example could be the urban water supply, where some water could be provided to the poor through public stand-posts. Such partnerships are not, however, easy to undertake in a democratic parliamentary framework: inevitably, questions arise about the award of concessions and contracts. Hence a great deal of work needs to be done to devise methodologies for the development of public-private partnerships.

Notes

1. Captive ports are ports dedicated to the use of one particular firm (for example, a fertilizer project or oil refinery). Captive facilities in general are dedicated facilities meant for specific user industries and not open to other users.
2. In India, civil service compensation is reviewed and reset roughly once every 10 years by a pay commission appointed for that purpose by the central government. The last such commission announced its recommended changes in 1997.
3. *Octroi* taxes are local taxes levied on goods as they enter a municipal jurisdiction.

References

Central Statistical Organisation. Various issues. *National Accounts Statistics*. New Delhi: Government of India.

Expert Group on the Commercialisation of Infrastructure Projects. 1996. *The India Infrastructure Report: Policy Imperatives for Growth and Welfare*. New Delhi: National Council of Applied Economic Research.

Expert Group on Indian Railways. 2001. *The Indian Railways Report: Policy Imperatives for Reinvention and Growth*. New Delhi: National Council of Applied Economic Research.

Ministry of Finance. Various issues. *Economic Survey*. New Delhi: Government of India.

Ministry of Power. Various issues. *Annual Report*. New Delhi: Government of India.

Ministry of Statistics and Programme Implementation. Various years. *Annual Report*. New Delhi: Government of India.

Telecommunications Regulatory Authority of India. Various years. *Annual Report*. New Delhi: Government of India.

Entrepreneurship, Innovation, and Growth

T. N. SRINIVASAN

The three basic sources of growth in any economy are growth in inputs of production; improvements in the efficiency of allocation of inputs across economic activities; and innovation that creates new products, devises new uses for existing products, and increases the efficiency of input use. Solow's path-breaking analysis of growth in the U.S. economy demonstrated that the share of growth attributable to innovation is substantial, although it varies among countries. In this context, the paper has four broad themes: the importance of innovation to growth; the importance of entrepreneurship to innovation; the role of financial intermediaries, particularly venture capital, in fostering entrepreneurship; and the influence of public policies relating to research and development, as well as other influences of policies, such as foreign trade policies, on productivity growth.

Introduction

The three basic sources of growth in any economy are growth in inputs of production; improvements in the efficiency of allocation of inputs across economic activities; and innovation that generates new products, devises new uses for existing products, and increases the efficiency of input use.[1] Solow's (1957) path-breaking analysis of growth in the U.S. economy during the first half of the 20th century showed that the contribution of growth in inputs of production, namely, labor and capital, to aggregate growth is around half; the remaining half, that is, the unexplained Solow residual, is commonly attributed to technical progress or the contribution of innovation. As, by definition, the residual growth is the difference between

Annual World Bank Conference on Development Economics 2004
© 2004 The International Bank for Reconstruction and Development / The World Bank

aggregate growth and the contributions to growth of factors of production, it is also called total factor productivity (TFP) growth.

Economists have debated whether or not TFP growth accounts for a similarly large share of output in the East Asian economies of the Republic of Korea, Singapore, and Taiwan (China) (Lau and Kim 1994; Young 1992, 1995). Their arguments have been based in part on methodological grounds (for example, the dependence of estimates on essentially arbitrary assumptions about scale economies and functional forms for the aggregate production function) and in part on grounds of possible errors of measurement and biases in the data (Pack 2001).

This paper does not enter into the debates on the empirics of estimating TFP growth, except to note that despite all its problems, it is a useful summary measure of the outcome of innovation or the lack thereof. For this reason, TFP estimates are used later in this paper on the assumption that TFP growth contributes to economic growth and is driven by useful ideas or increases in the stock of useful knowledge, as Lucas (2002) puts it. As he rightly points out, treating growth in technology or knowledge as exogenous for an economy implies that it originates from activities outside that economy. In other words, what is external to one economy must be internal or endogenous in some other economy. One is thus led to modeling the endogenous process that generates technological change in that economy. Even if one were to assume that for developing economies technical progress comes about through the adoption or imitation of technologies developed elsewhere, the adoption and imitation processes themselves could be endogenous. Therefore viewing processes that lead to innovation as endogenous processes that are influenced by incentives and institutions, as well as by the political economy, is essential.

This paper will deal with four broad themes. The first is the importance of innovation to economic growth. The next section reports on the analytics of the contribution of innovation to growth, selectively drawing from the vast recent literature on endogenous growth models.[2] An essential element of these models is that the endogeneity of growth is driven in part by the endogeneity of innovation. An earlier literature dealt with endogenizing the innovation process (Boserup 1965; Kaldor and Mirrlees 1962). A major strand of this literature goes under the rubric of induced innovation (see Ruttan 2001 for a recent treatment). I will not report on this body of work primarily because in the models of this genre, the market for innovation (or, more precisely, the process of compensation of innovators in contexts where benefits of innovation have spillover effects) and the link between innovation and growth are not transparent. An alternative to the induced innovation model is Arrow's (1962) celebrated learning-by-doing model. In this model, innovation as reflected in increases in labor productivity was the unintended or serendipitous effect of production. It was external to the firms who did the producing. Thus, strictly speaking, the process of innovation through learning-by-doing in the Arrow model is exogenous. Nordhaus (1969) and Shell (1973) endogenize innovation by modeling the incentives to innovate as arising from the prospect of monopoly rents from patented innovations.

Models of pure learning rule out sustained growth effects if there are diminishing marginal returns to learning in any single activity. For sustaining growth, new

activities with learning potential have to emerge and be adopted all the time, as in Lucas's (2002, chapter 3) model. The adoption of new products could be endogenous, thereby endogenizing learning and hence growth. However, learning processes involve an inherent ambiguity. As Lucas (2002, p. 84) points out:

> Is it the individual worker who is doing the learning? The managers? The organization as a whole? Are the skills being learned specific to the production process on which the learning takes place, or are they more general? Does learning accrue solely to the individual worker, manager, or organization that does the producing, or is some of it readily appropriable by outside observers?

The second broad theme is the importance of entrepreneurship to innovation. The dictionary meaning of the word entrepreneur is a person "who organizes, operates, and assumes the risk for a business venture" (*American Heritage Dictionary of the English Language* 2000, Fourth Edition) and "one who undertakes an enterprise; one who owns and manages a business; a person who takes the risk of profit or loss" (*Oxford English Dictionary* April 25, 2003, online database). Neither definition includes what economists usually think of as the essential characteristic of an entrepreneur, namely, the source of a new, potentially commercially viable idea. Entrepreneurs themselves, or others to whom they license or sell the idea, implement it in a productive enterprise. In this sense, it is not so much the generation of ideas, but their implementation in an enterprise, that is the center of attention from the perspective of growth.

The importance of resources needed to implement viable ideas leads to the paper's third broad theme: the role of financial intermediaries in fostering entrepreneurship. In particular, the ease of access to capital or finance and the chances of originators' ideas being appropriated or stolen if they do not have complete control over the enterprise are important issues. A particular form of organization, venture capital, has proved to be significant in financing high-tech start-up enterprises in the United States, but in other major industrial countries, such as Germany and Japan, venture capital is less important. Indeed, the equity market as a source of capital and as a market for corporate control is less developed in those countries compared with the United States. An important dimension of venture capital is that its return is mostly from the sale of equity at the time the start-up firm makes its initial public offering. Unless a deep equity market exists, venture capitalists will not have an attractive exit option through initial public offerings from the start-up firms they finance. The absence of a significantly deep equity market might explain in part why venture capital has not become important in Germany and Japan (Black and Gibson 1998). Also the U.S. government has supported start-ups through its Small Business Innovation Research Awards and Small Business Investment Company. The third section of the paper discusses entrepreneurship and the role of venture capital and other forms of financing for entrepreneurs.

Given that innovation is important to growth, entrepreneurship is important to innovation, and finance is crucial for entrepreneurial success, policy becomes an issue. Are there specific policies that developing countries can adopt so as to accelerate

growth by encouraging innovation? To what extent do other policies that influence the environment for resource allocation and accumulation affect innovation? Should governments themselves undertake and fund research that ultimately produces potentially commercially viable ideas? These policy issues compose the paper's fourth broad theme. Thus the fourth section of the paper is devoted to publicly funded research systems and to empirical findings on the importance of foreign trade and investment as mechanisms for transferring innovation-induced productivity gains across countries. The paper then wraps up with concluding remarks.

Some Illustrative Models of Entrepreneurship, Innovation, Imitation, and Growth

The literature on the economics of research, technological change, and innovation is vast. For example, between 1994 and 2003 just one publisher, Kluvwer Academic Publishers, brought out as many as 28 volumes on the theme of the economics of science, technology, and innovation. This is not the occasion to survey the vast literature, so let us focus on a few of the contributions.

Grossman and Helpman (1994) provide an accessible survey of endogenous innovation in the theory of growth drawing on their classic book (Grossman and Helpman 1992). Even though several contributions have been made to the literature since then (some of which are included in Aghion and Howitt 1998), for my expository purposes, concentrating on the simplest general equilibrium model of innovation and growth presented in Helpman (1990) is sufficient.[3] He considers a model in which a continuum of differentiated brands of consumer goods is produced only by labor. The products enter symmetrically in consumers' utility function. However, to produce each new brand, producers have to incur expenditure on product development. Product development or research and development (R&D) uses only labor. Once entrepreneurs have developed a brand, they enjoy indefinite monopoly power over its production and sale. Entry into product development is free. Then at each instant of time t, product development costs $c_n[w(t)]$ must equal the present value of future profits $\pi(\tau)$, where $w(t)$ is the wage rate and $n(t)$ is the number of brands in existence at t. Thus

$$c_n[w(t)] = \int_t^\infty \exp[R(t) - R(\tau)]\pi(\tau)d\tau, \qquad (1)$$

where the discount factor $R(\tau) = \int_0^\infty r(u)du$, with $r(u)$ being the interest rate. Differentiating both sides of equation (1) with respect to t, rearranging, and denoting by a dot over a variable, its derivative with respect to t:

$$\frac{\pi(t)}{c_n} + \frac{\dot{c}_n}{c_n} = \dot{R}(t). \qquad (2)$$

Consider a symmetric equilibrium in which the same amount of all $n(t)$ brands available at t are consumed and all brands are sold at the same price $p(t)$. With a constant elasticity σ of substitution among brands in the instantaneous utility function, utility $u(t)$ of the representative consumer (all consumers are identical) would be given by

$$u(t) = \frac{1}{\sigma - 1} Log n(t) + Log E(t) - Log p(t), \tag{3}$$

where $E(t)$ is expenditure on the $n(t)$ brands available at time t.

The consumer maximizes the discounted present value of $u(t)$, that is,

$$\int_0^\infty u(t)\exp(-\rho t)dt, \tag{4}$$

subject to the wealth constraint,

$$\int_0^\infty E(t)\exp(-R(t))dt = \text{wealth}, \tag{5}$$

by choosing $E(t)$, taking $n(t)$, $p(t)$, and $r(t)$ as given.

This maximization can be shown to imply that

$$\frac{\dot{E}(t)}{E(t)} = \dot{R}(t) - \rho. \tag{6}$$

Turning to the supply side, a unit of each brand requires a_{Lx} units of labor to produce. For producing a brand, a blueprint or design for production has to be developed, so that each new brand of product requires a_{Ln}/K units of labor for developing its design, where K is knowledge or R&D stock. Thus the larger this stock, the lower the labor per design produced (that is, per unit of R&D output). The model captures two features of innovation. The first is that innovation is driven by the expectation of future profits. This is exemplified by the requirement that the cost of development of a product equals the discounted present value of future monopoly profits from its production and sales, given free entry into the product development industry (see equation [1]). The second is the unintentional and uninternalized consequence of past product (that is, design) development for the cost of future design development. This can be viewed as a consequence of learning by doing in the product development or R&D industry along the lines of Arrow (1962). The benefits from lowered development costs accruing to future product developers does not enter the cost-benefit calculus (equation [1]) of present product developers. Thus these benefits do not accrue to, and hence are not internalized by, them. This leads to their undersupplying new product designs in a private market equilibrium relative to a social optimum.

Suppose the stock $K(t)$ of R&D reflects learning by doing according to $\dot{K} = n(t)$. Then by a proper choice of units we can set $K(t) = n(t)$.

Now the cost $c_n[w(t)]$ of each new brand design is $[a_{Ln}/n(t)]\, w(t)$. If we choose a new design as the numéraire, then free entry into the R&D sector will ensure zero profits in that sector, so that

$$1 \equiv c_n[w(t)] = \frac{wa_{Ln}}{n}, \tag{7}$$

omitting the time argument t from w and n. Under monopolistic competition in the symmetric equilibrium, the price $p(t)$ of a unit of output of each brand will equal its marginal cost $w(t)a_{Lx}$ times the markup or

$$p(t) = \left(\frac{\sigma}{\sigma - 1}\right) wa_{Lx}, \tag{8}$$

where σ (the elasticity of substitution in consumption) is the elasticity of demand.[4] Denoting $(\sigma - 1)/\sigma$ as α, we can rewrite

$$\alpha p = wa_{Lx}. \tag{9}$$

Equations (7) and (8) imply that the wage rate $w(t)$ and the price $p(t)$ grow at the same rate as $n(t)$. The factor market clearance condition, given that $\dot{n}(t)$ units of new brands are being developed, and denoting by $x(t)$ the aggregate output of all brands together, yields

$$[a_{Ln}/n]\dot{n} + a_{Lx}x(t) = L(t). \tag{10}$$

For simplicity, let us assume that $L(t)$ is a constant L. Aggregate consumer expenditure $E(t)$, by definition, equals $p(t)x(t)$. Substituting for $x(t)$ and using equations (7) and (9) in equation (10) we get:

$$\frac{\dot{n}}{n} = [L/a_{Ln}] - \alpha\eta, \tag{11}$$

where $\eta = E/n =$ expenditure per brand.

Using equation (9), profits $\pi(t)$ per brand equal price $p(t)$ minus cost wa_{Lx} times output x/n or

$$\pi(t) = (p - a_{Lx}w)\frac{x}{n} = (1 - \alpha)\frac{px}{n} = (1 - \alpha)\eta. \tag{12}$$

Because $c_n(t) \equiv 1$ from equation (7), it follows from equation (2) that $\pi(t) = \dot{R}(t)$. Using equation (6), it follows that

$$\dot{R}(t) = \frac{\dot{E}}{E} + \rho = \pi(t) = (1 - \alpha)\eta. \tag{13}$$

Because $\eta \equiv E/n$, it follows from equation (13) that

$$\frac{\dot{\eta}}{\eta} + \frac{\dot{n}}{n} + \rho = (1 - \alpha)\eta. \qquad (14)$$

Substituting $\dot{n}/n$ from equation (11) in equation (14), we get

$$\frac{\dot{\eta}}{\eta} = \eta - \rho - [L/a_{Ln}]. \qquad (15)$$

The only solution to the differential equation (15) that satisfies the transversality condition for maximizing intertemporal consumer welfare is the stationary solution $\dot{\eta} = 0$. This in turn means—using equations (11) and (15) and setting $\dot{\eta} = 0$—that the growth rate g of the stock of brands is given by

$$g = \frac{\dot{n}}{n} = (1 - \alpha)[L/a_{Ln}] - \alpha\rho. \qquad (16)$$

It is clear from equation (16) that g is higher if (i) the labor needed per unit of R&D (as indexed by a_{Ln}) is lower; (ii) the markup over marginal costs (that is, the value of $(1/\alpha)$) is higher; (iii) the impatience of consumers (as indexed by ρ) is lower; and (iv) the size of the labor force L is larger. This is not realistic, because it implies that R&D would grow faster in large, populous countries such as China and India. However, this unrealistic feature of the model can be and has been remedied (Matsuyama 1992) without weakening the other results.

The utility $u(t)$ at any point in time t of the representative consumer of this economy is given by equation (3). In the steady-state path, $n(t)$ grows at the rate g. Expenditure per worker is the wage rate w. Substituting in equation (3),

$$u(t) = Log\left[n(t)^{1/(\sigma-1)}\left(\frac{w}{p}\right)\right] = Log\left[n(t)^{(1-\alpha)/\alpha}\left(\frac{w}{p}\right)\right]. \qquad (17)$$

Because w/p is constant over time (see equation [9]) and $n(t)$ grows at the rate g, the representative consumer's utility $u(t)$ rises linearly over time at the rate $(1-\alpha)g/\alpha$. Thus consumers are better off living in a large economy that starts with the same number of brands of consumer goods than in a smaller economy, because their initial utility is the same in both countries but rises faster (that is, g is larger) in the larger country.

Helpman (1990) extends this model of innovation in a closed economy to a world of two countries, North and South. North is an industrial country and is where innovation takes place. However, Northern entrepreneurs know that they are not assured of indefinite monopoly, as they would have been were North a closed economy, because of the threat of imitation by the less developed South. Suppose at time t the total number of brands in existence and invented in North is $n(t)$, of which n_S have

already been imitated by South and $n_N = n - n_S$ have yet to be imitated. Let $\mu = \dot{n}_S/n_N$ be the endogenous instantaneous rate of imitation, with every Northern brand yet to be imitated having the same chance of being imitated. Let $F(t, \tau)$ be the probability that a brand developed at time $\pm$ will be imitated by South before $\tau \geqslant t$. Then

$$F(t, \tau) = [1 - e^{-\mu(\tau - t)}]. \tag{18}$$

Northern entrepreneurs maximize the expected present value of their profits, given $F(t, \tau)$ and given that they lose their market to Southern imitators once they have been imitated. Free entry into the R&D sector in North leads to the following analogue of equation (1):

$$c_n[w_N(t)] = \int_t^\infty [1 - F(t, \tau)] \left[\int_t^\tau \exp[R(t) - R(u)]\pi(u)du \right] dF(t, \tau)d\tau, \tag{19}$$

where $w_N(t)$ is the Northern wage rate.

Substituting for $F(t, \tau)$ and differentiating we get the analogue of equation (2)

$$\frac{\pi}{c_n} + \frac{\dot{c}_n}{c_n} = \dot{R} + \mu. \tag{20}$$

Comparing equation (2) with equation (19) reveals that the threat of imitation adds a risk premium μ to the interest rate $\dot{R}$, which the sum of the instantaneous rate of profit (π/c_n) and capital gains ($\dot{c}_n/c_n$) has to equal.

We now turn to the Southern process of imitation closely following Helpman (1990, pp. 31–33). Suppose that imitating a brand takes resources. Specifically, a Southern entrepreneur needs a_{LI}/n_S units of labor per brand to imitate a product that has yet to be imitated, given that n_S products have already been imitated. In other words, the stock of knowledge capital in South is the number n_S of products already imitated. Having imitated a variety, Southern imitators need $a_{L\chi}$ units of labor per unit of output in manufacturing (just like in North). Imitation takes place only if the present value of profits covers imitation costs. For imitators, however, the profit calculation is more involved. If they did not face competition from the original Northern innovators, they would mark up the price above marginal costs in the usual way. When the resulting price falls short of Northern marginal manufacturing costs, they can still charge this price without being threatened by Northern producers. This happens when South's wage rate is lower than the proportion $\propto$ of North's wage rate, and this is termed the wide gap case (the gap in relative wages is wide). Otherwise, Southern imitators charge a price that equals North's marginal manufacturing costs. Naturally, the imitators would lose money in either case if the wage rate were lower in North, so that active imitation requires a lower wage rate in South, and this is assumed hereafter. Free entry into imitation implies a no-arbitrage condition such as equation (2).

The labor market clearing in South (the analogue of equation [10]) implies that

$$a_{LI}\dot{n}_S/n_S + a_{Lx}x_S = L_S. \tag{21}$$

Now assume the wide gap case, so that South's pricing equations are similar to those of North. Together with the market clearing, no-arbitrage conditions, and the growth of spending equation (in which the subjective discount rate is the same in both countries), the pricing equations imply a steady-state growth equation that is analogous to equation (16), namely:

$$g = (1 - \alpha)L_S/a_{LI} - \alpha\rho. \tag{22}$$

A similar procedure for North, using equation (20), yields an equilibrium steady-state relationship between the rate of innovation and the rate of imitation as follows:

$$(1 - \alpha)(L_S/a_{LN} - g)(g + \mu)/\alpha g = g + \mu + \rho. \tag{23}$$

The left-hand side of equation (23) represents the profit rate, while the right-hand side represents the cost of capital, that is, the interest rate plus the risk premium. As g increases, the right-hand side increases, while the left-hand side declines. Therefore an increase in the rate of innovation g reduces profitability relative to the cost of capital. Alternatively, an increase in μ raises the right-hand side, but raises the left-hand side even more. Thus an increase in the rate of imitation increases the profitability of innovation relative to the capital cost. This explains the upward slope of curve NN in figure 1 along which equation (23) holds. As equation (22) holds along SS, the equilibrium levels of innovation and imitation are given by the intersection point 1.

Several implications of this model are worthy of note. First, observe that if innovation in South requires more resources than imitation, which is reasonable, then trade with North speeds up long-run growth in South. This can be seen from equation (22). Without trade, the growth equation is the same (see equation [16] for a closed economy), except that a_{LI} is replaced by a larger coefficient. Second, trade with South speeds up long-run growth in North. This is shown in figure 1 by the fact that the vertical intercept of NN identifies the autarky growth rate, so that both countries grow faster by trading with each other. Third, a larger South raises both the rate of innovation and the rate of imitation. A larger North, by contrast, does not affect the rate of innovation but reduces the rate of imitation. A lower rate of imitation is associated with longer average time periods during which Northern entrepreneurs command monopoly power. Fourth, the larger the country, the larger its relative wage rate.

Turning to policy implications, in the single country model, as noted earlier, growth of innovation (and hence growth of utility of the representative consumer) under laissez faire is not socially optimal, because private product developers do not take into account that by developing a product today, they generate a positive externality for

FIGURE 1.
North-South Product Cycle
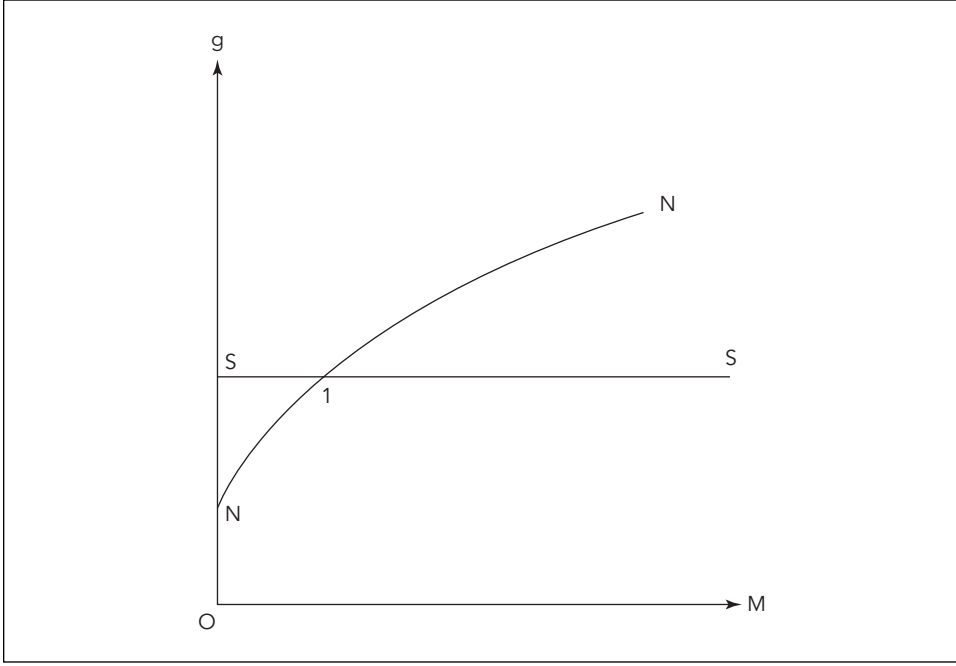
product developers in the future, because each product developed today adds to knowledge capital that reduces future product development costs. For this reason, some degree of growth promotion through subsidies for innovation is desirable, but how much?

Social planners of this economy would maximize the discounted present value of aggregate welfare subject to the constraint of aggregate labor availability (see equation [10]) $a_{Ln}g + a_{Lx}x = L$ through their choice of g and x. Aggregate consumption per brand is x/n and aggregate utility U is easily seen to be

$$U = \frac{1}{\sigma - 1}Logn + Logx = \left[\frac{1 - \alpha}{\alpha}\right]Logn + Logx. \tag{24}$$

Now

$$n = n_o e^{gt} \quad \text{and} \quad x = \frac{(L - a_{Ln}g)}{a_{Lx}}.$$

Thus the discounted sum of utility is

$$W = \int_o^\infty e^{-\rho t}U(t)dt = \frac{1}{\rho}\left[Log\left(\frac{L - a_{Ln}g}{a_{LX}}\right) + Logn_o\right] + \frac{g(1 - \alpha)}{\alpha\rho^2}. \tag{25}$$

Maximizing W, or equivalently ρW, leads to maximizing $Log(L - a_{Ln}g) + g(1 - \alpha)/\alpha\rho$ with respect to g. Thus the socially optimal g^* is

$$g^* = \frac{L}{a_{Ln}} - \left(\frac{\alpha}{1 - \alpha}\right)\rho. \tag{26}$$

Comparing g^* with the laissez faire market optimum g given by equation (16) shows that

$$g^* = \frac{g}{1 - \alpha} > g. \tag{27}$$

Hence subsidies that bring market optimum growth closer to the social optimum g^* are worthwhile, but further increases in the subsidy would reduce social welfare.

Helpman (1990) points out that foreign trade policy could be used to accelerate long-run economic growth, for example, if it succeeds in shifting resources toward innovation. However, the particular policy used would matter, because some policies are more prone to unleashing distortionary rent-seeking. Thus

> [e]ven in cases where the free-trade growth rate falls short of the optimal level and trade policy accelerates growth, trade policy may nevertheless be harmful. Monopolistic competition per se introduces a distortion that can be aggravated by a growth-enhancing trade policy. In addition, in the presence of rent seeking, growth is slower under quotas than under tariffs because quotas divert resources to rent seeking, thereby reducing employment in R&D. This effect is particularly strong when rent seeking uses entrepreneurial skills that are useful in product development. These examples highlight the role of policy in a dynamic context (Helpman 1990, p. 35).

Lucas (2002, chapter 3) describes a model in which learning occurs in individual product lines. New goods or product lines, which are better in a well-defined sense than other goods in existence, are introduced, and labor is reallocated between new and old goods all the time. The rate at which the new goods are introduced is endogenous. As in the Helpman model, the assumption is that the learning accumulated in producing a good reduces the cost of producing each good that is introduced later. Thus the spillover effect is loaded in the direction of improving the productivity of better goods introduced later. The spillover decays in the sense that the initial productivity of a good is a weighted average of learning on lower-quality goods, with the weights declining exponentially with the difference in quality. The equilibrium rate of the introduction of new goods is inversely proportional to the rate of decay of spillover experience, an increasing function of the spillover parameter and the learning rate, and increases as employment is more heavily concentrated on more advanced or better goods.

Lucas finds this learning spillover model attractive for at least two reasons, the main one being that it offers the potential to account for the great differences in productivity growth rates that are observed among low- and middle-income economies, although he admits that little is known empirically about crucial spillover parameters

of the model. The second reason is that the model is consistent with a strong connection between rapid productivity growth and trade or openness. For example, in two small economies facing the same world prices and with similar factor endowments, suppose that one somehow shifts its workforce into producing goods not formerly produced there, and continues to do so, while the other continues to produce its traditional goods. The learning spillover theory implies that the economy that shifts its labor force would grow more rapidly. For this to be plausible, the shifting economy has to open up a larger difference between the mix of goods it produced and the mix it consumes (recall that both economies are assumed to have similar incomes and tastes and to face the same world prices). Thus a large volume of trade is essential for a learning-based growth episode.

This reasoning can also show why import substitution policies ultimately fail, although they might initially succeed in stimulating growth. As Lucas rightly argues, an economy that exports agricultural products and imports manufactures could set up prohibitive trade barriers that succeed in closing the economy and in shifting workers to producing formerly imported manufactures, and indeed, rapid learning will occur initially. But this is just "a one-time stimulus to productivity, and thereafter the mix of goods produced in this closed system can change only slowly, as the consumption mix changes" (Lucas 2002, p. 94).

Thus there is no room for opening a large gap between what is produced and what is consumed in a relatively closed economy. This observation by Lucas is broadly consistent with the experiences of China and India after they opened up to foreign trade and investment, compared with the pre-reform period when their economies were insulated from world markets. The rate of TFP growth in China during the pre-reform period of 1953–78 was 1.1 percent per year, whereas during 1979–94, it jumped to 3.9 percent per year (Hu and Khan 1997). Various, and differing, estimates of TFP growth are available for India, but they all confirm that once India began to liberalize and open its economy, hesitantly and to a limited extent in the 1980s, and purposefully and across the board in the 1990s, TFP growth accelerated (Srinivasan forthcoming, table 7.6).

Others have modeled entrepreneurship in different ways. Consider a particularly insightful model by Acemoglu, Aghion, and Zilibotti (2003). The authors distinguish between financing and management, with managerial selection being important from the perspective of innovation, and draw out the implications of alternative development strategies for innovation. They consider an economy where firms owned by capitalists are run by managers hired by capitalists. Managers are of two types, high- and low-skill, and perform two tasks: engage in innovation, a task for which skills are important for success, and adopt technologies from the world frontier, a task for which skills are not as important. Each firm also invests, choosing between large and small levels of investment. Investment costs are financed either through retained earnings or borrowing from competitive intermediaries. A moral hazard problem exists in that managers could divert a fraction of firms' returns for their own use without being prosecuted. Capitalists make contracts with managers and financial intermediaries, with the contracts specifying the loan amount from

intermediaries, the payments to managers and intermediaries, and the level of investment.

In its dynamic equilibrium path, an economy starts with an investment-based strategy (high investment) and long-term relationships between firms and managers as it approaches the world technology frontier. A switch occurs to an innovation-based strategy with lower investment, shorter relationships, younger firms, and more managerial selection. However, depending on parameter values, an economy could fall into one of two traps: a stagnation trap, where the economy starts too far below and progressively falls further and further behind the world technology frontier, and a nonconvergence trap, where the economy grows at the same rate as the world frontier without ever converging to it. This occurs when the economy fails to switch out of the investment-based strategy. The authors show that the switch out of the investment-based strategy may occur too soon because firms do not internalize the greater consumer surplus from higher investment, or too late because the presence of retained earnings enables managers not to have to compete for investment funds, and thus shields them from competition. This prolongs the investment-based strategy.

Acemoglu, Aghion, and Zilibotti (2003) cite several policy implications. First, when the switch is too soon, policies restricting competition or subsidizing invest-ment would slow the switch. However, in situations where retained earnings shield insiders too much and economies never switch from the investment-based strategy (and hence do not converge to the world frontier), such policies might lead to non-convergence traps. The authors recognize that while a welfare-maximizing policy sequence consists of a set of policies that at early stages of development would encourage investment and protect insiders, thereby hindering competition, and at later stages would promote competition, such a sequence creates serious political economy problems. For example, as has occurred in economies that have been pro-tected from domestic and import competition through public policy interventions, as in India, beneficiaries of such policies would resist reforms. Thus a well-meaning attempt to encourage domestic investment at early stages of development through inward-oriented and anticompetition policies could end up preventing the economy from adopting policies that would speed up growth at later stages.

Entrepreneurship and Finance

Innovative ideas, successfully commercialized and implemented, are important sources of development and growth. Incentives for the generation of ideas and insti-tutions that evaluate them and finance those that are commercially viable in a way that adequately compensates innovators and suppliers of capital (while appropriately assigning the unavoidable risks of investment in new ideas to those most capable of bearing them) are necessary for entrepreneurship to succeed as a source of ideas and an engine of growth. Whether the public sector has to play a large role in developing such institutions and in providing finance, or whether private financiers such as

venture capitalists would largely suffice, depends on the stage of development of a country (particularly of its financial markets).

Role of Venture Capital

In the models presented in the previous section, innovators were directly rewarded by monopoly profits or managerial compensation. Financing of innovation was not an issue, and financial intermediaries had no role to play in the innovation process. There is no doubt that in reality, innovation also comes about from the implementation of ideas generated by entrepreneurs who lack finance being able to get it from those who can provide it. The problem of finance is particularly acute for start-up firms planning to use as yet unproven technologies. By definition, having just started operations or being about to do so, they cannot count on an established reputation in the market for obtaining finance at affordable costs. Also entrepreneurs, who are good at developing a potentially highly profitable product or process, may not have adequate managerial and sales skills. Thus for successful commercialization of entrepreneurs' idea, they may need more than finance from financiers. Alternative means of financing, such as issuing equity or debt in capital markets, borrowing from banks and other financial intermediaries, obtaining venture capital, and so on, are not all likely to be accessible on similar terms to entrepreneurs. Entrepreneurs' access to nonfinancial services will also differ. The informational asymmetries, reputation effects, moral hazard, and adverse selection that arise in financial markets are likely to be particularly severe for entrepreneurial finance.

Venture capitalists are important intermediaries, because they typically finance small and young firms with few tangible assets other than their products or processes (about which the firms know more than potential investors), and which often operate in rapidly changing markets. Venture capital organizations, unlike other financial intermediaries, have developed many mechanisms for overcoming the problems in financing such potentially high-return but high-risk firms (Gompers and Lerner 2001). Before turning to these mechanisms, consider a theoretical model developed by Amador and Landier (2002).

Amador and Landier, like Acemoglu, Aghion, and Zilibotti (2003), put firm managers along with financing constraints at the center of the innovation-implementation process. In Amador and Landier's model, managers are the ones who generate ideas for projects. They can implement their ideas either inside their firms or be financed outside with venture capital. Firms, because they own assets complementary to the project, could face a lower cost of implementation. On the other hand, a venture capital organization could offer a contract to the manager that is contingent on the project's outcome. The owner of the firm, that is, the capitalist, as in Acemoglu, Aghion, and Zilibotti (2003), will let the manager implement it within the firm if the cost of compensating the manager to do so does not exceed the cost of the competition the firm would face if the manager implements the idea outside the firm. The manager would opt for outside implementation if the expected payoff from doing so would exceed what the capitalist would be willing to offer to implement the project inside the firm.

The venture capital market influences these decisions by affecting both the projects that can be implemented outside and the payoff to managers to acquire new ideas. In the presence of asymmetric information, moral hazard, and differences in beliefs about the probable success of projects, Amador and Landier show that the most innovative projects are implemented in new ventures and that more focused firms innovate more. Also, if the marginal innovation is implemented under the outside threat, a better venture capital market increases the innovation rate. If the project would have been implemented even without the threat, a better venture capital market would reduce the innovation rate by reducing the firm's payoffs.

Lerner (1998) provides a succinct description of the problems start-up firms face, particularly firms in high-technology industries, and the role of venture capital and outside investors in addressing them. Typically such firms are characterized by significant uncertainties and information asymmetries that could result in opportunistic behavior by entrepreneurs. Serious incentive problems arise from conflicts between firm managers and investors, and these could affect the supply of debt and equity capital to firms. For example, with equity financing, managers have an incentive to engage in disproportionately (to them) beneficial but wasteful expenditures while bearing only part of their costs. Debt financing might induce them to undertake excessively risky activities. Because outside investors are aware of these incentives, they would demand a higher rate of return than the cost of funds internally generated by the firm. Even if managers are driven to maximize shareholder value, if they are better informed about investment opportunities and act in the interest of current shareholders, they would issue new shares only when the company's stock was overvalued. The information asymmetries and their consequences in debt markets have been well known since the work of Stiglitz and Weiss (1981).

Venture capital organizations address these information problems and alleviate capital constraints by scrutinizing firms before providing capital and by monitoring them afterward. The initial scrutiny is intense, with only a small proportion of firms seeking finance being funded. The funding is often syndicated, with more than one syndication partner having to approve funding before it is disbursed. In addition, approved funding is disbursed in stages, so that managers frequently have to go back to their financiers, thereby enabling the financiers to ensure that their funds are not being wasted. Finally, venture capitalists monitor managers intensively through their insistence on representation on the board of directors, preferred stock, and restrictive covenants on finance. For all these reasons, venture capital is the dominant form of financing for privately held, technology-intensive firms.

Developing a successful product requires heavy upfront investments in development and branding, which are facilitated if venture capitalists are present. The quality of venture capital, and not merely the quantity, is important. Interestingly, even though the Indian software industry in Bangalore has been extremely successful, the absence of a vibrant venture capital market has been one of a number of significant constraints to its growth. As Murthy and Raju (2002, pp. 200–01), two pioneers of industry, put it:

Though Bangalore has attracted several high-quality, technology-focused venture capitalists, it has a long way to go before matching the hands-on approach, commitment, relationships, and risk appetites of some of the leading VC [venture capital] funds in the [Silicon] Valley . . . Efficient commercialization of cutting-edge output from research labs, entrepreneurship forums at universities, highly efficient alumni networks, close links between leaders in academia and business, risk appetites of venture capitalists, synergies between science/engineering schools and business schools, collaborative research among universities, keiretsus bringing together business and venture capitalists, angels with willingness to nurture talent, the abundance of forums where youngsters may put forth their ideas and interact with industry leaders, opportunities for collective learning—all these are differentiators that put the Valley several notches above other high-tech habitats. In sum, Silicon Valley operates in a vibrant market economy that reveres innovation.

Indian software firms look for venture capital funding from sources outside India.[5] Indus Entrepreneurs, an organization of Silicon Valley information technology (IT) professionals from the Indian Subcontinent, has contributed significantly to this outcome. Indus Entrepreneurs has a network of chapters in the Subcontinent and elsewhere and plays a mentoring role for aspiring entrepreneurs. Its publication on its mentoring experiences is a practical guide to entrepreneurship (Indus Entrepreneurs 2003).

Empirical Findings on Venture Capital

Gans, Hsu, and Stern (2002) note that venture capital investments increased rapidly between 1991 and 1999, and firms backed by venture capital accounted for more than 8 percent of all domestic innovation in the United States. They examine whether the returns to innovation are earned through product market competition (the Schumpeterian gale of creative destruction) or through cooperation with established firms through several mechanisms such as alliances and acquisition. These mechanisms differ in their impact on future incentives to innovate but share the common feature of foreclosing product market competition. The strength of the intellectual property regime (IPR) influences the absolute and relative returns to competition and cooperation. There is an expropriation threat in either choice. Under competition, incumbents could attempt to reverse engineer and sell an imitation of the innovation. Under cooperation, negotiating the sale of an innovation runs the risk of disclosure, thereby eroding the innovator's bargaining position and reducing the incumbent's willingness to pay for the innovation. More generally, the costs of search and bargaining, as well as uncertainty about the value of the innovation, could affect the choice between competition and cooperation.

The authors' empirical analysis is based on a comparison of 55 firms backed by venture capital and 63 firms backed by funding from the U.S. government's Small Business Innovation Research (SBIR) Program. The principal dependent variable is a combination of two distinct measures of cooperation strategy through two dummy variables, one indicating whether the firm earned licensing fees from innovation and the other

indicating whether the firm was acquired since the project was funded. Explanatory variables included measures of IPR strength, measures of investment costs in acquiring assets necessary for effective competition, and firm-level and project-level control variables. The empirical results provide support for a model in which start-up innovators earn their returns to innovation through cooperation when IPR protection is strong and the costs of acquiring and controlling complementary assets for effective competition are high. If the IPR regime is weak, start-up innovators are more likely to pursue competition. Thus the results suggest a subtle role for the IPR. While earlier literature emphasized the role of the IPR in raising the absolute returns to competition and cooperation, Gans, Hsu, and Stern (2002) find that a stronger IPR raises the relative returns to cooperation.

Kortum and Lerner (2000) examine the influence of venture capital on patented inventions in the United States across 20 industries during roughly three decades (1965–92). They use annual data on U.S. patents issued to U.S. inventors by industry and date of application as the dependent variable and measures of venture capital funding and industrial R&D expenditures as the main explanatory variables. They find that increases in venture capital activity in an industry are associated with significantly higher patenting rates. Their univariate comparisons of 122 firms backed by venture capital and 408 firms not backed by venture capital suggest that firms backed by venture capital are more likely to patent, have previous patents, and engage in frequent and protracted litigation in relation to both patents and trade secrets. Their results are robust to different measures of venture capital activity; subsamples of industries; and representation of the relationship between patenting, R&D, and venture capital. Averaging across their preferred regressions, they find that a dollar of venture capital resulted in three times as much patenting as a dollar of R&D expenditure.

Gompers and Lerner (2001) point out that to understand the venture capital industry, one must understand the venture capital cycle. The cycle starts with the raising of venture capital funds, which is followed by investing in, monitoring of, and adding value to firms. The final step is the exit of venture capital from successful deals and the return of capital to investors in the fund. The authors summarize the findings of the empirical literature on each segment of the cycle, namely, fundraising, investing, and exiting.

Hellman and Puri (2002) present data on 170 young high-technology firms in Silicon Valley using a combination of surveys, interviews, commercial databases, and publicly available information and examine the impact of venture capital on the development of new firms. They empirically substantiate Lerner's (1998) description of several ways in which venture capitalists interact with the firms they finance. They find that

> [o]btaining venture capital is related to a variety of organizational milestones, such as the formulation of human resource policies, the adoption of stock option plans, or the hiring of a VP [vice president] of sales and marketing. Firms with venture capital are also more likely and faster to replace the founder with an outsider in the position of CEO [chief executive officer]. Interestingly, however, founders often remain with the

company, even after the CEO transition. The effect of venture capital is also particularly pronounced in the early stages of a company's development.

Lerner (1998) notes that the U.S. federal government played an active role in financing new firms, particularly in high-technology industries, since the former Soviet Union launched Sputnik in 1957, a role that other governments in Asia and Europe, as well as state governments in the United States, have tried to emulate. For instance, the Small Business Investment Company program provided US$3 billion to young firms between 1958 and 1962. In 1995, the SBIR program provided almost US$900 million to young technology-intensive firms. Many U.S. states and some foreign countries have adopted similar policies in recent years. According to Lerner, these policy interventions are based on two basic assumptions, namely, that the private sector provides insufficient capital to new firms and that the government can pick winners, that is, it can identify firms where investments will ultimately yield high social and/or private returns. Few attempts have been made to test either assumption. Lerner's empirical analysis of the long-run success of firms participating in the SBIR program is based on the employment and sales growth of 1,135 firms, approximately half of which received one of more awards of approximately US$500,000 in the first three cycles of the SBIR. The other half consisted of matching firms constructed to resemble the SBIR awardees. A decade later, he finds that

> the SBIR awardees have enjoyed substantially greater employment and sales growth than the matching firms. This superior performance, however, was not universal. The differentials in both employment and sales growth were confined to firms in zip codes that were simultaneously the site of substantial venture capital activity. The SBIR awards appear to have had much less impact on the performance of firms in other regions. The awards contributed both to the growth of firms that were or were not backed by venture capital, and that were or were not in industries heavily financed by venture capital. Some evidence suggests that the positive impact was strongest for firms in areas with many venture investments but in industries not frequently financed by venture capitalists (Lerner 1996, p. 6).

While these results are consistent with the hypothesis of capital constraints, they may also be consistent with two other hypotheses: that the selection process leads to successfully picked winners, that is, firms with superior long-run prospects, and alternatively, that the award of SBIR funding served as a favorable signal to other investors and potential customers of the firm. Lerner tests these alternative hypotheses and finds that the data do not support them. Another hypothesis that did not find support was that SBIR awardees grew because they established relationships with federal officials or politicians, which led to their receiving government contracts. Lerner's finding that the SBIR program seems to complement venture capital organizations and other private institutions that fund new firms, in the sense that the impact of SBIR awards in regions without private sector funding is important, leads to questions about the wisdom of programs that provide public financing of and guarantees for venture capital funds that invest in economically disadvantaged areas.

In an illuminating empirical study using household data from Thailand, Paulson and Townsend (2001) find substantial evidence that financial constraints play an

important role in determining which households start or expand businesses in rural and semi-urban areas. They identify running a business with entrepreneurial activity. They find that wealthier households and more talented potential entrepreneurs are more likely to start businesses. The evidence that households that eventually start businesses accumulate wealth more quickly than households that do not also suggests the severity of financial constraints. Finally, they conclude that financial constraints impose a greater constraint on entrepreneurial activity in the less developed northeast than in the wealthier central region of Thailand.

Financial Revolution and the Exploitation Growth Opportunities

Rajan and Zingales (2003) provide a superb analysis of the role of well-developed and deep financial markets in reducing risk and shifting risk to those who are able and willing to bear it, thereby enabling the economy to exploit inherently risky, but highly productive, growth opportunities. They illustrate their analysis with a number of examples from history and contemporary events to show the immense potential of the markets, both to reduce and allocate risks efficiently and to improve the efficiency of poorly managed firms through markets for capital control. They discuss the possibility of spectacular failures and collapses in financial markets that can be ruinous to individuals, and even more disastrous for economies, as recent financial crises in Russia, East Asia, and Latin American have shown. However, in their view, notwithstanding the spectacular failures, the overall performance of financial markets has been outstanding.

In a short but wide-ranging paper, Rajan and Zingales (2001) examine the implications of the so-called financial revolution on the organization of firms in the United States and other industrial countries. In their view, the main elements of the financial revolution are the ability of markets to (a) price a variety of financial instruments, such as derivatives; (b) assess, reduce, and spread risks; and (c) collect more and timely information on market participants by insisting on improvements in accounting procedures, disclosure requirements, and, above all, greater transparency in decision-making as a condition for participation. The revolution was in part a response to financial deregulation that reduced or did away with barriers that restricted the domains of different financial intermediaries. The resulting increase in competition among financial intermediaries and the emergence of new types of financial firms improved the functioning of markets with respect to pricing and risk allocation.

Turning to the organization of firms, Rajan and Zingales (2001) note that because contracts are incomplete, firm owners derive power from their ownership of unique, alienable assets that are critical to production. Such power could arise from owners' attributes, such as their particular talents, which are critical to their firms' success, or more generally from other critical resources, such as a set of clients and associates who rely on working with the owners and who would have been less productive without them. This means that, unlike the ownership of unique, alienable assets that can be sold, control over other critical resources has to be built up through other means, such as internal organization, work rules, and, above all, incentives.

The allocation of power within the firm or organization acquires its importance from its impact on incentives and the determination of the range of feasible actions for each member of the organization. Moreover, the allocation of power at a point in time determines the constellation of power in the future, and thus the future efficiency of the organization.

For the purposes of this paper, Rajan and Zingales's application of the implications of the financial revolution on the framework of power allocation within firms to the exploitation of new project ideas, which they characterize as growth opportunities, is particularly relevant. Before the financial revolution, the balance of power within a firm affected project choice and implementation. Complementarities between a firm's financial capacity (consisting of existing assets that not only generated cash flow but could also serve as collateral) and its human capital, meant that the interests of the firm's owners and insiders (particularly its managers) were aligned. Owners benefited from the property rights to growth opportunities exploited by the insiders through financing new investments from the firm's financial capital. Primarily because insiders lacked financing before the financial revolution, they were happy to remain with the firm and exploit the investment opportunities with its financial capital, thereby enhancing their own earnings potential and career prospects.

The financial revolution changed this situation in several ways. Greater transparency and voice for shareholders means that insiders had to convince outside shareholders that proposed investments were profitable. A variety of mechanisms (see Rajan and Zingales 2003 for details) emerged and were put in place that punished insiders financially if their case that proposed investments were profitable was found wanting. At the same time, if a project could convincingly be shown to be profitable to a firm's own bureaucracy and its outside shareholders, outside financiers could also be convinced of its profitability. Thus, with outside financing more easily available after the revolution, the need for insiders to rely on a firm's own financial capital for investment was diminished. As a result, from a technological perspective, much stronger complementarities than before were needed between a firm's financial capital and growth opportunities if new opportunities for profitable investment were to be undertaken within the firm. Clearly this meant that the financial revolution widened the scope for profitable investment opportunities being implemented within an existing firm or with outside financing.

Needless to say, the financial markets in most developing countries are thin, underdeveloped, and prone to manipulation. A revolution of the type Rajan and Zingales describe is likely only in the distant future. As they explain, robust financial systems started to emerge in the 19th century in contemporary industrial countries. The reason for this—which is relevant for developing countries—is that financial development does not take place unless those who control or influence the levers of power want it to happen and therefore create the institutional framework for it. The framework includes not merely respect for the property of each citizen in the sense of allowing and enforcing each citizen's right to own private property, but also respecting the rule of law, in particular, facilitating and enforcing private contracts and

preventing arbitrary coercion and taxation. Without such a framework, saving and investing savings in productive assets will not take place. Rajan and Zingales's recapitulation of the relevant periods of English history shows that the greatest obstacle to the development of free markets in general, and of financial markets in particular, was the rapacity of governments. How the rapacity was tamed by means of constitutional limits to the arbitrary exercise of power by the crown, how parliament gained ascendance, and how the accountability of governments to the governed evolved is a fascinating story. The lessons for the developing countries are obvious.

The Role of Government and Nonprofit Organizations in the Innovation Process

In addition to private actors, such as innovators, entrepreneurs, and firms, public actors not motivated by profit play an important role by spending on and doing research. These include universities and research institutes, many of which, though autonomous, are funded largely by governments and government agencies. Governments play an equally important role through public policies (fiscal policies, trade policies, regulatory policies, and so on) that influence the incentives of private actors. This section starts out by discussing publicly funded R&D systems, then presents some empirical evidence about the impact of openness to foreign trade on productivity growth, and finally turns to the lessons about public policies derived from the success of the India's IT industry.

Public Funding of R&D

Data from World Bank (2002, table 5.11) show that during 1989–2000, expenditures for R&D as a proportion of gross national income averaged as much as 3.7 percent in Israel and 2.0 to 3.0 percent in the other industrial countries. Among developing countries, the average proportions were as high as 2.7 percent in Korea and 1.9 percent in Egypt, and ranged from 0.5 percent to 1.0 percent in many others, such as Argentina, Brazil, Chile, and India. These expenditures are by no means negligible.

India spent 0.9 percent of gross domestic product on R&D in 2000, up from 0.4 percent in 1970. The share of the private sector in the total doubled from 15 percent in 1970 to 30 percent in 1990 (Forbes 2003). Publicly funded and run research institutes, such as the Council for Scientific and Industrial Research, the Indian Agricultural Research Institute, and several national laboratories have existed for decades. In addition, defense expenditures in India—and in some of the other larger developing countries—include significant research components. The contribution of the Indian Agricultural Research Institute in breeding and diffusing high-yielding varieties of cereals suited to India's agro-climatic conditions and the resulting green revolution are well documented. However, the contribution of the Council for Scientific and Industrial Research and the national laboratories to industrial development is, at best, modest.

The Council for Scientific and Industrial Research, set up by the colonial government in 1943, now employs more than 10,000 scientists in 40 research laboratories. Forbes (2003) states that it advertises itself as the world's largest publicly funded agency for industrial R&D. He notes that several parliamentary committees reviewed the functioning of the council and concluded that its contribution to Indian industry was negligible. In response to this adverse finding, as early as 1953 an independent organization, the National Research and Development Organization, was set up for transferring technologies from the council's laboratories to industry. Its performance was apparently no better, with successive studies continuing to show negligible contributions to industry. Forbes (2003) includes excerpts from several of these studies to substantiate this.

Although hard-headed cost-effectiveness analysis of R&D systems across countries using a common methodology and complete and reliable data does not exist, several studies have looked at national R&D systems in several countries. Nelson (1993) reports on a comparative analysis of seven large and three smaller high-income countries (Australia, Canada, Denmark, France, Germany, Italy, Japan, Sweden, the United Kingdom, and the United States) and five lower-income countries (Argentina, Brazil, Israel, Korea, and Taiwan [China]). An important objective of the project, (other than describing and comparing national innovation systems) was derived from concern that earlier studies, based on little evidence or analysis, had concluded that particular features were behind country performance differences. This conclusion was neither grounded in a strong conceptual understanding of what is or is not likely to be a causal factor, nor was it based on the requirement that asserted causal connections be consistent with a wide range of country observations. Nelson hoped to remedy these effects.

Nelson (1993) summarizes a number of findings. First, national policies targeted at helping high-tech industries through support of industrial R&D are extremely uneven. Second, affluent countries with large domestic markets are able to sustain a diverse range of productive activities, and this diversity enables them to support a significant R&D system. Clearly poor countries, especially small, open, and poor economies, cannot build a cost-effective domestic R&D system. Third, resource endowments and the resulting pattern of comparative advantage shape R&D systems at a basic level. For example, countries with abundant natural resources, including arable and fertile land, are unlikely to have a comparative advantage in manufacturing. As such, they are unlikely to develop an R&D system to support manufacturing. By contrast, countries that have to export manufactured goods to pay for their imports of natural resources and farm products are much more likely to do so. However, the comparative advantage based on resource endowments is only part of the picture: conscious decisions to develop and sustain economic growth in certain areas, that is, "creating" comparative advantage, also matter. Fourth, self-reliance in products and technologies related to national defense may dictate significant spending on related R&D, which often spills over into the civilian sector over time.

The most interesting findings relate to the features common to effective innovation systems across countries. These are mostly other features of the economy than

of its R&D systems as such. In countries that were successful at innovation in specific industries, firms in such industries were competent across the board in designing and producing products, managing their companies, responding to market demands, and forging links into upstream and downstream markets. Such countries have education and training systems that turn out individuals with the knowledge and skills their competent firms require. University research is also a major contributor to successful innovation: those countries with strong research communities in their universities engaged in research in the sciences associated with particular industries (for example, chemicals, pharmaceuticals, or agricultural science), tend to be innovative in those industries. Public research institutes complement rather than substitute for university research. Last, but perhaps most important, are public policies (fiscal, monetary, and trade) that promote competition, particularly with foreign firms in domestic and export markets.

Foreign Trade and Domestic Productivity Growth

In the Helpman model cited earlier, free trade by the innovating North with an imitating South accelerated the rate of innovation. The model also regarded foreign trade policy intervention as having the potential to accelerate innovation and growth over the free trade steady-state equilibrium as long as such intervention did not trigger distortionary rent-seeking. In the Lucas model, foreign trade results in the emergence of a gap between what is produced and what is consumed at home. This gap results in a continuous shift of labor into the production of goods not previously produced. In turn, the shift means that learning benefits are sustained as the economy moves its labor from producing goods in which learning peters out to new goods in which it is significant. Thus sustained growth emerges as an outcome. The Indian software industry's success is in large part due to its export orientation.

The positive effect on innovation of openness to foreign trade is due to direct as well as indirect mechanisms. One of the latter is through foreign direct investment (FDI) and the associated technology transfer by multinational corporations. Sidney Winter, in his comments in this volume, describes this mechanism and its significance. Gur Ofer in his comments draws attention to other channels. He points out that almost all critical inputs needed for growth, such as capital, managerial skills, knowledge, innovations, and even institutions, can not only be imported, but can be imported in greater amounts, faster, and more cheaply than just a couple of decades ago. Among the importables (through FDI), he includes key financial service institutions such as banks, insurance companies, and accounting and consulting firms. This leads Ofer to suggest that developing countries should concentrate on the nontradable segment of institutional development. These nontradables are the same ones that Rajan and Zingales (2003) stress: institutions of governance, law enforcement, and protection of property and contracts and a level playing field for all those doing business. A large literature, theoretical and empirical, is available on the roles of foreign trade, FDI, and multinational corporations in accelerating growth in developing countries. Here I focus on two empirical studies that substantiate the importance of

openness in promoting domestic productivity growth, which is an indirect measure of the effects of innovation.

Eaton and Kortum (1997) examine productivity growth since World War II in five leading research economies: France, the former Federal Republic of Germany, Japan, the United Kingdom, and the United States. They estimate a multicountry model of technological innovation and diffusion to address several controversies:

> One is whether countries that start out poor grow faster than initially rich countries, so that income levels are "converging." A second is whether sources of growth are primarily domestic or foreign in origin. A third, and perhaps most fundamental, is what causes growth rates in output per worker to differ among countries: differences in capital per worker or differences in available technology (Eaton and Kortum 1997, p. 1).

They conclude that

> as for the issue of foreign vs. domestic sources . . . growth is primarily the result of research performed abroad. We find that even the United States obtains over 40 per cent of its growth from foreign innovations. These findings seem to be consistent with historical accounts (Eaton and Kortum 1997, p. 30).

Eaton and Kortum's study was confined to just five leading R&D countries. Coe and Helpman (1995) and Coe, Helpman, and Hoffmaister (1997) test the impact of openness on the transmission of technical knowledge across countries, and hence on TFP growth. Both estimate variants of the following basic regression:

$$LogF_i = \alpha_{oi} + \alpha_1 LogS_i^d + \alpha_2 LogS_i^f + u_i, \qquad (28)$$

where F_i is the level of TFP in country i, S_i^d is the domestic knowledge stock of country i, S_i^f is the foreign knowledge stock relevant for country i defined as the sum of the import-share weighted domestic knowledge stock of countries from which i imports, and u_i is random disturbance.

Coe and Helpman's (1995) sample includes 21 Organisation for Economic Co-operation and Development countries and Israel for 1971–90. Their results suggest a statistically significant and similar quantitative impact of domestic and foreign knowledge stocks on TFP growth.

Coe, Helpman, and Hoffmaister (1997) analyze data from 77 developing countries for 1971–90. Because few developing countries undertake R&D, the variable S^d in equation (28) is not relevant. In addition to $LogS^f$, they include the secondary school enrollment rate E; the share of imports from industrial countries M; dummies for time periods 1971–75, 1975–80, 1980–85, and 1985–90, and the interaction (that is, the product) of S^f with each M and E. For various reasons of a primarily econometric nature, they settle on the specification

$$\Delta LogF = -\underset{(3.043)}{9.853}\Delta M + \underset{(0.252)}{0.837}\Delta(MLogS^f) + \underset{(0.096)}{0.247}\Delta LogE, \text{ Adjusted } R^2 = 0.208.$$

Using this equation, they estimate R&D spillovers from the industrial countries to the developing countries. These estimates suggest that such spillovers from North to

South are substantial, and in 1990 "may have boosted output in the developing countries by about 22 billion US dollars" (Coe, Helpman, and Hoffmaister 1997, p. 148). To put this figure in perspective, total official development aid from multilateral and bilateral sources in 1990 amounted to about US$50 billion.

Like all cross-country regressions, the two studies are subject to criticism on data, econometric, and analytical grounds (see Srinivasan and Bhagwati 2001 for a critique). What is more, Keller (1998) shows that replacing actual import-share weighted foreign knowledge stocks by random-share weighted ones reduces the impact of domestic stock and increases that of foreign stock in explaining TFP. Because the rationale for using actual import weights is that by trading more with a country having a large knowledge stock, a country can augment its productivity by importing a variety of intermediate and capital goods embodying that knowledge, Keller's results—based on random, rather than actual, import weights—could be interpreted either (1) as raising doubts about trade as a mechanism through which knowledge spillovers occur, or (2) that the growth of knowledge stocks of different industrial countries was highly correlated, so that alternative weighing schemes yield highly correlated values for S^f.

India's Software Industry and the Sources of its Success

India's software industry is widely viewed as high-tech and knowledge intensive. The output of IT and software services grew by an annual average rate of 50 percent during the second half of the 1990s, and in 2000–01 accounted for 2.9 percent of gross domestic product. Software exports increased from US$0.3 billion in 1993–94 to US$7.2 billion in 2001–02 (Reserve Bank of India 2003, pp. II-12, VII-18). Indeed, the success of the industry, particularly recognition abroad (especially in Silicon Valley) of the professionalism and competence of Indian software engineers, has brought with it a certain visibility and level of media attention that India rarely received in the past. Within India, the success of IT has transformed the mindset of the young from one of resignation and despair to one of confidence in their ability to compete with the world's best.

According to Murthy and Raju (2002), the revenues of the IT industry grew in excess of 50 percent a year after 1991, reaching US$7 billion in 2000. They cite a report by the consultants McKinsey, which in 1999 projected that the revenues would grow to US$87 billion by 2008, of which US$50 billion would be from exports. Not only has actual growth exceeded these expectations, but even during the slowdown in growth in the United States and Europe after 2000, Indian IT exports continued to grow. For the purposes of this paper, what is of interest is the extent to which India's education and R&D systems, public policies, and ease of access to finance contributed to the success of the IT industry and might possibly constrain its future growth. From this perspective, the paper by Saxenian (2002) and comments on it by Murthy and Raju (2002) and Desai (2002) are illuminating.

Saxenian points out that prior to 1984, India's then dominant import-substitution-led development strategy "stifled entrepreneurs and isolated India from the

global economy. As a result, efforts to promote software exports during the period never took off" (Saxenian 2002, p. 171). Export performance requirements, such as a guarantee to export a certain amount of software in return for a license to import state-of-the-art computers after paying high customs duties and obtaining foreign exchange allocations, prevented the takeoff. Only after the Rajiv Gandhi government rejected a strategy led by import substitution and the idea of self-reliance in software, allowed imports of any form of software under liberal rules, attempted to attract foreign investment, and tried to provide venture capital by the industry have a chance to develop. Yet Saxenian quotes Sen to the effect that "until 1991–92, there was virtually no policy support at all for the software sector. Even the term 'benign neglect' would be too positive a phrase to use in this connection" (Sen 1994, p. 55). The systemic reforms of 1991 following a balance-of-payments crisis led to the abandonment of development led by import substitution and opened the economy to foreign trade and investment. The creation of software technology parks in the early 1990s, which are, in effect, export processing zones for software, provided infrastructure and administrative support, two concessions that guaranteed access to high-speed satellite links and reliable electricity.

The contribution of the Indian education system, particularly the elite Indian Institute of Technology and other engineering colleges, to the growth of the IT industry in India and to the supply of IT professionals to Silicon Valley, cannot be underestimated. Of course, the emigration of highly trained professionals (trained with substantial subsidies from the government) has been viewed as a brain drain and pessimists like the United Nations Development Programme have estimated the net loss (costs of training minus emigrants' remittances to India) at US$2 billion or more. However, the real issue in this connection is not the brain drain, but whether subsidizing higher education rather than primary and secondary education would be socially worthwhile. Be that as it may, as Saxenian documents, the Indian education system generated trained programmers and systems analysts whose wages in 1994 were less than 10 percent of the wages of similar personnel in the United States. Indeed, Murthy and Raju (2002) identify having a pool of well-educated, high-quality, English-speaking professionals as the greatest advantage of India's IT industry. Among its several disadvantages is, as noted earlier, the absence of a vibrant venture capital market.

Desai (2002, p. 204) argues that "the domestic software industry—as distinct from body-shopping [that is, sending technicians abroad to service foreign enterprises in situ]—could not have become internationally competitive if India's computers had continued to cost 50 to 100 percent more than its competitor's." The abolition of quantitative restrictions on computers and peripherals in 1992 and the suspension of restrictions on FDI in the software industry not only contributed significantly to the success of India's IT, but are also testaments, according to Desai, to the damage that a lack of openness exemplified by high tariffs and the restrictions on FDI have done in the past and continue to do to other sectors of the Indian economy, though to a lesser extent than before.

Conclusions

The theory and empirical evidence reviewed here point to the vital roles that openness to foreign trade, investment, and technology plays in spurring entrepreneurship, innovation, and growth. Access to finance for entrepreneurs, particularly in the form of venture capital, is important. However, it is not just the quantity of capital but the quality aspects unique to venture capital that are also important in encouraging entrepreneurship, such as initial scrutiny of applications for finance and later continued monitoring. The financial revolution that Rajan and Zingales (2001) describe has yet to happen in many developing countries, and indeed, their financial sectors are rudimentary. As such, the traditional balance of power will probably continue to prevail. In addition, the creation of a vibrant venture capital market is problematic; however, as Desai (2002) suggests, citing the examples of Japan and Europe, the importance of venture capital markets in financing start-ups might be exaggerated.

The developing countries clearly have room to open their economies to trade and FDI to a greater extent, that is, to participate in globalization more. But whether they will or whether they will be deterred from doing so by gloom and doom predictions about globalization being propagated by assertions unsupported by rigorous analysis by such luminaries such as Stiglitz (2002) is an open question. The empirical literature on venture capital financing emphasizes the importance of a strong IPR; however, this finding has to be put in perspective: the literature shows that given that entrepreneurs' ideas are protected by patents, the stronger the IPR, the lower the chance that their ideas will be expropriated. But this by itself does not mean that patents and their strong protection necessarily increase the rate of innovation, and that even if they did, that they are the most cost-effective means of encouraging innovation.

The study by Lerner (2002) on the nearly 150 years of patent protection in 60 countries, including the United States; the paper by Boldrin and Levine (2002); and the paper by Sakakibara and Branstetter (2001) on Japanese patent law reforms throw doubt on the first hypothesis. Cohen, Nelson, and Walsh (2000) surveyed 1,478 R&D laboratories in the U.S. manufacturing sector in 1994 and found that of the many mechanisms available to firms to protect the profits they made from inventions, they tended to emphasize patents the least. These findings confirm an earlier survey by Mansfield (1986), who found that an absence of patent protection would not have affected the innovative efforts of a majority of firms in most industries, with the exception of pharmaceuticals. They also corroborate a study by Levin and others (1987) who report that for firms in most industries, including the most R&D-intensive industries (with the exception of pharmaceuticals), patents were not important for reaping profits from their innovations.

The evidence is also not conclusive for the second hypothesis. Indeed, researchers have debated whether profit-driven research has an inherent tendency to focus on those areas in which market rewards with patent protection would be highest and neglect other areas, for instance, curative and preventive medicines for diseases widely prevalent in poor countries. Unfortunately, without any serious analysis of

their global social value, IPR regimes have been strengthened through the agreement in Trade Related Aspects of Intellectual Property Services and enshrined in the World Trade Organization.

Notes

1. According to *Webster's New World Dictionary* (1970, Second College Edition), innovation means something newly introduced, such as a new method, custom device, and so on, and changes in the way of doing things. This is the sense in which the term is used in this paper. Traditionally, innovation is distinguished from invention, which according to the same source means something thought up or mentally fabricated. In other words, the process of invention generates ideas, while innovation implements them or commercializes them. This paper has nothing to say about the process of invention.

2. Economists often argue that (1) even though innovation is essential for growth in mature economies, developing countries can in part free-ride on innovation undertaken by mature economies, and that (2) developing countries have much greater scope for realizing productivity growth through efficiency gains from the reallocation of resources. However, free-riding is impossible in reality: resources and ingenuity are needed even to adapt and introduce what is developed elsewhere. This resource cost is explicit as an imitation cost in the Helpman model. In any case, according to my definition, such adaptation and introduction are innovations. Empirical estimates from applied general equilibrium models of efficiency gains from resource reallocation, for instance, following trade liberalization, are modest.

3. My exposition closely follows Helpman (1990, pp. 29–31).

4. Given that all products developed at any time have the same cost of development; that each unit of product, once developed, costs the same to produce; and that there are no other costs, differentiating (for the consumer) one product from another is clearly costless.

5. Patni Computer Systems, an Indian information technology consultancy and software firm, has emerged as the biggest beneficiary of venture capital funding in the Asia Pacific region. According to the Thomson Corporation of Hong Kong (China), a total of US$1.9 billion in venture capital investments was disbursed to Asia Pacific companies in 2002, with 232 venture capital firms participating in 381 rounds of financing that supported 362 companies. However, the total amount disbursed sharply declined from more than $5.0 billion disbursed a year earlier (Thomson Corporation press release January 20, 2003, http://www.thomsonfinancial.com).

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Acemoglu, D., Phillipe Aghion, and Fabrizio Zilibotti. 2003. "Distance to Frontier, Selection, and Economic Growth." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Aghion, Phillipe, and Peter Howitt. 1998. *Endogenous Growth Theory*. Cambridge, Mass: MIT Press.

Amador, Manuel, and Augustin Landier. 2002. "Entrepreneurial Pressure, Innovation, and Rent Cannibalization." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 58(2): 377–409.

Black, Bernard S., and Ronald J. Gibson. 1998. "Venture Capital and the Structure of Capital Markets—Bank Versus Stock Markets." *Journal of Financial Economics* 47(3): 243–77.

Boldrin, Michele, and David Levine. 2002. "The Case Against Intellectual Property." *American Economic Review, Papers and Proceedings* 92(2): 209–16.

Boserup, E. 1965. *The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure*. Chicago: Aldine.

Coe, David, and Elhanan Helpman. 1995. "International R&D Spillovers." *European Economic Review* 38(3–4): 859–87.

Coe, David, Elhanan Helpman, and Alexander Hoffmaister. 1997. "North-South R&D Spillovers." *Economic Journal* 107(446): 134–49.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not)." Working Paper Series no. 7552. National Bureau of Economic Research, Cambridge, Mass.

Desai, Ashok. 2002. "Comment on Saxenian (2002)." In Anne Krueger, ed., *Economic Policy Reforms and the Indian Economy*. Chicago and London: University of Chicago Press.

Eaton, Jonathan, and Samuel Kortum. 1997. "Engines of Growth: Domestic and Foreign Sources of Innovation." *Japan and the World Economy* 9(5): 235–59.

Forbes, Naushad. 2003. "Higher Education, Scientific Research, and Industrial Competitiveness: Reflections on Priorities for India." Paper presented at the Fourth Annual Conference on Indian Economic Policy Reforms, June 6–7, Stanford University, Palo Alto, California.

Gans, Joshua, David Hsu, and Scott Stern. 2002. "When Does Start-Up Innovation Spur the Gale of Creative Destruction?" *RAND Journal of Economics* 33(4): 571–86.

Gompers, Paul A. and Josh Lerner. 2001. "The Venture Capital Revolution." *Journal of Economic Perspectives* 15(2): 145–68.

Grossman, Gene M., and Elhanan Helpman. 1992. *Innovation and Growth in the Global Economy*. Cambridge, Mass: MIT Press.

_____. 1994. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives* 8(1): 23–44.

Hellmann, Thomas, and Manju Puri. 2002. "Venture Capital and the Professionalism of Start-Up Firms: Empirical Evidence." *Journal of Finance* 57(1): 169–97.

Helpman, Elhanan. 1990. *Monopolistic Competition in Trade Theory*. Special Papers in International Finance no. 16. Princeton, NJ: Department of Economics.

Hu, Z., and M. Khan. 1997. *Why Is China Growing Fast?* Economic Issues no. 8. Washington, D.C.: International Monetary Fund.

Indus Entrepreneurs. 2003. *The First Mile: Essentials of Entrepreneurship*. Indianapolis, Ind. Wiley.

Kaldor, Nicholas, and James Mirrlees. 1962. "A New Model of Economic Growth." *Review of Economic Studies* 29(3): 174–92.

Keller, Wolfgang. 1998. "Are International R&D Spillovers Trade-Related? Analyzing Spillovers among Randomly Matched Trade Partners." *European Economic Review* 42: 1469–81.

Kortum, Samuel, and Josh Lerner. 2000. "Assessing the Contribution of Venture Capital to Innovation." *RAND Journal of Economics* 31(4): 674–92.

Lau, L. J., and J. I. Kim. 1994. "The Sources of Economic Growth of the East Asian Newly Industrialized Countries." *Journal of the Japanese and International Economies* 8(3): 235–71.

Lerner, Josh. 1996. "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program." Working Paper Series no. 5753. National Bureau of Economic Research, Cambridge, Mass.

_____. 1998. "Angel Financing and Public Policy: An Overview." *Journal of Banking and Finance* 22: 773–83.

_____. 2002. "150 Years of Patent Protection." *American Economic Review, Papers and Proceedings* 92(2): 221–25.

Levin, R., A. Klevorick, R. Nelson, and S. Winter. 1987. "Appropriating the Returns from Industrial Research and Development." *Brookings Papers on Economic Activity* 1987(3): 783–820.

Lucas, Robert E. 2002. *Lectures on Economic Growth*. Cambridge, Mass.: Harvard University Press.

Mansfield, Edwin. 1986. "Patents and Innovation: An Empirical Study." *Management Science* 32(2): 173–181.

Matsuyama, Kiminori. 1992. "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory* 58: 317–34.

Murthy, N. R. Narayana, and Sandeep Raju. 2002. "Comment on Saxenian (2002)." In Anne Krueger, ed., *Economic Policy Reforms and the Indian Economy*. Chicago and London: University of Chicago Press.

Nelson, Richard R., ed. 1993. *National Innovation Systems*. New York: Oxford University Press.

Nordhaus, W. D. 1969. *Invention, Growth, and Welfare*. Cambridge, Mass.: MIT Press.

Pack, H. 2001. "Technological Change and Growth in East Asia." In Joseph E. Stiglitz and Shahid Yusuf, eds., *Rethinking the East Asian Miracle*. New York: Oxford University Press.

Paulson, Anna L., and Robert Townsend. 2001. "Entrepreneurship and Financial Constraints in Thailand." University of Chicago, Department of Economics, Chicago. Processed.

Rajan, Raghuram G., and Luigi Zingales. 2001. "The Influence of the Financial Revolution on the Nature of Firms." *American Economic Review, Papers and Proceedings* 91(2): 206–11.

_____. 2003. *Saving Capitalism from the Capitalists*. New York: Crown Business.

Reserve Bank of India. 2003. *Report on Currency and Finance, 2001–02*. Mumbai.

Ruttan, Vernon W. 2001. *Technology, Growth, and Development: An Induced Innovation Perspective*. New York: Oxford University Press.

Sakakibara, M., and L. Branstetter. 2001. "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms." *RAND Journal of Economics* 32(1): 77–100.

Saxenian, Annalee. 2002. "Bangalore: The Silicon Valley of Asia." In Anne Krueger, ed., *Economic Policy Reforms and the Indian Economy*. Chicago and London: University of Chicago Press.

Sen, Pronab. 1994. "Exports from India: A Systemic Analysis." *Electronics Information and Planning* 22(2): 55–63.

Shell, K. 1973. "Inventive Activity, Industrial Organization, and Economic Activity." In J. Mirrlees and N. Stern, eds., *Models of Economic Growth*. London: Macmillan.

Solow, R. M. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39(3): 312–20.

Srinivasan, T. N. Forthcoming. "Economic Reforms and Global Integration." In Francine Frankel and Harry Harding, eds., *The India-China Relationship: What the United States Needs to Know*. Washington, D.C.: Woodrow Wilson Center Press.

Srinivasan, T. N., and Jagdish Bhagwati. 2001. "Outward Orientation and Development: Are Revisionists Right?" In Richard Snape and Deepak Lal, eds., *Trade, Development, and Political Economy: Essays in Honour of Anne O. Krueger*. New York: Palgrave.

Stiglitz, Joseph E. 2002. *Globalization and Its Discontents*. New York: Norton.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Incomplete Information." *American Economic Review* 71: 393–409.

World Bank. 2002. *World Development Indicators*. Washington, D.C.

Young, Alwyn. 1992. "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore." *NBER Macroeconomics Annual*. Cambridge, Mass.: MIT Press.

_____. 1995. "Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *Quarterly Journal of Economics* 110(3): 641–80.

Toward a Microeconomics of Growth

ROBIN BURGESS AND ANTHONY J. VENABLES

The emergence of new economic activities is the driving force of economic development. The development of such activities is often "lumpy," manifesting itself in rapid growth of particular regions or sectors. Recognition of these facts requires a reorientation of the analytical frameworks and empirical approaches that are used to investigate growth. In particular, we need to understand what drives growth at the microeconomic level.

The paper starts by reviewing some of the evidence on the spatial and sectoral concentration of rapidly growing activities. It then outlines a framework to illuminate why particular locations experience rapid growth while others remain backward. We divide the factors that determine a location's growth performance into two groups that we term "first advantage" and "second advantage." First advantage refers to the conditions needed to provide an environment in which new activities can be profitably developed, such as access to inputs, access to markets, basic infrastructure, and an appropriate institutional environment. Second advantage factors increase returns to scale and can lead to cumulative causation processes. They can be acquired through learning, technological spillovers, and thick markets of suppliers and local skills. Furthermore, these increasing returns, which are often external to the firm and therefore associated with market failure, underlie the lumpiness of development.

The analysis suggests that empirical investigation of the drivers of growth must shift to a more microeconomic level. Such an analysis has become more feasible as data at the subnational level have become more available. By reviewing recent empirical evidence on drivers of growth using our analytical framework we can begin to sketch out a microeconomic agenda for growth. We emphasize that the manner in

which first and second advantages interact is what shapes the pattern of development. Policy must therefore take into account the quite different natures of these advantages.

We then turn to the specific example of structural change in India. During the postcolonial period, India has been characterized by strong heterogeneity in manufacturing growth across states and sectors. We study three examples of how policy has affected manufacturing growth. The examples are powerful. They suggest that to engender growth we need to think much more carefully about policy choices at the local level. Both theory and empirics need to downshift to the microeconomic level if we are to identify specific means of encouraging innovation and growth.

Two key policy lessons emerge from our analysis. First, in the absence of first advantage elements, rapid structural change and growth are unlikely to take place. Second, ensuring that first advantage elements are present is not a guarantee that rapid economic growth will take place in a given location. Consideration of second advantage factors is critical. The paper constitutes a modest step toward understanding which factors drive growth at the microeconomic level. Given the strong links between innovation, growth, and poverty reduction, few endeavors in economics are more important.

Introduction

Growth and development typically involve new economic activities. New productive sectors develop and old activities are performed in radically different ways. Evidence from both economic history and economic development suggests that this process of structural change is central to increasing growth and raising standards of living (Hirshman 1958; Rostow 1960; Sokoloff 1986). Indeed, many early development theories viewed development as a process of transformation from agriculture to manufacturing (Lewis 1954). Labor productivity is higher in the newly developing sectors, and the reallocation of labor also increases productivity in traditional sectors.

Not much is known, however, about the likely pattern of structural change in future development. What sorts of new activities will take place in the rapidly growing countries of the 21st century? Will new patterns of development be evident, analogous to those Chenery (1960) and others identified in the 1960s? And most important, what factors are conducive to fostering these new activities at the microeconomic level?

This paper explores some of the issues surrounding the development of new activities in low-income countries. Our central thesis is that this process is frequently "lumpy," manifesting itself in rapid growth of particular regions or sectors. Spatial inequalities tend to increase during periods of rapid economic development. At the sectoral level, many fast-growing cities, regions, and countries have export specializations in a narrow range of activities. Recognition of these facts requires a

reorientation of the analytical frameworks and empirical approaches that are used to investigate growth.

We divide the determinants of locations' growth performance into two groups that we term "first advantage" and "second advantage" (terminology adapted from economic geographers' notions of first-nature and second-nature geographies). First advantage refers to the conditions needed to provide an environment in which new activities can most profitably be developed. Second advantage refers to the fact that many aspects of growth are self-reinforcing. The lumpiness of the growth process suggests the presence of increasing returns, so that a location's advantage in an activity derives, in part, from the very fact that it has a presence in the activity. The interaction of first and second advantages is what yields the growth patterns we observe and makes it so difficult to pin down the causes of rapid economic growth. Increasing returns and cumulative causation have featured in many analyses of economic development, from the development economists of the 1950s through Murphy, Shleifer, and Vishny (1989). We think it important, however, to recognize that key aspects of returns to scale are often found at the micro level: within narrow sectors of production or small districts of a city. A microeconomics of growth is needed to capture these sectorally and spatially concentrated processes.

Cross-country comparisons, though useful in identifying some of the important first advantages, cannot provide policy insights into what will drive growth in a particular location. Instead a bottom-up approach is required, whereby microeconomic studies are used to build up evidence on what works in different countries, regions, and sectors. This kind of empirical work is facilitated by the increased availability of micro-level data: regional, city, firm, and household data are now available for a wide range of countries. The institutional context in which policies are implemented and accumulation decisions are made affects economic performance (see Djankov and others 2003; Rodrik 2003). This makes it natural to study single countries where the influence of these factors can be carefully examined. Furthermore, the actors involved in development work are highly diverse. Policy and institutional innovations are taking place all the time and are being implemented by central governments, local governments, nongovernmental organizations (NGOs), and local communities. Evaluating these innovations requires careful microeconomic work at the subnational level.

To investigate these issues the next section presents a selective overview of the facts concerning the growth of new productive activities in developing countries. The paper then outlines an analytical framework that can be used to examine the determinants of structural change and growth and develops the distinction between the first and second advantage groups of factors. The following section goes into more detail, outlining what we know about some of the key elements of first and second advantage based on the available evidence. The paper then turns to work on India to illustrate how our approach can be used to generate insights into the microeconomic determinants of structural change. The paper wraps up with a concluding section.

Aspects of Modern Economic Growth

Much of the recent work on the determinants of growth has focused on countries' aggregate growth performance. The most popular examples of this are cross-country growth regressions, which typically regress growth performance on a number of explanatory variables. While these regressions reveal something of the importance of first advantages, such as human capital, good policy and institutions, and open trade regimes, many of their results are not particularly robust (see Djankov and others 2003; Rodrik 2003). Furthermore, they fail to capture some key aspects of modern economic growth. An aggregate approach overlooks two of the most important aspects of the performance of many developing countries: the uneven performance of different regions within each country and the different performance across sectors.

Spatial Concentration

For many countries, development is associated with increasing spatial inequality. The finding of rising spatial inequality during development dates back at least to Williamson (1965), and many studies have confirmed it since (for example, the studies of urban concentration by Henderson 1999 and Shishido and Wheaton 1982). This increase in spatial inequality often arises from spatial concentration in the development of manufacturing. We see this most clearly in data for large countries; for example, later we show that states in southern India have become prominent in manufacturing. In Mexico, manufacturing has become highly concentrated in regions that border the United States, and spatial variation in per capita incomes has increased dramatically since the mid-1980s (Cikurel 2002). In China, Demurger and others (2002) chart increasing spatial inequalities in per capita gross domestic product (GDP) from the mid-1980s. The coastal provinces experienced the greatest decline in the share of agriculture in employment and output and the fastest growth of per capita income.

   The spatial concentration of new activities also occurs at a much finer level of disaggregation than state-level or province-level data suggest. Aranya (2002) documents the polarization of activity in a number of city regions in Asia. Bangalore is a good example: Estimates indicate that the city accounts for 25 percent of India's software exports, with some 100,000 workers in the city producing 3 percent of India's total exports. Within cities, particular sectors frequently cluster in particular districts (see, for example, the study of India by Chakravorty and Lall 2003).

   Spatial concentration is most dramatically demonstrated by the role of urbanization, and of megacities, in development. Worldwide the number of cities with a population of more than 1 million increased from 115 in 1960 to 416 in 2000, the number with a population of more than 4 million increased from 18 to 53, and the number with a population of more than 12 million rose from 1 to 11 (Henderson and Wang 2003). Henderson (1999) finds that national urban concentration (the dominance of the largest city) rises with growth from low-income levels, peaks at low-middle-income levels (1987 purchasing power parity income of around US$2,500 per

capita), and then declines. Thus despite the massive diseconomies associated with developing country megacities, even more powerful economies of scale make it worthwhile for firms to locate in these cities. Urbanization is one of the clearest features of the development of manufacturing and service activity in developing countries, yet discussion of urbanization is strangely absent from economic analyses of growth and development.

Sectoral Concentration

A substantial but somewhat old development economics literature derived from Chenery's "normal" pattern of sectoral change (Chenery 1960; Chenery and Taylor 1968) looks at the changes in countries' production structures typically associated with development. Using a trade framework, a number of authors (for example, Leamer 1984, 1987; Schott 2003) have researched the specialization of countries and find that so-called development paths link the structure of countries' production to their factor endowments. These studies are, however, based on sectorally aggregated data and seem to miss key aspects of modern specialization.

What do we know about specialization during development in the 21st century? Globalization has created the possibility of trade in new products and services and of a much finer pattern of specialization. The impact of new technologies is perhaps best illustrated by the experience of India (and of Bangalore in particular), which is able to export labor services embodied in information technology and services. The finer pattern of specialization arises with the growth of production networks. Component parts and semi-finished goods can cross borders multiple times, and countries are able to engage in vertical specialization (that is, producing just one narrowly defined part of a product). Data on such activities can be hard to obtain, as they do not correspond to the standard commodity classifications of trade. However, one of the fastest-growing elements of world trade has been trade in parts and components, which now accounts for around 30 percent of world trade in manufactures (Yeats 1998).

A striking feature of growth has been the fact that many countries have done well in a few extremely narrow product segments. Once again, India's success in software products is an example. Hausmann and Rodrik (2002) look in detail at exports to the United States from Bangladesh, the Dominican Republic, Honduras, the Republic of Korea, Pakistan, and Taiwan (China), using data at the six-digit level (for example, "hats and other headgear knitted or from textile material not in strips"). Even at this fine level of disaggregation, for each of these countries their top four product lines account for more than 30 percent of their exports to the United States. Furthermore, surprisingly little overlap is apparent in the top product lines of quite similar countries. For example, Bangladesh is successful in exporting shirts, trousers, and hats but not bed linen or soccer balls, while Pakistan does well in bed linen and soccer balls. Only 6 product lines are in the top 25 for both these countries. Hausmann and Rodrik (2002, p. 26) conclude that "for all economies except possibly the most sophisticated, industrial success entails concentration in a relatively narrow range of high productivity activities."

These findings are not inconsistent with the broad pattern of development paths identified by earlier researchers; however, they indicate that these paths are only part of the story. Aggregate studies can conclude that a country has a comparative advantage in a labor-intensive activity such as textiles, but they provide no basis for predicting what particular product line a country will produce. To understand the economic success stories we must also know why spatial and sectoral concentration occurs.

A corollary of spatial and sectoral concentration is that exports are essential, and narrow sectoral specialization can develop only if output is exported from the city, region, or country. Increasing shares of exports in income are observed for many countries. According to the World Bank (various years), between 1980–82 and 1998–2000, the export to GDP ratio rose from 22 to 41 percent for the East Asia and Pacific region and from 8 to 14 percent for South Asia, compared with an increase from 20 to 23 percent for the world as a whole and 28 to 29 percent for Sub-Saharan Africa.[1]

Analytical Framework: First and Second Advantage

To understand the microeconomic determinants of growth we divide the driving factors into two sets of factors that operate in distinct ways. We refer to one set as first advantage. These are the factors that can be viewed as preconditions for the development of new activities; they are necessary, but not sufficient, conditions for growth. First advantage factors include most of the considerations that traditional theory has focused on, such as access to inputs (labor skills and capital), access to markets, provision of basic infrastructure, and presence of an appropriate institutional environment. These factors shape the business environment in which firms make investment and location decisions. We refer to the other group of factors as second advantage. Second advantage factors inherently have some element of increasing returns to scale and may therefore give rise to processes of cumulative causation. They are a function of the past and present level of activity in a location or sector and include such factors as technological capability, knowledge spillovers, thick market effects, and networks. For example, a particular city may be a good place to set up a business precisely because of spillovers from existing similar businesses. The increasing returns are typically sector- and/or location-specific and are often external to the firm. This combination of sector- and location-specific increasing returns to scale and external effects is what gives rise to the lumpy performance described in the previous section.

Before looking at the elements of first and second advantage in detail, the rest of this section outlines some theory on how they interact to shape the growth process. To this end, we draw on work on subnational inequalities (Venables 2003). However, the rather broad lessons we draw have wider applicability to international and intersectoral development, as well as to subnational development (see, for example,

Fujita, Krugman, and Venables 1999 and Puga and Venables 1999 for more fully developed models).

Suppose that a particular activity or set of activities could take place in many potential locations. For our purposes a location can be a town or a city and the activity can be a particular industrial sector, although broader interpretations are also possible. The private return to creating a job in this activity in location i is π_i, where $\pi_i = q_i a(n_i) - (w(N) + t_i)$.

The first term on the right-hand side of this expression is the value of output from the job. The first part, q_i, measures the first advantage of location i, capturing all aspects of the environment that, while specific to the location, are exogenous for the firms and sector under study. We rank locations such that location 1 has the highest value of this term, location 2 the next highest, and so on. For an export activity an example of this might be the distance of the location from the port, with more remote locations having worse first advantage (a lower value of q_i).

The second part is a function $a(.)$ of the level of activity in location i as measured by employment, n_i. This function measures second advantage, and we assume that it is increasing and concave, capturing increasing returns to scale that become progressively exhausted. In contrast to first advantage, which is location specific, second advantage is represented by the same function at all locations, although the function is evaluated at location-specific values of the endogenous variable, n_i. Thus if one location has a higher employment level in the activity than another, then, because of increasing returns, it will also have a higher value of $a(n_i)$ and higher output per worker. Second advantage can originate from different mechanisms, including the presence of knowledge spillovers, the acquisition of technical know-how, the development of a dense network of local suppliers of specialist inputs to production, and the presence of thick local labor markets in particular sector-specific skills.

The remaining term, $w(N) + tn_i$, is the cost of employing labor. $w(N)$ is the opportunity cost of labor, that is, its marginal product (wage) in some alternative use. This wage is a function of total industrial employment, $N \equiv \Sigma_i n_i$, and rises as workers are drawn out of agriculture into the new activities, $w'(N) > 0$. tn_i is the cost of living in a city of size n_i. This could be a congestion cost or, in the tradition of the urban economics literature, the rent plus commuting cost that urban workers have to pay. It is a source of diminishing returns, preventing all activity from concentrating in a single location.

We are now in a position to address the following question. Suppose that total modern sector employment, $N = \Sigma_i n_i$, is exogenous and increasing through time. As jobs are created they go to locations that offer the highest return (highest value of π_i). What form does growth take in this economy? Figures 1 and 2 illustrate equilibrium patterns of development. The horizontal axis is total employment, N, so the economy moves to the right along this axis through time. The vertical axis is employment in each location, and the curves on the figure illustrate equilibrium employment in each location, n_i.[2]

FIGURE 1.
Development without Second Advantage
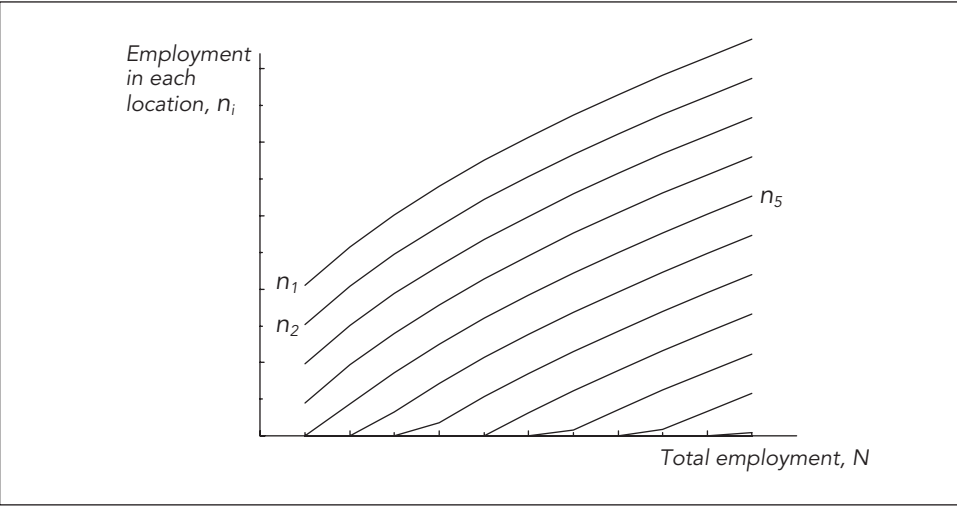
Source: Authors.

FIGURE 2.
Development with First and Second Advantage
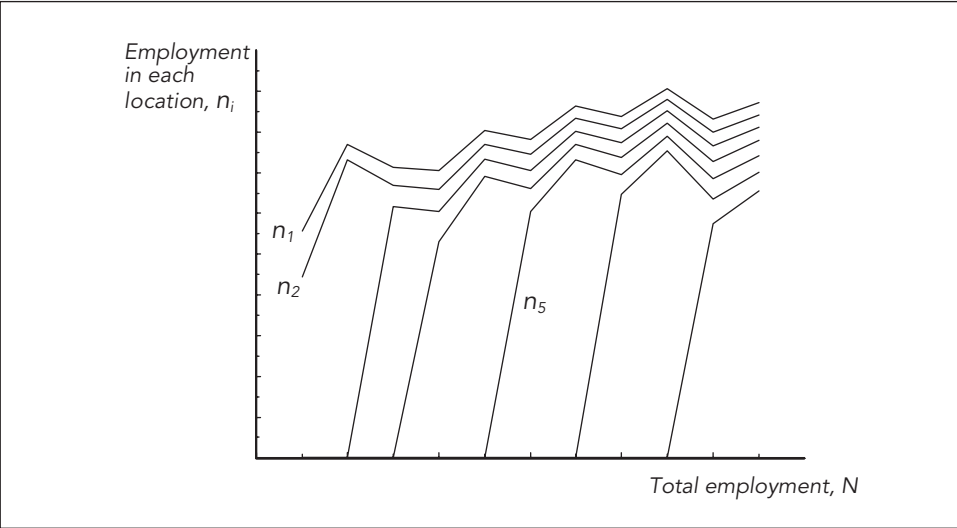
Source: Authors

Figure 1 illustrates the case where there is no second advantage, so $a(n_i)$ is a constant, the same for all locations. There is, however, first advantage heterogeneity, and this maps in a continuous way into outcomes. Better locations (those with low i and hence high q_i) get activity sooner and are larger than worse ones. Thus the curve n_1 gives employment in the location with the best first advantage, n_2 in the second

best, and so on. Increasing N causes both intensive and extensive growth, enlarging existing locations and making it profitable for new locations to gain activity. Figure 2 gives the equilibrium when second advantage returns to scale are present. Development is much less continuous. Once a particular location starts to attract activity, increasing returns cut in and lead to rapid growth. Outcomes are determined by a combination of the first advantage of each location, which determines the sequencing, and the second advantage of increasing returns, which determine the growth path of each location.[3]

We learn several things from a comparison of figures 1 and 2. First, in the presence of second advantage, development is uneven: locations do not develop in parallel, but in sequence. Second, development is more spatially concentrated. For any value of total employment, N, activity is concentrated in fewer and, on average, larger locations. Third, first advantage determines the order of development and, at any point in time, predicts which locations are successful and which are not. However, the dependence of performance on first advantage is not continuous. For locations just at the threshold of development, small differences in first advantage can translate into large differences in outcomes; if two locations have similar first advantage, then a small improvement could cause one to overtake the other and gain industry sooner. Conversely, for locations well below the current threshold, an improvement in first advantage does not confer any immediate payoff in terms of attracting the activity. Finally, among the set of active locations, differences in first advantage have little effect on employment, as they are dominated by acquired second advantages.

What about the efficiency of the equilibrium development pattern? In the absence of second advantage, the development path is efficient: the simple structure we have outlined does not include market failures. However, the presence of second advantage adds two distinct types of market failures. Within existing locations, private agents do not take into account the fact that an expansion of employment raises the productivity of workers already in the city, meaning that private incentives to expand employment in the location are too small. In addition, a coordination failure is involved in the development of new locations and activities. In figure 2, these new locations and activities are initiated when they become privately profitable for an individual small firm. However, the profits from the development of a new location or activity would be higher if firms could coordinate their decisions and act collectively.

The two market failures combine to have two effects. One is that cities tend to be too large, because of the difficulty of initiating activities in new locations. The other is that the overall return to job creation, π_i, is too small. If the rate of growth of jobs depends on π_i, then employment growth is less than optimal. In an extension of the present model, multiple equilibria are possible, and countries can become stuck in a low-level equilibrium. Essentially, the slower employment growth is, the worse the coordination failure is (as a new city only achieves increasing returns relatively slowly), but a worse coordination failure means lower returns to job creation, and hence slower employment growth. Thus second advantage at the local level causes cumulative causation at the aggregate national level.

Summing up, the combination of first and second advantage offers an explanation of the lumpiness of development. According to the theory, both advantages need to be considered together: changes in first advantage alone do not necessarily translate into changes in outcomes. The theory also indicates the presence of market failure, and hence potential scope for policy. Of course, the form any such policy should take depends on the detailed ingredients of first and second advantage.

A Microeconomic Agenda for Growth

This section confronts our analytical framework with some evidence on the drivers of growth. This allows us to begin to sketch out a microeconomic agenda for growth. We do not intend to be comprehensive, but instead focus on several areas, including institutions, skills, technology, and trade. We have several objectives. One is to illustrate the distinction between first and second advantages in each of these contexts and thereby make concrete the somewhat abstract discussion of the previous section. Another is to provide evidence for some of the increasing returns mechanisms and give a sense of their importance. This section also points to the sort of detailed microeconomic information that policymakers need to encourage growth within a narrow geographical jurisdiction.

Property Rights and Contracts

Observers often see the strengthening of property rights as critical for promoting innovation, investment, and growth (Aghion and Howitt 2003; North 1990). Recent cross-country papers provide evidence in this direction. Acemoglu, Johnson, and Robinson (2001) find that insecure property rights are associated with lower income per capita. Hall and Jones (1999) show that a social infrastructure variable (which captures both the quality of government institutions and openness to trade) is positively associated with income per capita. How to map from these findings into concrete policy suggestions, however, is not clear.

Here microeconomic evaluation is pointing the way. For a given location we want to understand which elements of the property rights regime are central to engendering innovation and growth. Strengthening property rights over land appears to be an important element of first advantage. For example, the 1978 rural reforms in China, which entailed a shift from collective to household farming, are credited with engendering increases in agricultural productivity (Lin 1992) and an explosion in town and village enterprises, which were the engine of growth in China until the mid-1990s (Qian 2003). In a similar vein, Banerjee, Gertler, and Ghatak (2002) show that a government program that increased tenurial security in West Bengal had a large, positive effect on agricultural productivity. Field (2002) finds that the issuance of property titles to urban households in Peru led to an increase in labor hours and a shift in labor supply from work at home to work in the outside market. Besley and Burgess (2000) find that land reform acts passed in

Indian states account for about 10 percent of the overall fall in poverty between 1958 and 1992.

Effective institutions for enforcing contracts are another source of first advantage. Djankov and others (2002) examine court procedures for evicting tenants or collecting on bounced checks in 109 countries. They find that common law countries exhibit greater procedural formalism, which is associated with slower judicial proceedings, more corruption, less consistency, less honesty, less fairness in judicial decisions, and inferior access to justice. Johnson, McMillan, and Woodruff (2002a) surveyed small manufacturing firms in Poland, Romania, Russia, and Ukraine and found that the perceived effectiveness of courts in resolving commercial disputes affected entrepreneurs' propensity to reinvest their profits. Well-functioning courts also encouraged entrepreneurs to form new trading relationships with their customers and suppliers (Johnson, McMillan, and Woodruff 2002b). Besley and Burgess (2003a) find that long backlogs in state high courts in India are associated with lower output in informal sectors, where contract enforcement problems may be acute.

Where contract laws are inadequate, informal relationships such as business networks may substitute for courts in allowing deals to be made. Typically proximity and frequency of interaction foster the development of such networks. This is a second advantage. A locality with a tightly knit business sector develops this advantage, helping to explain why new economic activities tend to be clustered. For example, McMillan and Woodruff (1999) use the amount of trade credit extended by small manufacturing firms to suppliers and customers in Vietnam as a measure of business trust. They find that relationships of longer duration, visits to premises, and being part of the same business network all enhanced the propensity to extend trade credit. They argue that relational contracting within business networks has enabled manufacturing to flourish in Vietnam even in the absence of formal institutions; however, as firms grow, formal institutions to support finance and contracting will be required (McMillan and Woodruff 2002).

Regulation

A second major element of first advantage concerns the degree to which government regulation impedes entrepreneurship and the emergence of new activities. Economic theories that saw the state as playing a central role in directing and coordinating economic development led to a huge variety of regulations being put in place. However, increasing empirical evidence indicates that—no matter how well intentioned the architects of regulation may have been—regulation has been neither an engine of economic development nor a boon for the poor.

One problem is that officials can use government regulations to extract rents from entrepreneurs. For example, recent surveys of start-up firms reveal that more than 90 percent of Russian managers had to make extralegal payments to secure government services or a business license (Johnson, McMillan, and Woodruff 2002a, b). Those firms that were most concerned about corruption invested 40 percent less than those that were least concerned. Government regulation of entry can

also impede entrepreneurship and innovation. Djankov and others (2002) collect data on the number of procedures an entrepreneur must complete to officially open a business and the time the entire process takes in 85 countries. They find pronounced variation across countries, with the number of regulations involved in setting up a business being greater in poorer countries. Heavy regulation of entry is associated with less democratic governments, more corruption, and larger unofficial economies. They argue that this is in line with a public choice view of regulation as being put in place by officials intent on extracting rents.

Findings for labor regulation point in a similar direction. Botero and others (2003) study labor regulations in 85 countries and find that heavier regulation of labor is associated with a larger unofficial economy, with lower labor force participation, and with higher unemployment. Besley and Burgess (2004) find that pro-worker labor regulations in Indian states are associated with lower output, employment, investment, and productivity in registered (formal) manufacturing, but higher output in informal manufacturing. A survey of 1,000 manufacturing establishments across 10 Indian states revealed that managers would be willing to reduce their workforce by 16 to 17 percent if labor market flexibility were greater, indicating the negative impact of labor regulation on firm productivity (Dollar, Iarossi, and Mengistae 2001).

Of course, government regulation is called for in a number of areas. For example, in the area of finance, legal and regulatory institutions that protect investors and creditors from expropriation are critical for engendering financial development. Glaeser, Johnson, and Shleifer (2001) argue that in Poland, strict enforcement of the securities law by a highly motivated regulator was associated with a rapidly developing stock market. In contrast, in the Czech Republic, hands-off regulation was associated with a moribund stock market. La Porta and others (1998) find that common law countries provide investors and creditors with greater protection than do civil law countries, and this results in larger debt and equity markets (La Porta and others 1997) and faster economic growth (Levine 1998). Given that external finance becomes increasingly important as firms grow, identifying the set of institutions that support financial development in a given location is critical.

Nevertheless, removing the impediments to entrepreneurship via regulatory change can frequently be a powerful mechanism for attracting new firms to a location. As employment in a location increases, various types of second advantage kick in as a result. Thicker markets imply better access to inputs and customers, skilled workers and external finance become more available, and firms benefit from technological upgrading. The emergence of entrepreneurs as a powerful political class can also lead to demand for institutions that protect property and support contracting and finance. Stern (2003) argues that improving the investment climate can be a powerful way to engender increasing returns to scale along these different dimensions.

Education and Skills

Both theory and empirical evidence suggest that investment in human capital is central to promoting innovation and growth (Aghion and Howitt 1998). The best

estimates of the rate of return to education in developing countries (from micro-econometric studies that take endogeneity and measurement errors issues seriously) line up with those for industrial countries: each additional year of schooling is associated with a 6 to 10 percent increase in earnings. Duflo (2001), for example, uses a large project to build primary schools in Indonesia to establish that the economic returns to extra schooling this program provides lie in this range. This evidence appears robust across both methods and locations (Card 1999). After accounting for measurement error, the effects of changes in educational attainment on income growth in cross-country data are at least as great as microeconometric estimates (Krueger and Lindhal 2001). These results suggest that access to education is an important source of first advantage in a location.

This still leaves open the central question of how such an expansion in education is best achieved. New work in the area is addressing this, looking at the market conditions under which it is provided and the incentives education providers face. One strand of research focuses on policy design. Banerjee and others (2003) use randomized experiments to look at whether remedial education and computer-assisted learning programs implemented in India by NGOs affected learning. Investigators have also used randomized intervention in western Kenya to evaluate whether improving the quality of education by increasing the supply of textbooks and improving child health affect attendance and attainment (Glewwe, Kremer, and Moulin 2000; Kremer and Miguel 2001). Another strand focuses on whether the organization of policy delivery needs to be changed. Political representation, decentralization, and involvement by NGOs and the private sector are major issues here (Chattopadhyay and Duflo 2001; Hsieh and Urquiola 2002).

Education levels raise the return to creating a job in a particular location and are therefore an element of first advantage. The human capital embodied in skills that workers acquire while doing a job can also spill over to other firms, for example, via job turnover: workers may become familiar with a particular technology and be able to use this knowledge in other firms. The magnitude of such spillovers will increase as the level of economic activity in a given location increases, and constitute a source of second advantage. Both micro-level and more aggregate studies of the importance of these effects are available. Evenson and Westphal (1995) review firm-level studies and find extensive evidence of learning by doing within firms and of spillovers between firms. They also report that the diffusion of technology between firms comes largely from the turnover of skilled workers and managers. Human capital externalities are also evident, whereby increasing the number of skilled workers in a location results in external benefits for other workers (Klien 2003; Moretti forthcoming).

Trade and Market Access

We have already argued that exporting is an important aspect of modern economic growth. Barriers to exports are, in our terminology, a source of first *dis*advantage, reducing the return to job creation. While this suggests a need for liberal trade policy,

several caveats are called for. Indeed, the role of trade policy in economic development has been the subject of intense and passionate debate over the last decade.

Our central thesis is that the mapping from first advantages to outcomes is not straightforward. The removal of trade barriers is not a sufficient condition for growth, as the empirical literature has confirmed. While samples of successful countries indicate strong export performance (Bhagwati and Srinivasan 1999; Dollar and Kraay 2000), cross-country regressions attempting to establish a causal link between open trade policy and growth have failed to come up with robust findings (Rodriguez and Rodrik 2000). This is partly because of the complex interaction between first and second advantages. Thus some authors argue that in some instances, especially in extremely poor economies, liberalization actual impedes growth by inhibiting infant industries and local accumulation of knowledge (Hausmann and Rodrik 2002; Krugman 1981; Young 1991). Others emphasize that, in the presence of capital market imperfections, liberalization exacerbates income inequality within countries, imposing high costs on less favored regions, social groups, or sectors of activity, and has ambiguous effects on average performance (Banerjee and Newman 2003; Trefler and Zhu 2001).

Once again, micro-based research is going to be more useful than broad-brush, cross-country work in informing the design of policy in particular countries. For example, Aghion and others (2003) show that technological capability and institutional conditions determine whether industries in India benefited from or were harmed by liberalization in 1991. Using Chilean plant-level data, Pavcnik (2002) finds that import liberalization led to productivity improvements within plants. Other recent work using firm-level data from a number of countries has also drawn attention to the productivity gains from exporting (Hallward-Driemeier, Iarossi, and Sokoloff 2002). McCulloch, Winters, and Cirera (2002) have undertaken detailed studies of the impact of trade reform on poverty.

In addition to trade policy, geography is also an important determinant of access to markets in which to sell output and from which to source intermediate and capital goods. Much of the success of Mexico's northern border region is a direct consequence of its proximity to the United States. Similarly, coastal regions of China do well simply because they are on the coast. A number of studies have tried to quantify the importance of these geographical first advantages. Gallup and Sachs (1999) regress national per capita income on four variables: a measure of the endowment of hydrocarbons; a dummy variable for the incidence of malaria; and two measures of geography, namely, internal geography as captured by the proportion of the population living within 100 kilometers of the coast, and external geography as measured by shipping costs (the ratio of cost, insurance, and freight to free on board). They find that these four variables account for 69 percent of the per capita variation in incomes across countries. Frankel and Romer (1999) and Redding and Venables (2004) also show how geography-based measures are important determinants of levels of per capita income.

In this area, as with others, microeconomic detail is important. The market access of a location depends on its infrastructure and the quality of its internal and external transport links. Limao and Venables (2001) construct an index of infrastructure

quality and show how a deterioration of infrastructure from the median to the 75th percentile raises transport costs by 12 percent and reduces trade volumes by 28 percent. The effects are even greater for landlocked countries. Numerous studies of port quality point to the damaging effects of delay, be it because of capacity constraints or maladministration. Second advantage effects are apparent in this area also. Both the costs of shipping and the frequency of service are subject to increasing returns to scale, reducing costs for ports with a high turnover.

Thick Market Effects

Thick market effects arise when increased volumes of trade in a particular market increase the efficiency with which the market operates. A number of mechanisms give rise to such effects, including an improvement in the matching of buyers and sellers, a reduction in the level of monopoly or monopsony power, and the development of specialist suppliers (a finer division of labor). These effects are all fundamentally second advantages: increasing the scale of activity in a market creates benefits for other consumers, workers, or firms.

Such thick market effects are important in many markets. As already noted, in relation to the supply of finance, trade credit can be extended more readily in tightly knit business networks. Thickening financial markets in a particular location enable external finance to become more available and less costly, allowing firms to expand in size and to upgrade the technologies they employ (Rajan and Zingales 1998). This in turn leads to greater demand for financial services. Complementary institutions, such as credit rating agencies, and more specialized finance providers, like venture capital firms, often develop as economic activity in a location increases.

In the labor market, the thick market effect is usually known as labor market pooling, and it operates in two ways. One is that workers and firms are better able to match their specific skills and needs the thicker the market. The other is that scale reduces the risk of an excess supply of or demand for specific skills (if, for example, there are firm-specific shocks). This in turn makes it more worthwhile for workers to invest in skill acquisition. Rosenthal and Strange (forthcoming) present some evidence in support of these hypotheses.

Turning to the product market, firms generally derive benefits from being located close to markets for their output and to suppliers of intermediate inputs. Such input-output linkages are a source of second advantage, as firms seek to locate close to other firms with which they are transacting. Forward and backward linkages between these firms constitute a pecuniary externality that raises productivity in the sector and/or location concerned and gives rise to activity clustering. Fujita, Krugman, and Venables (1999) model this theoretically, and on the empirical side. Holmes (1999) presents evidence of U.S. firms locating so as to realize the benefits of proximity to input suppliers.

A good deal of case study evidence points to the importance of dense networks of specialist suppliers. For example, Hobday (1995) studies a number of situations in which initial multinational investments in developing East Asia created backward

linkage effects to local suppliers. Examples include computer keyboards, personal computers, sewing machines, athletic shoes, and bicycles in Taiwan (China). The initial foreign investments created demand for local suppliers and improved their quality, productivity, and product diversity. Large numbers of local firms entered to supply components or assembly services to multinational firms. This growth in the supply and productivity of component and other intermediate goods in turn created a forward linkage effect to the final goods producers, drawing in more multinationals and domestically-owned firms. A second-round backward linkage effect followed, and so forth. In some cases, for instance, bicycles and computers, local firms eventually displaced the original multinational entrants.

Technology and Productivity

Success in a particular product line requires mastery of the technology. Sutton (2002) emphasizes this point, arguing that to compete successfully, a firm must be within a relatively small window as defined by production costs and product quality. Acquiring capability requires fixed outlays on research and development and related activities, and is therefore associated with increasing returns to scale. Learning by doing may amplify such effects. For Hausmann and Rodrik (2002), the development of successful activities requires experimentation, with firms having to incur upfront expenditures to learn about the comparative costs of their location.

Knowledge spillovers mean that productivity depends on the activities of others, and are a source of second advantage. These spillovers are often sector and location specific. Hausmann and Rodrik (2002) argue that the results of experiments become public knowledge, but the knowledge revealed is specific to the activity and location combination. If knowledge spillovers occur through job turnover, they are also likely to be spatially concentrated. Sutton (2002) argues that knowledge is embodied in teams of workers and that moving an entire team between locations is difficult.

Many studies establish a positive relationship between spatial concentration and activity, confirming the spatial dimension of spillovers. For example, from work on U.S. states, Ciccone and Hall (1996) find that doubling employment density raises labor productivity by 6 percent. A recent survey by Rosenthal and Strange (forthcoming) reports a consensus view that doubling city size increases productivity by around 3 to 8 percent. Authors who have tried to ascertain the spatial range of such effects find that information spillovers attenuate rapidly over space (see, for example, Jaffe, Tratjenberg, and Henderson 1993). Face-to-face communications remain important, despite technological changes (Gasper and Glaeser 1998).

As for the sectoral spread of spillovers, the economic geography literature is engaged in a long-standing debate between the relative importance of "Jacob's externalities" and "Marshall-Arrow-Romer externalities." The former derive from the diversity of a city, the argument being that serendipitous interaction in a large and diverse economic environment, such as New York City, stimulates innovation and creativity. By contrast, Marshall-Arrow-Romer externalities are knowledge spillovers within narrowly defined sectors, as information on how to produce is transmitted

from firm to firm. Detailed empirical studies of productivity spillovers in developing countries confirm the importance of Marshall-Arrow-Romer externalities. A number of studies confirm the positive dependence of productivity on employment levels in the same city and industry for a number of countries (for example, Henderson 1988 for Brazil and the United States; Henderson and Kuncoro 1996 for Indonesia; Henderson, Lee, and Lee 2001 for Korea).

The acquisition and spread of knowledge is important in selling as well as in production. Roberts and Tybout (1997) establish the importance of learning effects in firms' export behavior, and Evenett and Venables (2003), using a fine commodity and geographical disaggregation, show that learning to export also has a destination-specific component.

Structural Change in India

One key objective of this paper is to argue that evidence accumulating at the microeconomic level on what drives structural change and growth should form the basis of policy formulation. This section focuses on one country, India, to illustrate this process in action. Even though our evidence base remains incomplete, the work described illustrates how microeconomic analysis at the subnational level is beginning to provide detail about what is important at the local level. It is this detail that is directly useful to policymakers.

Figure 3 shows the basic pattern of structural change in India. It graphs real agricultural and nonagricultural output per capita during 1960–97 for India's 16 main states. Agricultural output per capita remained relatively flat over the period, and economic growth was driven in large part by the rise in nonagricultural output. In 1960, average real nonagricultural and agricultural outputs per capita were similar across the states: Rs 406 and Rs 425, respectively. By 1997, at Rs 2,814, average nonagricultural output was more than double average agricultural output (Rs 1,266). Note, however, that the pattern differs markedly across states.

Figure 4 reveals a similar pattern in relation to manufacturing, which policymakers and economists often see as the engine of structural change.[4] Certain states—Andhra Pradesh, Gujarat, Haryana, Karnataka, Maharashtra, Punjab, and Tamil Nadu—show rapid growth in registered and unregistered manufacturing, while states such as Assam, Bihar, Jammu and Kashmir, Kerala, Uttar Pradesh, and West Bengal display relative stagnation (albeit from different base levels).

These differences in economic performance have profound consequences for welfare. India is unusual among developing countries in having carried out household expenditure surveys on a regular basis since the 1950s, which permits the construction of consistent and comparable poverty series (see Ozler, Datt, and Ravallion 1996). Figure 5 graphs rural and urban poverty headcounts (which capture the percentage of the rural and urban population in poverty) for the 16 main Indian states during 1958–2000. Comparing figures 3, 4, and 5, what is striking is that states that experienced rapid structural change also experienced faster rural and urban poverty reduction. In short, structural change and growth appear to matter for poverty reduction.[5]

FIGURE 3.
Output Per Capita in India
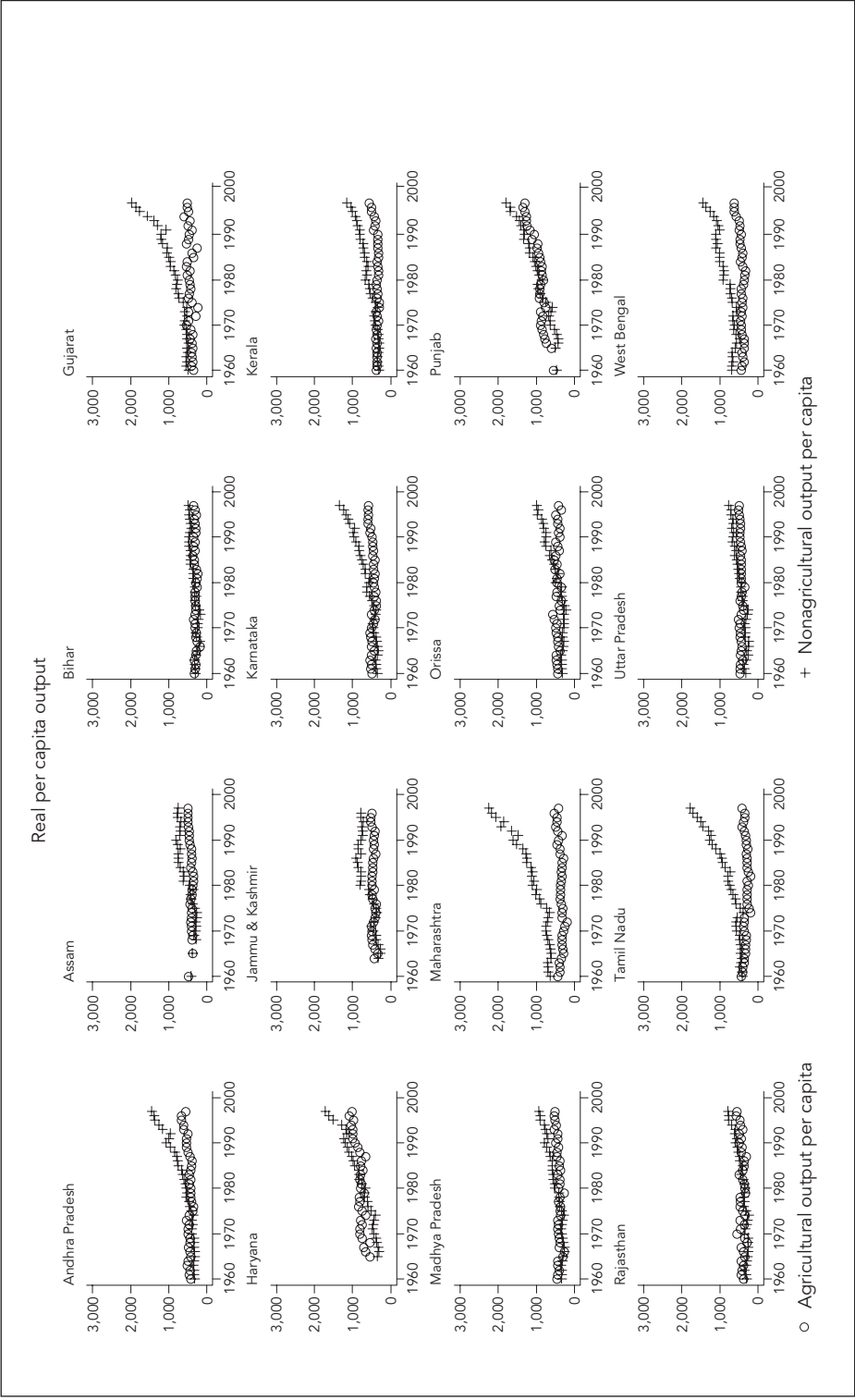
Source: Authors.

FIGURE 4.
Manufacturing Output Per Capita across Indian States, 1960–97

Real per capita output

o Registered manufacturing output   + Unregistered manufacturing output
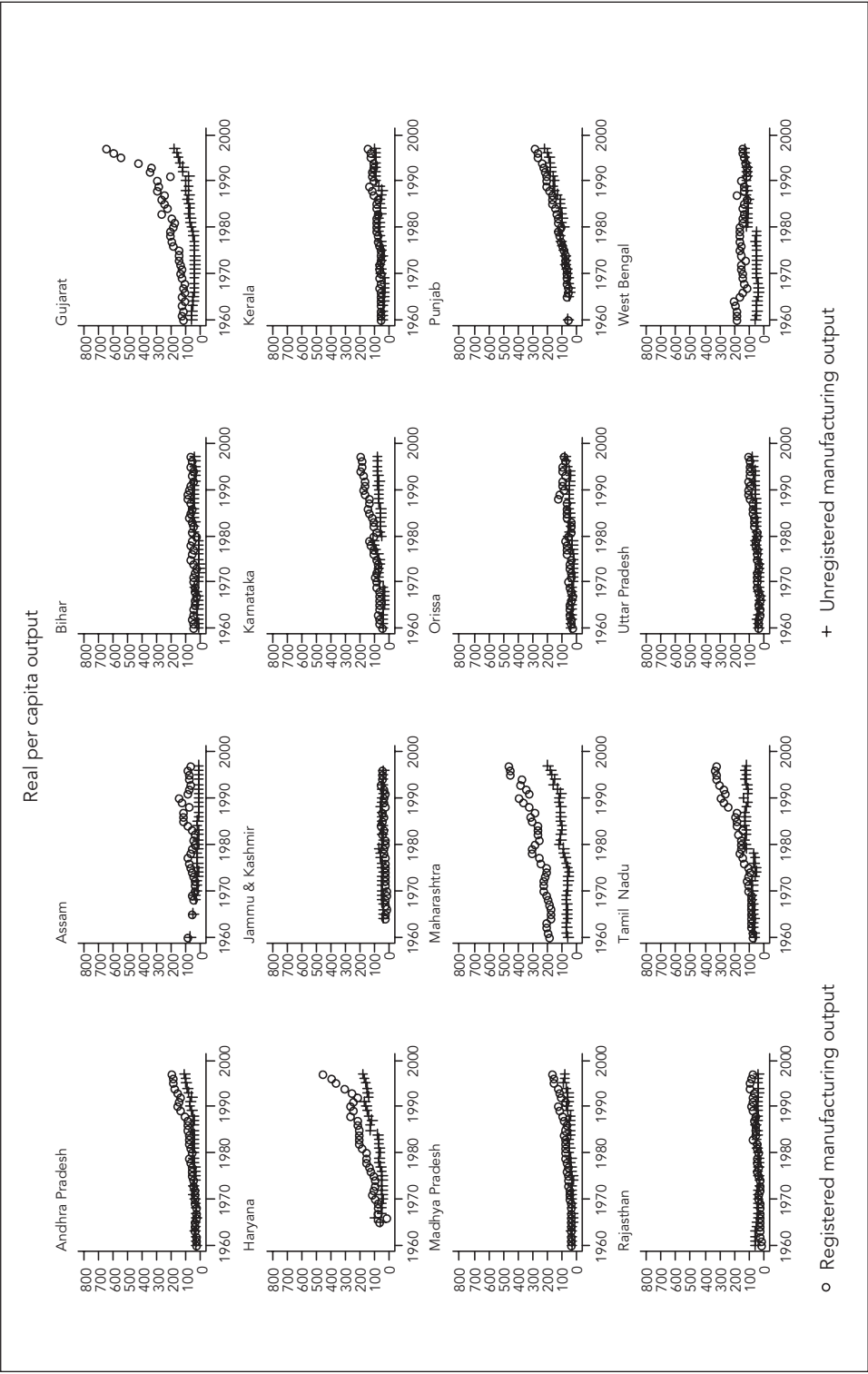
Source: Authors.

FIGURE 5.
Rural and Urban Poverty across Indian States, 1958–2000
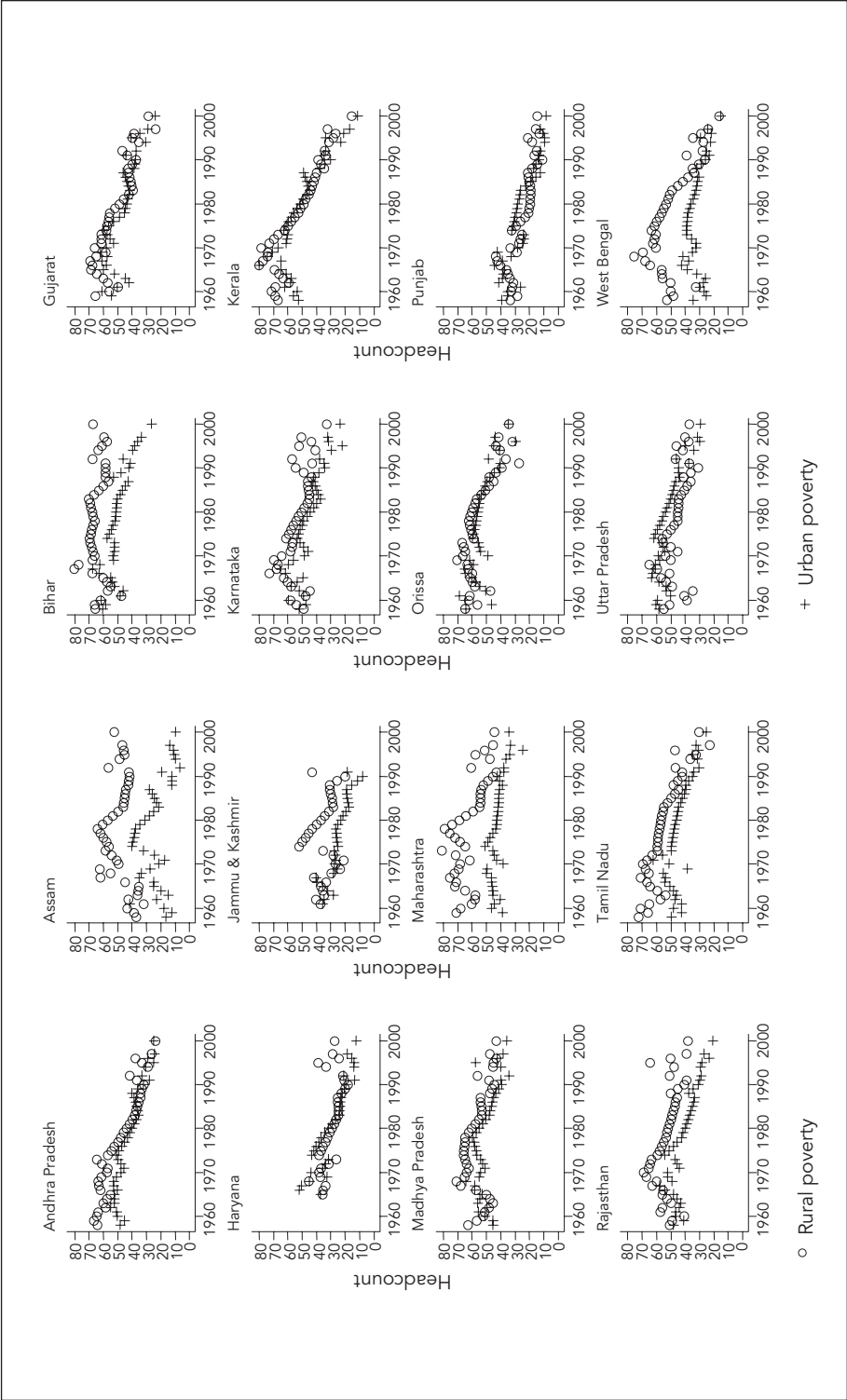
Simple regression analysis helps to bring this point home. Table 1 presents regressions of poverty measures on output per capita measures using panel data for the 16 main Indian states for 1958–1997. Column (1) shows that increases in the levels of non-agricultural output are associated with significant falls in overall poverty. In contrast, changes in agricultural output do not affect poverty. A similar pattern is apparent when we break poverty into its urban and rural components in columns (2) and (3): The growth of nonagricultural output is what is exerting a significant effect on urban and rural poverty. Columns (4) and (5) look at the influence of registered and unregistered manufacturing.[6] Column (4) indicates that increases in per capita registered manufacturing output reduce urban, but not rural, poverty. That makes sense, as those larger firms are located mainly in urban areas. In contrast, column (5) shows that increases in output by firms in the unregistered manufacturing sector are significantly negatively associated with rural poverty, but not urban poverty. Thus the entry and growth of small businesses is important for explaining the pattern of rural poverty reduction.

TABLE 1.
Poverty and Output across Indian States, 1958–97

	(1)	(2)	(3)	(4)	(5)
	Total poverty headcount	Urban poverty headcount	Rural poverty headcount	Urban poverty headcount	Rural poverty headcount
Model	OLS	OLS	OLS	OLS	OLS
Log agricultural OPC	4.22	1.46	4.60	1.53	3.49
	(1.37)	(0.48)	(1.27)	(0.56)	(0.91)
Log nonagricultural OPC	−16.27***	−9.34***	−16.18***		
	(3.00)	(2.42)	(2.23)		
Log registered manufacturing OPC				−4.51***	−0.40
				(3.20)	(0.15)
Log unregistered manufacturing OPC				−1.91	−5.99***
				(1.65)	(2.52)
State effects	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES
Adjusted R^2	0.83	0.88	0.79	0.89	0.79
Number of observations	318	318	318	318	318

OLS   Ordinary least squares estimation.

OPC   Output per capita

*** Significant at the 1 percent level.

Note: Absolute standard errors are shown in parentheses. Standard errors are adjusted for clustering by state. The poverty headcount is the percentage of the population below official Indian poverty lines, which are defined separately for rural and urban areas. Agricultural, nonagricultural, registered manufacturing, and unregistered manufacturing output are in log real per capita terms. The sample is a panel of the 16 main Indian states. Regressions only include years when national sample surveys were carried out.

Source: Authors, based on state-level poverty data provided by Ozler, Datt, and Ravallion (1996).

We have established that structural change matters for economic growth and poverty reduction in India. We now turn to the question of which institutions and policies account for different rates of structural change across Indian states, focusing on manufacturing. The fact that states within a centrally planned economy performed so differently in terms of manufacturing is striking (figure 4). To date, much of the literature has focused on whether India's industrial strategy, which involved extensive use of industrial licensing and tariff and nontariff barriers, led Indian manufacturing to perform poorly relative to other countries, in particular, those in East Asia (Bhagwati and Desai 1970; Bhagwati and Srinivasan 1975).[7]

National policies, however, have limited scope in explaining the marked cross-state variation in manufacturing performance. To understand these patterns we need to focus on state-specific policies and conditions. We examine three examples of research on how first advantage factors and the interplay between first advantage and second advantage factors determine the pattern of manufacturing growth in India. These examples illustrate how the framework we have developed can be used to build up a microeconomic evidence base on what matters.

Example 1—First Advantage: Labor Regulation and Registered Manufacturing

Researchers and policymakers have identified labor regulations as an important element of India's investment climate (Dollar, Iarossi, and Mengistae 2001; Stern 2001). Besley and Burgess (2004) examine whether labor regulation is an important source of first advantage that could help explain differences in manufacturing performance across Indian states. To do this they exploit two key facts, namely: (a) the labor regulations only apply to firms in the registered manufacturing sector, and (b) the Indian Constitution empowers state governments to amend central government legislation. The main piece of central legislation is the Industrial Disputes Act of 1947, which state governments have amended extensively since independence. Besley and Burgess (2004) read the text of each amendment and coded each as pro-worker (+1), neutral (0), or pro-employer (−1). Figure 6 shows the pattern of change across Indian states between 1958 and 1992 and reveals that the direction of regulation varies across states and time.

Besley and Burgess (2004) then check whether the pattern of regulatory changes shown in figure 6 affects the pattern of manufacturing development shown in figure 4. Column (1) of table 2 shows that moving in a pro-worker direction is associated with lower per capita levels of manufacturing output. This occurs because pro-worker labor regulation led to less output in registered manufacturing, as shown in column (2). Investment in this sector is lower in states with more pro-worker labor regulation. Column (3) shows that the effect goes the other way for unregistered manufacturing; that is, states with more pro-worker labor regulations tend to have larger informal manufacturing sectors. This makes sense, because where workers are able to extract more of the rents from production in the registered sector, capitalists will prefer to remain in the unregistered sector where labor has no power. Columns (4) and (5) indicate that regulating in a pro-worker direction is also associated with

FIGURE 6.
Labor Regulation across Indian States, 1960–2000
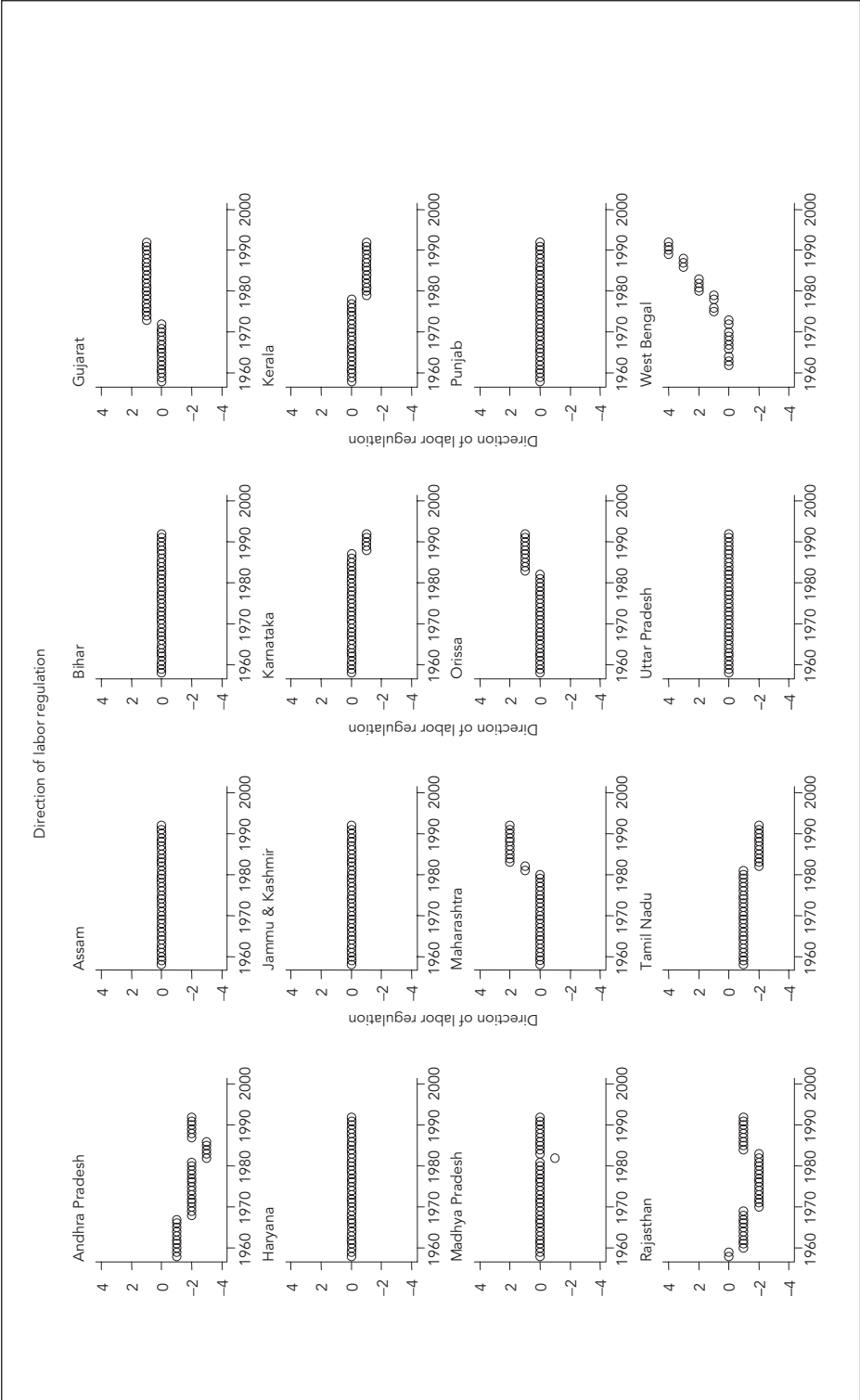
increases in urban poverty but does not affect rural poverty. This reflects the fact that the adverse effects of labor regulation are mainly felt in the registered sector, which is located primarily in urban areas. These results suggest that attempts to redress the balance of power between capital and labor can end up hurting the poor.

As Besley and Burgess (2004) show, the policy choices of state governments in India as regards labor regulation have strongly affected manufacturing performance. First advantage factors like regulation, which are, in part, under the control of sub-national governments, will have a strong bearing on whether or not manufacturing develops in areas under their jurisdiction, and this in turn will have welfare consequences for citizens in those regions. Note that the large differences in manufacturing performance were present well before liberalization in 1991.[8] This suggests that countries or regions wishing to develop manufacturing must pay attention to the first advantage factors that affect the business climate that firms face, and not only to the trade regime. We have clear evidence here that the institutional environment affects entrepreneurs' investment and location decisions.

Example 2—First Advantage: Rural Banks and Unregistered Manufacturing

Our focus so far has been on the registered manufacturing sector. However, most individuals in low-income countries do not have the choice of working in a factory, especially those residing in rural areas. For them the relevant choice might

TABLE 2.
Labor Regulation and Manufacturing Output across Indian States, 1958–92

	(1) Log total manufacturing OPC	(2) Log reg manufacturing OPC	(3) Log unreg manufacturing OPC	(4) Urban poverty headcount	(5) Rural poverty headcount
Model	OLS	OLS	OLS	OLS	OLS
Labor	−0.073**	−0.186***	0.086**	2.288***	−0.821
regulation [t − 1]	(2.05)	(2.90)	(2.46)	(3.31)	(0.48)
State effects	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES
Adjusted R^2	0.93	0.93	0.75	0.89	0.80
Number of					
observations	509	508	509	547	547

OLS   Ordinary least squares estimation.

OPC   Output per capita

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: Standard errors are adjusted for clustering by state. Absolute t statistics are in parentheses. Registered and unregistered manufacturing output is in log real per capita terms. The poverty headcount is the percentage of the population below official Indian poverty lines, which are defined separately for rural and urban areas. State amendments to the Industrial Disputes Act are coded 1 = pro-worker, 0 = neutral, and −1 = pro-employer, and then cumulated over the period to generate the labor regulation measure. The data are for the 16 main states.

Source: Besley and Burgess (2004).

be between remaining in agriculture or starting a small business. These new non-agricultural activities are typically more productive than basic agriculture, and thus engender both growth and poverty reduction. Understanding which factors drive structural change by facilitating the emergence of small businesses is a key challenge.

Burgess and Pande (2003) try to make some inroads into this issue by evaluating whether a massive expansion of rural bank branches in India affected structural change and growth. During 1961—2000, more than 30,000 new branches were opened in rural areas. The rationale for the program was simple. The government had identified lack of access to finance as a key reason why growth was stagnant and poverty was persistent in rural areas. It saw the failure of banks to enter rural areas as a brake on entrepreneurship and the emergence of new activities. To address this the central bank first nationalized commercial banks in 1969 and then imposed a license rule in 1977 that stated that for each branch opened in a location that was already served by a bank (typically urban), banks had to open four branches in locations without banks (typically rural). As a result states that had fewer banks per capita before the program in 1961 received more bank branches between 1977 and 1990, leading to both a reduction and an equalization in population per bank branch. Entrepreneurs, small business people, and agriculturalists were explicitly targeted in banks' mandated lending practices. The program is an example of the government targeting a key first advantage factor, access to capital, as a way to encourage the spread of new economic activities across rural India. The government rescinded the one to four rule in 1990, and branch building in rural areas came to a halt.

To evaluate the program, Burgess and Pande (2003) use the 1977 and 1990 trend breaks in the relationship between initial financial development and rural branch expansion attributable to license regime shifts as instruments for the number of branches opened in rural locations without banks. Some results are contained in table 3. Column (1) shows that rural branch expansion is associated with higher total manufacturing output per capita. Columns (2) and (3) show that the impact on manufacturing was felt mainly in the unregistered manufacturing sector, which is an important source of income in rural areas. Column (3), in particular, suggests that the arrival of banks in rural areas helped people to start and expand small businesses. By addressing a key first advantage factor, the (albeit forced) entry of banks spurred entrepreneurship and structural change. In contrast, as column (2) indicates, registered manufacturing, which is located mainly in urban areas, was unaffected. Column (4) reveals that the structural change engendered by rural branch expansion had positive consequences in terms of reducing rural poverty, and column (5) shows that, in contrast, urban poverty was unaffected.

This example brings home how a first advantage factor, access to finance, may be critical in enabling poor, rural residents to begin new economic activities. The arrival of rural banks in India enabled people to take on new production and employment activities and to exit poverty. The results highlight how important access to finance is for encouraging entrepreneurship, structural change, and growth.

TABLE 3.
Rural Banks and Manufacturing Output across Indian States, 1961–2000

	(1)	(2)	(3)	(4)	(5)
	Log total manufacturing OPC	Log reg. manufacturing OPC	Log unreg. manufacturing OPC	Rural poverty headcount	Urban poverty headcount
Model	IV	IV	IV	IV	IV
Number of branches opened in rural, unbanked locations	−0.15* (2.10)	0.04 (0.57)	0.28* (1.85)	−5.02** (2.14)	−0.82 (0.92)
State effects	YES	YES	YES	YES	YES
Year effects	YES	YES	YES	YES	YES
Adjusted R^2	0.94	0.94	0.81	0.75	0.92
Number of observations	579	579	579	627	627

IV   Instrumental Variables estimation.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

Note: Standard errors are adjusted for clustering by state. Absolute t statistics are in parentheses. Registered and unregistered manufacturing output is in log real per capita terms. The poverty headcount is the percentage of the population below official Indian poverty lines, which are defined separately for rural and urban areas. The number of bank branches opened in rural, unbanked locations is normalized by the 1961 population. The two instruments for this variable are the number of locations with banks in 1961 per capita interacted with (a) an indicator variable that equals 1 if the year is later than 1976 and a post-1976 time trend, and (b) an indicator variable that equals 1 if the year is later than 1989 and a post-1989 time trend, respectively. The sample is a panel of the 16 main Indian states.

Source: Burgess and Pande (2003).

Example 3—Interplay between First and Second Advantage: Technological Capability and Trade Liberalization

How do second advantage factors, which are intrinsic to firms and industries, interact with first advantage factors, which affect the environment in which firms function? We have argued that the interplay of factors is critical to understanding modern economic growth. Such interactions are difficult to study empirically, as second advantage factors are themselves endogenous.

One way to get some purchase on this issue is to look at how the technological capability of firms or industries affects how they respond to a liberalization shock. That is, in line with the theory presented earlier, we might expect firms in the same industry to respond differently to the threat of entry depending on how competitive they are with the new firms and products. How firms respond will in part depend on the technological choices they have made in the past. Firms that are close to the technological frontier are in a better position to compete with new firms and products. They may respond by investing in new technology to retain their existing markets and to benefit from increased access to new markets. In contrast, investment incentives are likely to be blunted in laggard firms, and liberalization is likely to harm those

firms. Thus responses to the same liberalization shock are likely to be heterogeneous across geography and within sectors. Heterogeneity in post-reform performance is also likely to be magnified if the institutional environment in which firms are located varies. That is, interactions between market access and other first advantage factors may lead to differential performance in two firms that have identical technological capabilities.

To look at this issue of interplay between first and second advantage factors, Aghion and others (2003) exploit India's massive trade liberalization in 1991. This shock was common across firms in the same industry; however, firms in different states in the same three-digit industry varied in terms of their level of pre-reform productivity, which can be taken as proxy of their technological capability. As discussed earlier, on dimensions such as labor regulation, the institutional environment in which firms functioned also varied across Indian states.

Table 4 reports how interactions between the liberalization reform (captured by a dummy variable that switches on in 1991), the pre-reform technological capability (captured by the ratio of labor productivity in a three-digit state industry relative to the most productive state industry in India), and the pre-reform investment climate at the state level (captured by the labor regulation measure) affected performance in India's registered manufacturing sector. We look at these relationships using a panel of three-digit state industries for 1980–97. The first row of table 4 indicates that state industries with higher pre-reform technological capability (that is, closer to the Indian productivity frontier for that industry) saw greater increases in output, employment, labor productivity, and total factor productivity following reform. This shows that state industries with greater technological capability benefited more from liberalization. In contrast, laggard state industries that were below India's median productivity experienced below trend rates of increase in output, employment, labor productivity, and total factor productivity following liberalization in 1991. Manufacturing performance is therefore a function of the interplay between a first advantage factor (market access) and a second advantage factor (technological capability).

In line with the results presented in table 2, the second row of table 4 shows that state industries located in states with more pro-worker labor regulation saw less growth in output, employment, labor productivity, and total factor productivity. Thus the institutional environment (as captured by first advantage factors such as labor regulation) in which three-digit industries function affects how well they perform.

Moreover, the third row of table 4, which reports results for the interaction between labor regulation and liberalization, reveals that the negative impact of pro-worker regulation is magnified following liberalization. That is, for state industries to be located in a pro-worker state when market access is increased is even more damaging in terms of output, employment, labor productivity, and total factor productivity growth.

The findings presented in table 4 emphasize that the initial level of technology and institutional context mattered for whether and to what extent industries and states in India benefited from liberalization. These findings have interesting implications for

TABLE 4.
Technological Capability, Liberalization, and Manufacturing Performance in India, 1980–97

	(1)	(2)	(3)	(4)
	Log total registered manufacturing output	Log registered manufacturing employment	Log registered manufacturing OPW	Log total factor productivity
Model	OLS	OLS	OLS	OLS
Pre-reform technological	0.439***	0.211***	0.228***	0.159***
capability* reform	(4.66)	(2.85)	(3.37)	(2.44)
Labor regulation	−0.090***	−0.035*	−0.055**	−0.070***
	(3.88)	(1.69)	(1.98)	(3.13)
Labor regulation* reform	−0.061***	−0.052***	−0.010	−0.036***
	(4.38)	(4.39)	(0.17)	(4.12)
State industry fixed effects	YES	YES	YES	YES
Industry time trends	YES	YES	YES	YES
Year effects	YES	YES	YES	YES
R^2	0.94	0.96	0.94	0.64
Number of observations	22,883	22,883	22,883	22,883

OLS   Ordinary least squares estimation.

OPW   Output per worker

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: Standard errors are adjusted for clustering by state industry. Absolute t statistics are in parentheses. Pre-reform technological capability is pre-reform state industry labor productivity relative to the state with the highest level of pre-reform labor productivity within the industry. Reform is a dummy that equals 0 before 1991 and equals 1 from 1991 onward. State amendments to the Industrial Disputes Act are coded 1 = pro-worker, 0 = neutral, and −1 = pro-employer, and are then cumulated over the period to generate the labor regulation measure. The sample is a three-dimensional, unbalanced panel of three-digit industries in the 16 main Indian states.

Source: Aghion and others (2003).

policy. First, in line with the theory outlined earlier, the level of productivity before reform clearly matters for the trajectory after liberalization. The key importance of the interplay between first advantage and second advantage factors is that it has a critical bearing on the pattern of growth following liberalization. This, of course, begs the question of why levels of productivity varied in different three-digit industries in different states in the first place. While that is beyond the scope of the current analysis, it does indicate that what determines firms' industrial capabilities should constitute a central area of future research (see Sutton 2002).

Second, the institutional environment affects the extent to which liberalization will be growth enhancing. For instance, rigidities in the labor market may limit the positive impact of trade liberalization. The institutional and policy choices that state governments make have a central bearing on the extent to which the regions they govern will benefit or not from liberalization.

Third, liberalization may have adverse effects on industries and regions that are initially less developed. This may call for complementary measures to offset the negative distributional consequences of reforms, for example, investing in infrastructure and supporting knowledge acquisition in backward areas.

Conclusions

Determining the factors that drive structural change and growth has always been central to development economics. As growth theory has developed and disaggregated data have become available, this work has been moving in a microeconomic direction. Some broad correlations are emerging in new cross-country work that tries to explain the different growth experiences of countries. These serve as useful signposts for more detailed microeconomic work, although the process is still at an early stage. For example, despite some consensus that access to finance matters for entrepreneurs, how such access should be provided (for example, through private banks, development banks, microfinance, and so on) is unclear, and what is appropriate may depend on the type of business and the extent of development of the legal system and other institutions.

The microeconomic approach described in this paper represents the frontier for identifying the first advantage preconditions for growth. This approach, based on careful building of microeconomic evidence, has become standard in industrial economies and is now taking off in research on low-income countries. Consensus is growing that evidence accumulating via the evaluation of policy and institutional reforms represents the best approach for identifying what works. The first advantage examples provided from India illustrate this process at work.

As this paper makes plain, however, identifying sources of first advantage is not the whole story. The spatial and sectoral pattern of modern economic growth suggests that it is an outcome of a complex interaction of first advantage factors and second advantage factors. These latter factors, which include technological capabilities, technological spillovers, market linkages, and networks, are the results of investment and location decisions made by entrepreneurs. The externalities that they create shape future growth. As the example from India brings out, the interaction between first advantage and second advantage is what determines where and in which sectors growth takes place. This observation brings into sharp relief the fact that modern economic growth is the result of interactions between institutions and organizations. Understanding the factors that drive firms' technological, investment, and location decisions is therefore central to understanding modern economic growth. Those decisions will themselves be a function of institutions and other first advantage factors, which makes the empirical study of growth challenging.

Two key policy lessons emerge from our analysis. First, in the absence of first advantage elements, rapid structural change and growth are unlikely to take place. Much recent work therefore focuses on uncovering the microeconomic roots of backwardness. Second, ensuring that first advantage elements are present is not a

guarantee that rapid economic growth will take place in a given location. Consideration of second advantage factors at the microeconomic level is therefore central to understanding economic growth, and here our knowledge base is extremely thin.

This paper constitutes a modest step toward understanding which factors drive growth at the microeconomic level. Given the strong links between innovation, growth, and poverty reduction, few endeavors in economics are more important.

Notes

1. Differing levels across regions mainly reflect country size; it is the change over time to which we draw attention.
2. The figures are generated from a simple numerical example, available on request.
3. The main point is the rapid growth of new locations. The jagged saw-tooth effect occurs as some workers are drawn from existing locations, and is not central to the argument.
4. In India, manufacturing is divided into a registered or organized sector, which is regulated and encompasses firms with employment levels above 10 if they have electricity and above 20 if they do not, and an unregistered or unorganized sector, which encompasses firms below these thresholds.
5. Using Indian state-level data, Datt and Ravallion (2002) establish a robust link between income growth and poverty reduction. They find that the elasticity of the incidence of poverty with respect to net domestic product per capita was –0.75 for the period 1958–91. This figure is close to the elasticity of –0.76 that Besley and Burgess (2003b) find using cross-country data for developing economies.
6. Over the period, the registered sector accounted for about 9 percent of state output and the unregistered sector for 5 percent.
7. Between 1960 and 1995, manufacturing as a share of GDP grew from 9 to 24 percent in Indonesia, 8 to 26 percent in Malaysia, and 12.5 percent to 28.0 percent in Thailand. In contrast, manufacturing output in India constituted 13 percent of GDP in 1960, but grew to only 18 percent of GDP by 1995 (World Bank various years).
8. Besley and Burgess (2004) restrict their econometric analysis to the 1958–92 period to better identify the impact of state-level policies.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–1401.

Aghion, Philippe, and Peter Howitt. 1998. *Endogenous Growth Theory.* Cambridge, Mass.: MIT Press.

_____. 2003. "Growth with Quality Improving Innovations." Harvard University, Cambridge, Mass. Processed.

Aghion, Philippe, Robin Burgess, Stephen Redding, and Fabrizio Zilibotti. 2003. "The Unequal Effects of Liberalization: Evidence from India." London School of Economics, London. Processed.

Aranya, Rollee. 2002. "Spatial Disparity in the Globalizing City: The Case of the Silicon Plateau Emerging in Bangalore." University of Trondheim, Trondheim, Norway. Processed.

Banerjee, Abhijit, and Andrew Newman. 2003. "Inequality, Growth, and Trade Policy." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Banerjee, Abhijit V., Paul J. Gertler, and Maitreesh Ghatak. 2002. "Empowerment and Efficiency: Tenancy Reform in West Bengal." *Journal of Political Economy* 110(2): 239–80.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2003. "Improving the Quality of Education in India: Evidence from Three Randomized Experiments." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Besley, Timothy, and Robin Burgess. 2000. "Land Reform Poverty and Growth: Evidence from India." *Quarterly Journal of Economics* 115(2): 389–430.

_____. 2003a. "Court Efficiency and Economic Performance in India." London School of Economics, London. Processed.

_____. 2003b. "Halving Global Poverty." *Journal of Economic Perspectives* 17(3): 3–22.

_____. 2004. "Can Labor Regulation Hinder Economic Performance? Evidence from India." *Quarterly Journal of Economics* 119(1): 91–134.

Bhagwati, Jagdish, and Padma Desai. 1970. *India: Planning for Industrialization.* Delhi: Oxford University Press.

Bhagwati, Jagdish, and T. N. Srinivasan. 1975. *Foreign Trade Regimes and Economic Development.* Delhi: McMillan.

_____. 1999. *Outward Orientation and Development: Are the Revisionists Right?* Discussion Paper no. 806. New Haven, Conn: Yale University, Economic Growth Center.

Botero, Juan Carlos, Simeon Djankov, Rafael La Porta, Florencio López-de-Silanes, and Andrei Shleifer. 2003. "The Regulation of Labor." Working Paper no. 9756. National Bureau of Economic Research, Cambridge, Mass.

Burgess, Robin, and Rohini Pande. 2003. "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment." London School of Economics, London, and Columbia University, New York. Processed.

Card, David. 1999. "The Causal Effect of Education on Earnings." In Orley Ashenfelter and D. Card, eds., *Handbook of Labor Economics.* Amsterdam: Elsevier.

Chakravorty, S., and S. V. Lall. 2003. "Diversity Matters: The Economic Geography of Industry Location in India." Working Paper no. 3072. World Bank, Washington, D.C.

Chattopadhyay, Raghabendra, and Esther Duflo. 2001. "Women as Policy Makers: Evidence from an India-Wide Randomized Policy Experiment." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Chenery, Hollis B. 1960. "Patterns of Industrial Growth." *American Economic Review* 50(4): 624–54.

Chenery, Hollis B., and Lance J. Taylor. 1968. "Development Patterns: Among Countries and Over Time." *Review of Economics and Statistics* 50(4): 391–416.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86(1): 54–70.

Cikurel, Daniel Chiquiar. 2002. "Why Mexico's Regional Convergence Broke Down." University of California, San Diego. Processed.

Datt, Gaurav, and Martin Ravallion. 2002. "Is India's Growth Leaving the Poor Behind" *Journal of Economic Perspectives* 16(3): 89–108.

Demurger, S., J. Sachs, W. Woo, S. Bao, G. Change, and A. Mellinger. 2002. "Geography, Economic Policy, and Regional Development in China." University of California, Davis. Processed.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117(1): 1–37.

Djankov, Simeon, Edward L. Glaeser, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer. 2003. "The New Comparative Economics." Discussion Paper no. 3882. Centre for Economic Policy Research, London

Dollar, David, and Aart Kraay. 2000. "Growth Is Good for the Poor." World Bank, Development Research Group, Washington, D.C. Processed.

Dollar, David, Giuseppe Iarossi, and Taye Mengistae. 2001. "Investment Climate and Economic Performance: Some Firm Level Evidence from India." Paper prepared for the Economists' Forum, May 3–4, World Bank, Washington, D.C.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91(4): 795–813.

Evenett, Simon, and Anthony J. Venables. 2003. "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows." London School of Economics, London. Processed.

Evenson, R. E., and L. E. Westphal. 1995. "Technological Change and Technology Strategy." In Hollis B. Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 3a. Amsterdam: North-Holland.

Field, Erica. 2002. "Entitled to Work: Urban Property Rights and Labor Supply in Peru." Princeton University, Princeton, N.J. Processed.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3): 379–99.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, Mass.: MIT Press.

Gallup, John L., and Jeffrey Sachs (with Andrew Mellinger). 1999. "Geography and Economic Development." In Boris Pleskovic and Joseph 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(1): 136–56.

Glaeser, Edward L., Simon Johnson, and Andrei Shleifer. 2001. "Coase vs. the Coasians." *Quarterly Journal of Economics* 116(3): 853–99.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2000. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Harvard University, Department of Economics, Cambridge, Mass. Processed.

Hall, Robert E., and Chad Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114(1): 83–116.

Hallward-Driemeier, M., G. Iarossi, and K. L. Sokoloff. 2002. "Exports and Manufacturing Productivity in East Asia: A Comparative Analysis with Firm-Level Data." Working Paper no. 8894. National Bureau of Economic Research, Cambridge, Mass.

Hausmann, R., and D. Rodrik. 2002. "Economic Development as Self-Discovery." Working Paper no. 8952. National Bureau of Economic Research, Cambridge, Mass.

Henderson, J. V. 1988. *Urban Development: Theory, Fact, and Illusion*. Oxford, U.K.: Oxford University Press.

_____. 1999. "How Urban Concentration Affects Economic Growth." Policy Research Working Paper no. 2326. World Bank, Washington, D.C.

Henderson, J. V., and A. Kuncoro. 1996. "Industrial Centralization in Indonesia." *World Bank Economic Review* 10(3): 513–40.

Henderson, J. V., and H. G. Wang. 2003. "Urbanization and City Growth." Brown University, Providence, R.I. Processed.

Henderson, J. V., T. Lee, and Y.-J. Lee. 2001. "Scale Externalities in Korea." *Journal of Urban Economics* 49(3): 479–504.

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

Hobday, M. 1995. *Innovation in East Asia: The Challenge to Japan.* Aldershot, U.K.: Edward Elgar.

Holmes, T. J. 1999. "Localisation of Industry and Vertical Disintegration." *Review of Economics and Statistics* 81(2): 314–25.

Hsieh, Chang-Tai, and Miguel Urquiola. 2002. "When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program." Princeton University, Princeton, N.J. Processed.

Jaffe, A., M. Tratjenberg, and J. V. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108(3): 577–98.

Johnson, Simon, John McMillan and Christopher Woodruff. 2002a. "Courts and Relational Contracting." *Journal of Law, Economics, and Organization* 18(1): 221–77.

_____. 2002b. "Property Rights and Finance." *American Economic Review* 92(5): 1335–56.

Klien, Shira. 2003. "Human Capital Externalities in India." London School of Economics, London. Processed.

Kremer, Michael, and Edward Miguel. 2001. "Worms: Education and Health Externalities in Kenya." Working Paper no. 8481. National Bureau of Economic Research, Cambridge, Mass.

Krueger, Alan B., and Mikael Lindhal. 2001. "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39(4): 1101–36.

Krugman, P. 1981. "Trade, Accumulation, and Uneven Development." *Journal of Development Economics* 8(2): 149–61.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1997. "The Legal Determinants of Finance." *Journal of Finance* 52(3): 1131–50.

_____. 1998. "Law and Finance." *Journal of Political Economy* 106(6): 1113–55.

Leamer, E. 1984. *Sources of Comparative Advantage.* Cambridge, Mass.: MIT Press.

_____. 1987. "Paths of Development in the Three Factor N-Good General Equilibrium Model." *Journal of Political Economy* 95(5): 961–99.

Levine, Ross. 1998. "The Legal Environment, Banks, and Long-Run Economic Growth." *Journal of Money, Credit, and Banking* 30(3): 596–613.

Lewis, W. A. 1954. "Economic Development with Unlimited Supplies of Labour." *Manchester School of Economic and Social Studies* 22(May): 139–91.

Limao, N., and A. J. Venables. 2001. "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade." *World Bank Economic Review* 15(3): 451–79.

Lin, Justin Yifu. 1992. "Rural Reforms and Agricultural Growth in China." *American Economic Review* 82(1): 34–51.

McCulloch, N., L. A. Winters, and X. Cirera. 2002. *Trade Liberalization and Poverty: A Handbook*. London: Centre for Economic Policy Research.

McMillan, John, and Christopher Woodruff. 1999. "Interfirm Relationships and Informal Credit in Vietnam." *Quarterly Journal of Economics* 114(4): 1285–1320.

———. 2002. "The Centrality of Entrepreneurs in Transition Economies." *Journal of Economic Perspectives* 16(3): 153–70.

Moretti, E. Forthcoming. "Human Capital Externalities in Cities." In V. Henderson and J. Thisse, eds., *Handbook of Regional and Urban Economics*, vol. 4. New York: Elsevier Science.

Murphy, K., A. Shleifer, and R. W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97(5): 1003–26.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge, U.K.: Cambridge University Press.

Ozler, B., G. Datt, and M. Ravallion. 1996. "A Data Base on Poverty and Growth in India." World Bank, Washington, D.C. Processed.

Pavcnik, N. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies* 69(1): 245–76.

Puga, D., and A. J. Venables. 1999. "Agglomeration and Economic Development: Import Substitution Versus Trade Liberalisation." *Economic Journal* 109(455): 292–311.

Qian, Y. 2003. "How Reform Worked in China." In D. Rodrik, ed., *In Search of Prosperity: Analytic Narratives of Economic Growth*. Princeton, N.J.: Princeton University Press.

Rajan, Raghuram G., and Luigi Zingales. 1998. "Financial Dependence and Growth," *American Economic Review* 88(3): 559–86.

Redding, S., and A. J. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics* 62(1): 53–82.

Roberts, Mark J., and James R. Tybout. 1997. "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review* 87(4): 545–64.

Rodriguez, F., and D. Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In B. Bernanke and K. S. Rogoff, eds., *Macroeconomics Annual 2000*. Cambridge, Mass.: MIT Press.

Rodrik, Dani. 2003. "Growth Strategies." Harvard University, Cambridge, Mass. Processed.

Rosenthal, S. S., and W. C. Strange. Forthcoming. "Evidence on the Nature and Sources of Agglomeration Economies." In V. Henderson and J. Thisse, eds., *Handbook of Regional and Urban Economics*, vol. 4. New York: Elsevier Science.

Rostow, W. W. 1960. *The Stages of Economic Growth*. Cambridge, U.K.: Cambridge University Press.

Schott, P. 2003. "One Size Fits All? Heckscher-Ohlin Specialization in Global Production." *American Economic Review* 93(2): 686–708.

Shishido, H., and W. Wheaton. 1982. "Urban Concentration, Agglomeration Economies, and the Level of Economic Development." *Economic Development and Cultural Change* 30(1): 17–30.

Sokoloff, Kenneth. 1986. "Productivity Growth in Manufacturing in Early Industrialization: Evidence from the American Northeast, 1820—1860." In S. Engerman and R. Gallman, eds., *Long Term Factors in American Economic Growth*. Chicago: Chicago University Press.

Stern, Nicholas. 2001. *A Strategy for Development*. Washington, D.C.: World Bank.

_____. 2003. "Poor People and Economic Opportunity: The Dynamics of Development." Draft write-up of Munich Lectures in Economics, November 19–21, 2002, Ludwig-Maximilians-Universität, Centre for Economic Studies, Munich.

Sutton, John. 2002. "Rich Trades, Scarce Capabilities: Industrial Development Revisited." *Economic and Social Review* 33(1): 1–22.

Trefler, D., and S. Zhu. 2001. "Ginis in General Equilibrium: Trade, Technology, and Southern Inequality." Working Paper no. 8446. National Bureau of Economic Research, Cambridge, Mass.

Venables, A. J. 2003. "Spatial Disparities in Developing Countries: Cities, Regions, and International Trade. London School of Economics, London. Processed.

Williamson, J. G. 1965. "Regional Inequality and the Process of National Development." *Economic Development and Cultural Change* 13(4): 3–45.

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

Yeats, A. 1998. "Just How Big Is Global Production Sharing." Policy Research Working Paper no. 1871. World Bank, Washington, D.C.

Young, A. 1991. "Learning by Doing and the Dynamic Effects of International Trade." *Quarterly Journal of Economics* 106(2): 369–406.

Comment on "Entrepreneurship, Innovation, and Growth" by T. N. Srinivasan and "Toward a Microeconomics of Growth" by Robin Burgess and Anthony J. Venables

GUR OFER

This is my second visit to India; the first was about a dozen years ago. The difference is striking. On both occasions I came on an El-Al flight from Tel Aviv with a first stop in Mumbai. During the first trip, the plane was full of young Israeli backpackers coming to discover the historical and spiritual treasures of India on a dollar-a-day basis. This time the plane was full of slightly older business people and engineers, busy with their laptops, and with tightly scheduled meetings in Mumbai, Bangalore, and Delhi. During this short timespan between my two visits, India joined the global economy. A short tour in Bangalore also testifies to this. The two papers that I review here describe and discuss the main tools whereby India has achieved this goal.

The papers by T. N. Srinivasan and by Robin Burgess and Anthony J. Venables are important papers that apply various aspects of microeconomic approaches to growth and to the study of economic development. The following is a discussion of each paper separately, but there is a significant amount of overlap. Being mostly a student of transition economics, I tried to make some comparative references, but constrained by space, I had to delegate most of them to footnotes and references. Naturally the discussion covers only a selection of the issues raised in the papers.

Comments on "Entrepreneurship, Innovation, and Growth"

The first paper, by T. N. Srinivasan, surveys a number of key issues pertaining to the theory of entrepreneurship and innovation in the framework of endogenous growth and their implications for the process of economic development. Srinivasan extends the discussion to development mostly through the concept of the product cycle, whereby the South imitates new products already invented and developed in the North for its own use and for exports. The product cycle works through trade, foreign direct investment (FDI), and two-way movements of professionals; therefore, in

Gur Ofer is emeritus professor of economics at the Hebrew University of Jerusalem in Israel.

almost all cases, openness is demonstrated as the preferred policy option. This is so with respect to the basic innovation process (mostly in accordance with Helpman 1990) and to a number of major recent changes and innovations in the character and structure of the modern firm (along the lines of Acemoglu, Aghion, and Zilibotti 2003) and in the sphere of financial services (per Lerner 1996, 1998; Rajan and Zingales 2001, 2003), symbolized by the growth of venture capital and of intellectual property rights (based on Gans, Hsu, and Stern 2002).

I agree that openness, even more so in recent years than before, is a major source of growth for developing countries. Many of the critical inputs needed for growth—capital, managerial skills, knowledge, innovations, and even institutions—can be imported in greater amounts, faster, and more cheaply than just a couple of decades ago. The dynamic development of information technology, in all its aspects, has made movement across borders faster, cheaper, and less risky. Thus many products, mostly services, that were considered nontradable just a generation ago are now tradable, and their volume is growing. Now included under tradables, I also note the relative ease with which many services can be moved across borders and the improved capabilities for doing so as multinational corporations that have developed and specialized in such activities are involved in FDI (Keren and Ofer 2002). To be able to take advantage of these developments, two conditions have to be met: the suppliers and producers of capital, knowledge, services, and so on must be willing or tempted by price and other incentives, and the countries at the receiving end must be ready to pay the price. The price is indeed high, and the demand (the potential benefits) and supply and the risk level and obstacles to entry will determine its level.

The high costs of FDI and of the transfer of innovations and services, both the potential benefits thereof and the risks and obstacles on the way, depend on the institutional infrastructure in the receiving country. Proper institutional development therefore becomes a major factor in obtaining the most important inputs, and must thus accompany and complement opening up. Two types of institutions are pertinent in this regard: first, those that develop human capital, and second, those that secure the proper rules of the market game. Hence developing economies have to invest more in education, training, and research and development (R&D) than their present levels, which are typically equivalent to just a small percentage of their gross domestic product (East Asia was an exception). This is necessary merely to be able to properly absorb knowledge-intensive imports and to imitate the production of advanced products.

The second type of institution is proper market institutions, which are needed to facilitate FDI in financial and innovation services and to attract capital. Some of those institutions, such as appropriate government, law enforcement, protection of property and contracts, and a level playing field for all those engaged in business, must be created and/or improved domestically, as large parts thereof are still nontradable or not importable through FDI.[1] However, some key institutions could be imported from abroad by means of FDI, such as banks and other financial institutions, insurance services, accounting, and consulting.

One of the main obstacles to growth in developing economies is the lack of proper financial institutions. With the development in industrial economies of even more sophisticated versions of financial institutions, such as venture capital firms, the difficulties for developing economies have become even greater.[2]

Also the new more flexible and dynamic forms of firms discussed by Srinivasan and many others are much more difficult for countries with a weak institutional infrastructure to establish and handle, but they could possibly be more easily moved to or be established as parts of larger global organizations in developing economies. Those modern business services now account for a considerable part of the new products produced in the North as their contribution to the new and dynamic product cycle. Indeed, one can visualize a new type of product cycle in which industrial economies produce new institutional and market services that the developing economies can import and imitate to accelerate their modernization and growth. Therefore concentrating on the needed complementary, nontradable segment of institutional development is probably the most cost-effective investment in developing economies today.

The foregoing calls for more government intervention in developing economies at a time when they are also weaker, less capable than the industrial countries, and infested with corruption and state capture. The market failures that call for government intervention confront equally serious government failures. This is also why the policy of selecting winners (industries with a good chance of success), which can be shown, theoretically, to be justified in some cases during early stages of development, is so much in dispute. This is the main source of the debate between people such as Stiglitz (2002) on the one side, and Bruno (Stiglitz 1997) and Easterly (2001) on the other.

All available evidence shows that the institutional environment in developing economies (and in transition economies) is, in general, weak. An analysis based on materials collected by Kaufmann, Kraay, and Zoido-Lobaton (2001, 2002) across six clusters of governance indicators gives low grades to developing economies, even those at middle-income levels (Ofer 2003).

This is not the place to go into an analysis of the reasons for this except for one, considered major, that is also mentioned in passing in Srinivasan's paper. This is the difficulty, for political economy reasons, of shifting the rules of the game from an economic regime based on extensive government intervention and semi-autarky to a less regulated, open system. These two institutional setups are orthogonal to each other with respect to the nature of incentives and the targets of entrepreneurial energies. Even if extensive government intervention could be justified for some period of time, for example, during an early imitation or import substitution period (in accordance with Acemoglu, Aghion, and Zilibotti 2003), the shift to the new system is usually delayed well beyond the optimum timing, and then the transition is extremely costly in terms of both political and economic costs. A glaring example of this is Japan, where, following a period of rapid growth of more than 30 years, the country sank into a prolonged stagnation, fully explained by its failure to shift from one strategy to the other, including across two different institutional settings.[3] There is also the

dismal experience of the former Soviet Union and other transition economies.[4] Considering the high cost of the transition, the choice between a sequence of different growth strategies to fit requirements at different stages of development (even if justified) and of a single, uniform strategy throughout, though somewhat less fit for any stage, becomes a difficult one (Ofer 2004).

Improved institutions will also help bring down the high price developing economies have to pay for FDI and imported services and in the form of unfavorable terms of trade.[5] The high price of imported services is one of the main grievances of developing economies in the recent trade rounds sponsored by the World Trade Organization (subsidies to agriculture in industrial economies is another) and in the more general debate about globalization. Speeding up the institutional transition is something developing economies can do to lower this price. Part of this high price is a risk premium, and the lower the risk, the lower the price. The high price also reflects in part the potentially large joint profits and welfare gains, and is therefore worth paying.[6] However, the high price may also be partly due to the monopoly power, or just the superior economic power, of the multinationals involved, and they may be collecting too high a share of the total benefit that is being created.

The disagreements about the opening up of trade in services and about intellectual property rights between developing economies and industrial economies (including in relation to pharmaceuticals) reflect this feeling on the part of the developing economies. The industrial economies' refusal to reduce their trade barriers for agricultural, textile, and other products in which developing economies have a comparative advantage is another major grievance. While there is some competition in those markets (among the multinationals), the developing economies probably do receive a smaller share of the incremental benefit.

As Srinivasan notes, debate is also ongoing about the fair share that developing economies should pay for the R&D costs of developing new products (and drugs). Note that many citizens of industrial economies receive most drugs with small co-payments, and most poor citizens of these countries do not even pay any of the taxes that finance part of the R&D effort.

The global economy lacks a global authority that can redistribute income across countries to support, for example, the education and R&D efforts of developing economies. Soros (2002) and Stiglitz (2002), among others, point to the downward pressure of the globalization process on public budgets for public goods and services. In developing economies, these budgets are much too small even without globalization and for obvious reasons. A global arrangement that would facilitate the transfer of funds for these institutional needs could benefit all parties.

Endogenous growth is usually associated with externalities that their creators do not tap. In Srinivasan's paper, the main emphasis is on the intergenerational externalities of the learning-by-doing paradigm,[7] hence the justification for government intervention to support more investment in innovation than would otherwise take place. This can be done by providing a general subsidy and by supporting R&D in nonprofit organizations, such as universities and research institutes. In addition, as mentioned earlier, another reason for advocating openness for developing

countries is that they can obtain the benefits of industrial countries' R&D efforts by importing knowledge-intensive goods, and, of course, through FDI.

One of the costs of openness mentioned in the paper is that of the brain drain, mostly of university graduates. This is certainly a problem for developing economies and a price paid for openness. However, two elements might modify the price paid. First, the chance that some may be able to emigrate may create an incentive for many more to try by entering domestic universities or by preparing for foreign universities' entrance examinations, in both cases resulting in more learning and a larger stock of human capital at home. Second, some of those that emigrated will return and others who stay abroad will help form networks that will benefit their home country. Even though impersonal transactions dominate in developed markets (see North 1990), nationals situated in key industries abroad provide a clear advantage for the transfer of knowledge and the facilitation of trade and business (Israel is an obvious example of this).

Comments on "Toward a Microeconomics of Growth"

Robin Burgess and Anthony J. Venables take a somewhat different road into the microeconomic foundations of growth. They examine the potential effects on growth of scale and of increasing returns originating in the innovation process and benefiting from a concentration of activities of narrowly defined industries in specific geographical locations (cities, regions). The main vehicle through which this process takes place is structural change: the movement of resources across space and of activities across industries. In both cases the move is from less productive locations and industries to ones where the innovation process is taking place, and hence productivity is higher. The high productivity of the new activities is manifested mostly in the growth of exports. The process is accompanied by shifts in the distribution of income in favor of the new industries and centers of activity.

The authors are aware that their story resembles the classical one, as developed by, for example, Schumpeter (1942); Chenery and Taylor (1968); and let me add Nobel Prize Laureate Simon Kuznets (1971), whom they fail to mention; and before even all of them, Marx and Lenin (on the development of the producer goods industry as the main vehicle of growth). The way the foregoing scholars explain growth through innovation and structural change is not really different from what Burgess and Venables offer. The new element in their paper is the focus on narrowly defined industries—that is, subsectors of the major industries rather than more broadly defined sectors—as the main carriers of innovation, and thus of growth. Nevertheless, the authors dwell, as so many before them, on the classical argument of the importance of manufacturing to technological change, and thus to growth.

The innovation in the paper is more in its declared ability to test the hypotheses on the basis of disaggregated data for subsectors than the theory itself. While concentrating on the export manifestation of narrowly defined product groups, the authors also bring their argument close to what used to be called new international

trade (see, for example, Grossman and Helpman 1992, 1994), where comparative advantage, and thus trade, was based not on endowments of principal inputs, but on acquired advantages, technological and/or historical, in specialized products. The new theory defines the trade that evolves as a result as intra-industry in contrast with the traditional inter-industry conventional comparative advantage, generated by noncompetitive mechanisms and scale effects of various kinds, also similar to the mechanisms of increasing returns discussed in this paper. I wonder to what extent intra-industry trade really becomes an important feature of trade and of growth of developing economies.[8] The paper does not, however, provide enough evidence for this.

A third element of the paper combines all the foregoing with the growing emphasis of the literature on locational aspects, that is, specific attributes of locations, such as coastline, and the tendency to concentrate particular industries in cities. All these new elements of the theoretical framework are interesting and important, and are most likely correct, but there is still a long way to go before the framework can be demonstrated empirically with microeconomic data. I say this because most of the empirical evidence cited by the paper is for entire countries or for large states within one country, like in India, where each state is larger than many independent countries, or is for the traditional major branches of the economy, that is, agriculture or manufacturing.[9] However, the paper's analysis of the micro data of Aghion and others (2003) is a first step in the micro direction. The findings of both studies are interesting, though not entirely surprising: better firms under strong state regulation also do better when these regulations are lifted. In transition economies, which suffered from over-industrialization, many state-owned enterprises produced negative or low value added and had to shut down if and when subsidies were discontinued.

The story of the rapidly expanding software industry in Bangalore (also mentioned by Srinivasan) is probably the best example of the rapid development of a subsector that really follows the logic of Burgess and Venables's model, including the concentration in one or a few centers, except for the locational aspect, which seems to be less relevant for this kind of industry.

As the authors recognize, the income distribution aspects of the model offered in the paper have early precedents. They cite Williamson (1965) on the time and development patterns of the spatial distribution of income. Inequality initially rises in favor of the centers of growth, but later it tapers off. This is, however, only one dimension along which inequality changes with development. The general pattern was first offered by Kuznets (1971) as the Kuznets curve. The main idea is exactly as shown in Burgess and Venables's paper, that is, the first group of innovators—a branch of industry, a service, a group of traders, a city, a professional group—and their immediate vicinity take off in terms of income and reap large quasi rents. Over time the innovation spreads, and so its benefits and the inequalities are mitigated. A vast literature is available on the significant increase in inequality in transition economies such as Russia and in other large countries between urban and rural areas and between regions.

In the second part of the paper, the authors concentrate on the important contribution of the decline in poverty to growth. They illustrate it in the paper by

examining the effects on poverty of the spread of micro financial services in rural India. Put together with the assumed increase in inequality accompanying the process of structural change, one has to accept that growth and the decline in poverty can take place at the same time that the level of inequality rises.

Given what was said earlier about the crucial role of improvements in the institutional environment as the main stimulant for growth, I wonder why the authors chose to list them among the group of factors included in the so-called group of first advantage. The paper divides the factors contributing to growth into two groups, first and second advantage. This division corresponds roughly to basic inputs, like labor and capital, and to those that contribute to increased productivity under the traditional growth literature. I prefer the distinction between extensive factors, such as labor and capital, and intensive ones, that is, those that contribute to sustained growth by providing externalities and increasing returns. This distinction was common in discussing growth under the communist regime. Under such a division, institutional improvements should be included under the second group.

In a global world where most inputs can move freely in great quantities across countries, not to speak of across regions in one country, institutional improvements can attract large quantities of the first advantage inputs (capital, market services, technology and know-how, and so on). In this way they can bring about some sort of increasing returns. Given the constraints on a country's or region's own resources, institutional improvements can play, at least for a while, the same role as any other dynamic growth-producing innovation or networking. Such improvements may not be able to endogenously generate sustained growth over the long run, but they can clearly serve such a role for a substantial period during the transition from traditional institutions to modern and market-oriented ones. As mentioned in the discussion of the first paper, an improvement in institutions can attract foreign banks and other financial institutions (including venture capital firms, which can revolutionize the financial and production sectors in developing and transition economies alike) and can use the new forms of firms to move some activities to developing economies.[10]

Conclusion

These are two extremely rich papers that apply a number of recent innovations in growth theory and in institutions in industrial organization theory to economic development and to developing economies. They both emphasize various elements that may be placed under the heading of a dynamic global product cycle in the service of economic development. I enjoyed reading them and learned a good deal.

Notes

1. When the transition economies of Eastern and Central Europe join the European Union, this will make even these institutions somewhat more tradable.

2. Even industrial economies reveal institutional weakness in dealing with more sophisticated financial institutions, as the recent U.S. experience (for example, with Enron) has showed.

3. One of the main causes of the stagnation is the crisis in the financial sector, which cannot adjust. In relation to the transition in financial services in transition economies, see Keren and Ofer (2003).

4. On the cost of the shift in transition economies compared with developing economies, see Ofer (2001). I also wonder if India's success in the high-tech realm following the reforms is explained by the fact that this is a new industry with relatively few vested interests to push aside.

5. Between imported new goods and exported imitated ones.

6. In a recent paper I argue that transition economies, with their more sophisticated production sectors, need modern financial services even more than developing economies, and therefore ought to be ready to pay almost any price for them. Indeed, in most East-Central European countries the bulk of the banking sector is already foreign owned, while Russia and most countries of the Commonwealth of Independent States maintain high barriers to the entry of such services (Keren and Ofer 2002; Ofer 2002).

7. There is also an ambiguous effect of the scale of production.

8. In the context of transition, the share of intra-industry trade is considered a sign of making the transition.

9. Similar analysis for Russian regions (*oblasts*), equivalent to states, can be found in, for example, Yudaeva and others (2004).

10. Though this may be more important for transition economies with their more sophisticated production sectors.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Acemoglu, D., Phillipe Aghion, and Fabrizio Zilibotti. 2003. "Distance to Frontier, Selection, and Economic Growth." Massachusetts Institute of Technology, Cambridge, MA. Processed.

Aghion, Philippe, Robin Burgess, Stephen Redding, and Fabrizio Zilibotti. 2003. "The Unequal Effects of Liberalization: Evidence from India." London School of Economics, London. Processed.

Chenery, Hollis B., and Lance J. Taylor. 1968. "Development Patterns: Among Countries and Over Time." *Review of Economics and Statistics* 50(3): 391–416.

Easterly, William. 2001. *The Elusive Quest for Growth*. Cambridge, MA: MIT Press.

Gans, Joshua, David Hsu, and Scott Stern. 2002. "When Does Start-Up Innovation Spur the Gale of Creative Destruction?" *RAND Journal of Economics* 33(4): 571–86.

Grossman, Gene M., and Elhanan Helpman. 1992. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

_____. 1994. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives* 8(1): 23–44.

Helpman, Elhanan. 1990. *Monopolistic Competition in Trade Theory*. Special Papers in International Finance no. 16. Princeton, N.J.: Department of Economics.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 2001. "Governance Matters." Policy Research Working Paper no. 2196. World Bank, Washington, D.C.

_____. 2002. "Governance Matters II: Updated Indicators for 2000/01." Policy Research Working Paper no. 2772. World Bank , Washington, D.C.

Keren, Michael, and Gur Ofer. 2002. "The Role of FDI in Trade and Financial Services in Transition: Is Transition Different Than 'Development'?" *Comparative Economic Studies* 44(1): 15–45.

_____.2003. "Globalization and the Role of Foreign Banks in Economies in Transition." In Grzegorz W. Kolodko. ed., *Emerging Market Economies: Globalization and Development*. Aldershot, U.K.: Ashgate.

Kuznets, Simon. 1971. *Economic Growth of Nations: Total Output and Production Structure*. Cambridge, Mass.: Harvard University Press.

Lerner, Josh. 1996. "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program." Working Paper Series no. 5753. National Bureau of Economic Research, Cambridge, Mass.

_____. 1998. "Angel Financing and Public Policy: An Overview." *Journal of Banking and Finance* 22: 773–83.

North, Douglas C. 1990. *Institutions, Institutional Change, and Economic Performance*. New York and Cambridge, U.K.: Cambridge University Press.

Ofer, Gur. 2001. "Development and Transition: Emerging but Merging?" *Revue d'Economie Financiere* (Special Issue): 107–46

_____. 2003. "Transition and Development: Comparing the Quality of Governments." In Nauro Campos and Jan Fidrmuc, eds., *Political Economy of Transition and Development*. Boston: Kluwer Academic Publishers.

_____. 2004. "Switching Development Strategies and the Costs of Transition: The Case of the Soviet Union and Russia." Paper presented at the American Economic Association meetings, January 2–5, San Diego, Calif.

Rajan, Raghuram G., and Luigi Zingales. 2001. "The Influence of the Financial Revolution on the Nature of Firms." *American Economic Review, Papers and Proceedings* 91(2): 206–11.

_____. 2003. *Saving Capitalism from the Capitalists*. New York: Crown Business.

Schumpeter, Joseph, A. 1942. *Capitalism, Socialism, and Democracy*. Cambridge, Mass.: Harvard University Press.

Soros, George. 2002. *On Globalization*. New York: Public Affairs.

Stiglitz, Joseph, E. 1997. "The Role of Government in Economic Development." In Michael Bruno and Boris Pleskovic, eds., *World Bank ABCDE 1990*. New York: Oxford University Press.

_____. 2002. *Globalization and Its Discontents*. New York and London: Norton.

Williamson, J. G. 1965. "Regional Inequality and the Process of National Development." *Economic Development and Cultural Change* 13(1): 3–45.

Yudaeva, Ksenia, Maria Gorban, Vladimir Popov, and Natalia Volchkova. 2004. "Down and up the Stairs: Paradoxes of Russian Economic Growth." In Gur Ofer and Richard Ponfret, eds., *The Economic Prospects of the CIS: Sources of Long Term Growth*. Cheltenham, U.K.: Edward Elgar.

Comment on "Entrepreneurship, Innovation, and Growth" by T. N. Srinivasan and "Toward a Microeconomics of Growth" by Robin Burgess and Anthony J. Venables

SIDNEY G. WINTER

Individually and in combination, these two papers present a broad range of significant comments and insights regarding the microeconomic context of economic growth. Broadly speaking, they invite us to think about some of the contextual factors that an entrepreneur or manager has to take into account when considering whether to undertake some particular business initiative, and then pursue this line of thinking further, into the systemic and long-run implications of the view the entrepreneur sees. For example, the papers emphasize the ability to appropriate gains from innovation, the conditions affecting access to finance, and the regulatory framework or business climate. All of this is welcome and certainly congenial to me personally, inasmuch as some of these firm-level issues are the same ones that I think about "for a living." (I do not really make my living thinking about economic development, and I hope you will forgive me if I stumble unannounced into one of my substantial zones of ignorance.)

I do recognize in my reactions to these papers a sense of frustration that I have experienced in previous encounters with the growth and development literature. The papers fail somewhat to get the big picture in focus. Sometimes individual contributions get pieces of the picture in sharp focus, but then one is nagged by the concern that it is often such a small piece. In other cases, the treatment is more kaleidoscopic: look, see a nice picture; shake, see another nice picture, but one quite different from the first. (In an alternative metaphor, we could say that there is a "blind men and the elephant" problem.) These two papers are more toward the kaleidoscopic end of this spectrum, which is certainly appropriate in a setting that is broadly oriented toward policy concerns. The sense of frustration, however, remains.

Why is that? What explains the persistent sense of frustration? To frame my more specific comments, I will address this big question briefly. First, there is, of course, an

Sidney G. Winter is the Deloitte & Touche Professor of Management at the University of Pennsylvania's Wharton School in Philadelphia.

inherent tension between "big picture" and "focus." How serious that tension is depends on the quality of the equipment you bring to the task (think of photography). Beyond that, however, the key point is that in trying to understand a developing economy we are dealing with a complex, highly interactive, dynamical system. It is basically a tangle of feedback loops, many of them positive feedback loops. Worse yet from the puzzled observer's standpoint, human agency is involved. This means, in particular, that whatever apparent obstacle to growth the observer may identify, participants might see it too and react constructively through individual or collective action. When this happens, the resources to undertake that remedial action likely derive from prior success, while the motivation and confidence to do so derive from an image of desired future success, which is also largely derivative of past success. In a great many ways, success breeds success and thus explains itself, and so does failure. This is not an illusion or a veil to be penetrated; this is the way it works. More precisely, it is an apt characterization at an aggregate level of the behavior of a big system of simultaneous dynamical equations. There is no a priori reason for confidence that an observer who can see only a limited number of aggregate variables can understand the causal structure of such systems, especially if experimental intervention is largely barred as an instrument of inquiry. Not only can symptoms be confused with causes, symptoms feed back and function as causes themselves, often producing mysterious symptoms remote from themselves as well as the proximate ones that bear their characteristic signature.

Another notable feature of the development process that partly explains the positive feedback loops is a fairly strong complementarity among contributing factors. If we contemplate a list of success factors, such as a healthy and educated workforce, secure property rights, strong infrastructure, means of enforcing agreements, access to capital, a liberal trade regime, and limited and well-administered regulation, few would support the claim that "any one of these will do," and most would agree that the truth is better approximated by "all are required." But that is not really the truth. The complementarity is not as strong as that, because, as noted previously, some obstacles can be attacked and removed, or if not, they can be circumvented or compensated for. All those key factors are matters of degree in any case. My sense is that more success has been achieved in identifying factors and proving that they matter in a statistical sense than at characterizing the complementarity and mapping the limited tradeoffs that exist. The latter would probably be more helpful for policy guidance than establishing that some shortlist of factors captures a lot of variance in a cross-national growth rate regression.

In light of these considerations, the increasing emphasis on microeconomic mechanisms and institutional contexts reflected in these papers is extremely welcome. Such inquiry has more promise of cutting through the feedback loops and revealing causal mechanisms than cross-sectional comparisons at more aggregate levels can ever have. Of course, an even more powerful tool is deliberate, randomized experimentation, as noted by Burgess and Venables and featured elsewhere on the conference program, but that is a tool of narrow scope relative to the development problem as a whole.

Turning away from these general issues, I want to focus on some issues concerning the role of knowledge in economic development—the sort of knowledge that always underlies productive activity and the advance of which is crucial in innovation.

Influenced significantly by a classic paper by Arrow (1962), for a long time economists have tended to explore a particular view of knowledge that emphasizes the ease with which it leaks out, the difficulty of sustaining a property right in it, and the resulting limits on incentives to create it in the first place. A common subtext, occasionally explicit, is that the intellectual property rights problem is serious because there are thieves out there, actors who would endeavor to seize the knowledge so as to enrich themselves. Less attributable to Arrow, but also common in the economic literature, is the tendency to think of knowledge and innovation in terms of "ideas"; economists are evidently familiar with the value of good ideas (see, for example, Romer (1993). These themes are visible in the endogenous growth model that Srinivasan, drawing on Helpman (Grossman and Helpman 1992, 1994; Helpman 1990), expounds in the early part of his paper.

All the elements of this familiar view of knowledge are found in economic reality. They are important in the economics of various types of copyrighted works, in software, and in pharmaceuticals, for example. What is much less clear is whether these particular knowledge issues are anywhere near as typical or fundamental in economic reality as many economists seem to believe. One consequence of this exaggerated emphasis on expropriation is the tendency for economists to offer unhesitating support for proposals to strengthen intellectual property rights. As Srinivasan observes: "Unfortunately, without any serious analysis of their global social value, IPR [intellectual property rights] regimes have been strengthened through the agreement in Trade Related Aspects of Intellectual Property Services and enshrined in the World Trade Organization."

I have a further argument to support the general point that the economists' familiar view of knowledge, emphasizing expropriation hazards, misses important aspects of the role of knowledge in development. The first thing to understand about the organizational knowledge domain is its enormous diversity. There are indeed secrets that are hard to protect, and for some, to both protect and exploit is impossible. However, there are also secrets that are, for various reasons, hard to convey. While knowledge thieves are a genuine problem in some contexts, closed minds, sometimes evincing a smug commitment to willful ignorance, are a vast problem in others. Here we can reference the fact that Nobel Laureate George Akerlof had difficulty in getting his famous "lemons" paper published (Akerlof 1970). What a wonderful irony that there was difficulty in giving away valuable ideas on the economics of information, even though the overwhelming expressed concern of economists addressing the subject has been with the difficulty of protecting valuable ideas, not the difficulty of giving them away.

In short, a spectrum of knowledge types exists, ranging from the difficult-to-protect to the difficult-to-convey. Economic activities of genuine significance lie near both ends of the spectrum as well as in between (see Winter 1987 for more details about the nature of this spectrum).

Let me bring this point into the venture capital context, which Srinivasan addresses. As he observes, the transactional context between entrepreneur and venture capitalist is fraught with potential hazards. The entrepreneur may behave opportunistically and devote the funding to ends the venture capitalist does not favor, which the venture capitalist can try to resist by means of monitoring. When intellectual property protection is of dubious reliability (which is typical), the entrepreneur faces a big hazard of expropriation by the venture capitalist, who can place the valuable idea in the hands of someone who will exploit it more advantageously (at least in relation to the venture capitalist). Alternatively, however, the entrepreneur may have difficulty in locating a venture capitalist who appreciates the proposal, for much the same reason that Akerlof had difficulty in placing his article. People resist new ideas because they are unfamiliar, and entrepreneurs obviously cannot clinch the argument by pointing to a developed example of the very idea they need resources to develop. I suspect that such communication difficulties often underlie the apparently opportunistic actions by entrepreneurs, who may simply be pursuing the optimal development path as they conceive it. Such actions may actually be in the venture capitalist's interest, but the entrepreneur cannot convince the venture capitalist that this is the case.

Even though sufficiently innovative and unfamiliar ideas may be hard to convey, ideas generally lie toward the end of the spectrum where the characteristics and hazards Arrow (1962) originally highlighted tend to prevail. Ideas, however, do not create goods and services any more than they create publishable papers. Ideas are merely the starting point for what is typically an extended effort to bridge the gap between conception and achievement. (As Edison put it: "Genius is 1 percent inspiration, 99 percent perspiration.") What does create goods and services, aside from the inputs acknowledged in the production function, is command and exercise of the appropriate skills and routines.[1] The point here is that skills and routines typically lie at the hard-to-convey end of the spectrum of knowledge types, especially when they are highly complex and partly tacit, as they often are. Managers struggle to spread desirable practices through their organizations for reasons intrinsic in the character of organizational knowledge, as well as for incentive reasons (Szulanski 1996).

A key conceptual point needs to be made here. Knowledge that is difficult and/or costly to reproduce or transfer still has the fundamental increasing returns character that economists have recognized in relation to information and knowledge: it is nonrivalrous in use. Grasping that central commonality is one part of getting the subject right. The other part is recognizing that situations cover an enormous range with respect to difficulty or cost of transfer. Unfortunately, many economists have tended to package the nonrivalry aspect of information with the low-cost aspect of transfer, and treat the two as if some intrinsic linkage existed between them. This is not the case. It is indeed true that there is one cost that nonrivalry rules out, namely, the opportunity cost of the knowledge as such is zero, but there are plenty of other sources of transfer costs.[2]

This brings me to two observations that apply to both papers, but perhaps especially to Burgess and Venables. The first observation is that the skills and routines

aspect of the development problem is largely missing, though both papers mention learning effects and Burgess and Venables cite Sutton (2002) in relation to team-embodied knowledge. This neglect is, of course, quite typical in the literature. The second observation is about another neglect: the term "multinational corporation" does not appear in these papers [as they were presented at the conference]. This one is more surprising to me. The two observations have an important connection. The activities of multinational corporations are an important mechanism by which the skills and routines needed in the contemporary global economy are transferred to businesses, individuals, and groups in the developing and emerging economies. In particular, the operation of tightly coordinated global supply chains in an increasing number of industries has that effect. Much of that transfer is deliberate. It is costly, but the cost is borne voluntarily because it is a necessary part of an efficient approach to accessing the resources that the developing world has to offer, particularly low-priced labor. In other cases it is non-deliberate. The transfer costs are borne in the form of substandard performance while learning takes place in an appropriate facilitating setting provided by the multinational corporation,[3] and the fact that a quasi-permanent asset transfer is taking place is not focal for the multinational corporation.

Thus the source firm's concern in these processes is not with protecting valuable knowledge, but with overcoming the difficulties of exploiting it on a large scale. To the extent that it seeks to recapture the costs of original creation, the large-scale exploitation by the creator is what provides the main answer, and the difficulty of making the transfer is what limits the losses from exploitation by others. Note that while transfer difficulty protects the source from direct competition by imitators (and thereby helps to solve the incentive problem otherwise impeding the transfer), it does not preclude the use of the transferred skills and routines in other settings in combination with other skills and routines that have undergone a similar process of costly transfer. Or to put it another way, the individual firm in the advanced country is not creating a competitor for itself and would not transfer knowledge in this fashion if it were. Collectively, however, the advanced country firms are definitely creating competitors for themselves by liberally seeding the developing world with a multitude of specific skills and routines developed in the advanced countries and proven in the global marketplace. That, of course, is a good thing, assuming that we want the developing economies to grow.

Let me wind up these comments by mentioning one other problem that does not, in my, opinion, get due attention in these papers. The firm, city, or region that is trying to prosper by joining the world economy has to figure out what specifically it is going to try to offer on the world market. It has to make or facilitate a number of specialized investments to make that participation possible. It has to be sensitive to current and prospective prices, particularly world market prices for exported outputs and required imported inputs, which together define the value added gap that may be exploitable. It has to do that while confronting a global context that, in many sectors, changes continuously and rapidly, and in which many thousands of other actors are presumably attempting similar calculations and initiatives. It must also stand ready to do the whole thing over again in at most a few years, when the activities of rivals and other change processes will likely have reshaped the competitive landscape.

For a firm or region to succeed in such an effort over a period of time is a considerably more impressive achievement than might be suggested by models featuring, for example, reliable mechanisms for increasing returns at the regional level. Note that the game can involve failures and temporary successes as well as sustained successes, and that even sustained successes generally give way eventually to waves of firm failure and depressed regional economies. The lumpiness that Burgess and Venables describe is not manifested just in the domain of growth; it is also manifested in stagnation and decline. While increasing returns are a fundamental consideration, the specificity of activities is also fundamental and often tends to pull in the opposite direction. Reflecting the importance of specificity would be a valuable enrichment of the formal model that Burgess and Venables have offered, permitting it to give a more complete account of the relevant phenomena.

The kind of entrepreneurship required in such settings is an elaborated version of the kind discussed in the Austrian tradition (Kirzner 1997). It is a matter of searching for so-called arbitrage opportunities, albeit in a sense of that term so stretched as to make it almost unrecognizable. The virtue of the Austrian formulation is its emphasis on the broader price-search aspects of the problem of achieving business success. Consistent with the general orientation of this session, I suggest that considerations that are so plainly important at the firm level must also deserve some attention for the larger-scale purposes of understanding economic development. Treatments of growth grounded in the production function framework certainly do not throw much light on this. While the skills and routines perspective has something relevant to offer, it too is not up to the full task. This is an area where, I believe, we need a great deal of progress.

Of course, we can always say that the market will take care of it, as in a sense it will, but the market that will do the work is the real market, not the mythical Walrasian auctioneer, nor any of the other highly abstract versions of markets that economists typically consider. As far as I know, none of those theoretical markets can quote the price of a nonexistent product, yet that is the kind of information the entrepreneur is often looking for (see Denrell, Fang, and Winter 2003). We have much to learn about how this works. Indeed, the two papers considered are an important contribution to building the more general case that the microeconomics of growth is a growth sector with great promise.

Notes

1. As Nelson and I argue (Nelson and Winter 1982), organizational routines may be viewed as the multiperson counterpart of individual skills. In management discourse, the term "practices" covers the ground of our "routines" somewhat imperfectly, but still usefully.

2. Because of the tendency to ignore them, Szulanski and I label the high transfer cost cases as "nonstandard examples" of information economics. These are discussed further in Winter and Szulanski (2002).

3. The situated cognition view in cognitive psychology emphasizes that the knowledge people use to perform concrete tasks is typically highly dependent on aspects of the specific context of performance, including artifacts and social relations (see, for example, Hutchins 1995).

References

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84(3): 488–500.

Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Invention." In R. Nelson, ed., *The Rate and Direction of Inventive Activity*. Princeton, N.J.: Princeton University Press.

Denrell, Jerker, Christina Fang, and Sidney G. Winter. 2003. "The Economics of Strategic Opportunity." *Strategic Management Journal* 24(10): 977–90.

Grossman, Gene M., and Elhanan Helpman. 1992. *Innovation and Growth in the Global Economy.* Cambridge, Mass.: MIT Press.

_____. 1994. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives* 8(1): 23–44.

Helpman, Elhanan. 1990. *Monopolistic Competition in Trade Theory.* Special Papers in International Finance no. 16. Princeton, N.J.: Department of Economics.

Hutchins, Edwin. 1995. *Cognition in the Wild.* Cambridge, Mass.: MIT Press.

Kirzner, Israel M. 1997. "Entrepreneurial Discovery and the Competitive Market Process: An Austrian Approach." *Journal of Economic Literature* 35(1): 60–85.

Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change.* Cambridge, Mass.: Harvard University Press.

Romer, Paul. 1993. "Idea Gaps and Object Gaps in Economic Development." *Journal of Monetary Economics* 32(3): 543-73.

Sutton, John. 2002. "Rich Trades, Scarce Capabilities: Industrial Development Reconsidered." *Economic and Social Review,* 33(1): 1–22.

Szulanski, Gabriel. 1996. "Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within the Firm." *Strategic Management Journal* 17 (Winter Special Issue): 27–43.

Winter, Sidney G. 1987. "Knowledge and Competence as Strategic Assets." In D. J. Teece, eds., *The Competitive Challenge: Strategies for Industrial Innovation and Renewal.* Cambridge, Mass.: Ballinger.

Winter, Sidney G., and Gabriel Szulanski. 2002. "Replication of Organizational Routines: Conceptualizing the Exploitation of Knowledge Assets." In C. W. Choo and N. Bontis, eds., *The Strategic Management of Intellectual Capital and Organizational Knowledge.* New York: Oxford University Press.

Challenges of Development in Lagging Regions

World Poverty: Causes and Pathways

PARTHA DASGUPTA

In this paper I offer an account of extreme poverty in the world's poorest regions by studying the interplay between poverty, reproduction, and the state of the local environmental resource base. The persistence of poverty in a world that has elsewhere been enjoying economic progress is traced to socioeconomic and ecological mechanisms involving positive feedback.

Introduction

Even though there probably are only a few pathways to economic prosperity, the number of routes societies can take to experience stagnation—even decay—are many. I have been asked to talk at this conference about those countries that have been the laggards in the race toward contemporary economic development. I want to do this not only by identifying the various senses in which certain regions of the world have at best remained where they were decades ago, but also by trying to understand how they have managed to do so.

Outline of a View of Poverty Traps

Only a few years ago, a paper with my intentions would have identified weaknesses in public policy—including the choice of wrong investment projects—as the cause of economic failure. Today the temptation would be to point to institutional weaknesses. One can even see such a change in the way the World Bank's annual *World Development Report* has evolved since its inception in 1978. But even when existing

---

institutions are progressive, good policies and sound investment projects cannot be plucked from thin air. Institutions, policies, and investments are so dependent on each other that if you want to probe one, you must simply keep an eye on the others.

With these dependencies in mind, growth theorists in the late 1980s identified resource allocation mechanisms harboring the kind of positive feedback that brings about prosperity. The models encouraged growth experts in the early 1990s to seek evidence of convergence in the economic performances of nations. The believable results have been negative (Pritchett 1997). One aim of growth experts now is to explain why only a few of the countries that were considered "underdeveloped" in the early 1950s have experienced economic progress.

For more than 20 years I have been studying resource allocation mechanisms involving a different kind of positive feedback (Dasgupta 1982, 1993, 1996, 1997, 1998a,b, 2000, 2001a, 2003; Dasgupta and Mäler 1991; Dasgupta and Ray 1986, 1987). In contrast to the ones studied by growth theorists, the mechanisms I have been studying permit hunger and poverty to be a persistent experience for large groups of people in poor regions even while others there and elsewhere are able to prosper. The mechanisms harbor poverty traps, and they operate at a disaggregated level. One category, involving metabolic pathways, works at the level of the individual person. They are based on physiological links connecting nutritional status and work capacity among adults and nutritional status and physical and mental development among children (Dasgupta 1993, 1997; Dasgupta and Ray 1986).[1]

A second category, operating at a spatially localized level, is site specific. It involves a combination of ecological and socioeconomic pathways, sustaining reproductive and environmental externalities. In contrast to the mechanisms underlying modern growth models, these mechanisms are influenced by the local ecology. The theory based on such mechanisms acknowledges that the economic options open to a poor community in, say, the African savannahs, are different from those available to people in the Gangetic plain of India. To put it simply, policies matter, as do institutions, but the local ecology matters too.

Among the ecological and socioeconomic pathways I have been studying are some that reflect synergies between undernourishment and a person's vulnerability to infectious diseases (Dasgupta 1993),[2] but other pathways have a more extensive reach. They involve positive feedback between poverty, population growth, and degradation of the local natural resource base. In the latter mechanisms, however, neither poverty, nor population growth, nor environmental degradation is the prior cause of the others: over time each influences, and is in turn influenced by, the others. The two broad categories of positive feedback mechanisms are able to coexist in a society, because nutritional status does not have much of an effect on fecundity except under conditions of extreme nutritional stress (Bongaarts 1980).[3]

In speaking of an economy, I shall cast a wide net here. The economy could be a household, or it could be a village, a district, a province, a nation, or even the whole world. Note, however, that a village could be in the grip of a poverty trap even if the country is not. In such a situation, aid from outside the village would be needed if the villagers were to lift themselves out of the mire. By contrast, to say that a country

is in the grip of a poverty trap is only to say that without external assistance, it would not be possible for all groups in the country to lift themselves out of poverty, at least not in the foreseeable future. It is not to say that there are no rich people in the country.

Those who are caught in poverty traps do not necessarily spiral down farther. For most there is little room below to fall into: many are already undernourished and susceptible to diseases. Modern nutrition science has shown that relatively low mortality rates can coexist alongside a high incidence of undernutrition and morbidity. To be sure, many die because of causes traceable directly to their poverty, but large numbers continue to live under nutritional and environmental stress. Moreover, people tend not to accept adverse circumstances lying down, so it is reasonable to assume that they try their best to improve their own lot.

The theory I outline later identifies conditions under which the coping mechanisms people adopt are not enough to lift them out of the mire. For example, Turner and Ali (1996) show that in the face of population pressure in Bangladesh, small landholders have periodically adopted new ways of doing things so as to intensify agricultural production. The authors demonstrate, however, that this has resulted in only an imperceptible improvement in the standard of living and a worsening of the ownership of land, the latter probably because of the prevalence of distress sales of land. These are the kinds of findings that the perspective I have been exploring anticipated and was designed to meet.[4]

The externalities associated with people's coping strategies can amount to significant differences between private and social returns to various economic activities. Thus, where reproductive behavior is "pro-natalist," the private returns to having large numbers of children are high, in contrast with the social returns. Similarly, where communities degrade their natural resource base, the collective endeavors to maintain the base are unable to withstand the pressure of private malfeasance.

Grumbles

The presence of synergies between undernutrition and infection is widely accepted today. Development economists have also learned to accept the nonlinear relationships between nutritional status and the capacity to work. Yet econometric models incorporating human health and labor productivity are often linear. There is a dissonance here, one that could lead policymakers astray when they assess the finances required to eliminate absolute poverty.[5]

That ecological processes are usually nonlinear is even less appreciated. For example, economists commonly think that, to quote an editorial in the United Kingdom's *The Independent* (December 4, 1999): "[Economic] growth is good for the environment because countries need to put poverty behind them in order to care." Similarly *The Economist* notes (December 4, 1999, p. 17): "[T]rade improves the environment, because it raises incomes, and the richer people are, the more willing they are to devote resources to cleaning up their living space."

The view's origin can be traced to the World Bank (1992), which observed an empirical relationship between gross national product (GNP) per head and atmospheric concentrations of industrial pollutants. Based on the historical experience of countries of the Organisation for Economic Co-operation and Development, the report's authors suggest that when GNP per head is low, concentrations of such pollutants as sulfur oxides increase as GNP per head increases, but when GNP per head is high, concentrations decrease as GNP per head increases further (see also Cropper and Griffiths 1994; Grossman and Krueger 1995). Among economists this relationship has been christened the environmental Kuznets curve.[6] The moral that would appear to have been drawn from the finding is that resource degradation is reversible: degrade all you want now; Earth can be relied upon to rejuvenate it later if you require it.

This presumption is false. The underlying basis of economic activity is the possibility of transforming goods and services into further goods and services, both in time and across time. The problem is that the transformation opportunities permitted by ecological processes form nonconvex sets.[7] Nonconvexities are often a reflection of thresholds, but the existence of ecological thresholds implies that major damage to an ecosystem is, to all intents and purposes, irreversible. The environmental Kuznets curve was detected for mobile pollutants. Mobility means that, so long as emissions decline, the stock at the site of the emissions declines. But reversal is the last thing that would spring to mind should a grassland flip to become covered by shrubs, should the Atlantic gulf stream come to a halt, should a source of water disappear, or should a fishery become a dead zone. As a metaphor for the possibilities of substituting manufactured and human capital for natural capital, the relationship embodied in the environmental Kuznets curve has to be rejected.[8]

Although nonconvexities are prevalent in global ecosystems (for example, ocean circulation, global climate), it is good to emphasize the spatial (and, as we have seen, even personal) character of many of the positive feedback mechanisms associated with nonconvex processes. Aggregation at the regional or national level can therefore mislead hugely (see the interchange between Johnson 2001 and Dasgupta 2001b). Cross-country regressions, on which much contemporary understanding of the pathways to economic progress and stagnation is based, can also mislead. Societal features that could be important, but have not yet found expression in quantitative form on a national basis, are overlooked in them. For example, in an early study (Dasgupta 1990), economic failure among the poorest countries was viewed in terms of restrictions in political and civil liberties.

More recently, Easterly and Levine (1997) sought to explain Sub-Saharan Africa's appalling economic performance in terms of ethnic diversity, quantitative measures of which are available at the country level. However, one may ask why ethnic allegiance should have played such a role in Africa as the authors discovered, or why political and civil liberties have been so scarce there. One may ask the more basic question why the public sphere of life in Sub-Saharan Africa continues to elicit so little trust and commitment. The approach I take in this paper leads me to offer an account not altogether at odds with Easterly and Levine, but it provides something of an explanation

for the way ethnic membership is able to play a damaging, coordinating role in Sub-Saharan Africa. Nor is the theory I develop here at odds with Dasgupta (1990), but it has the virtue of explaining why certain other socioeconomic variables, such as the fertility rate, are so different in Africa than in other poor regions.

The analytical basis of the resource allocation mechanisms I discuss here are not dissimilar to the ones Collier and Gunning (1999) study in their survey paper on Sub-Saharan Africa, but the evidence I collate here is based on a wider set of considerations, and thus the mechanisms I track differ from the ones Collier and Gunning report. The pathways I study give prominence to population growth and natural resource degradation in the world's poorest regions, matters that go mostly unnoted by Collier and Gunning. Nor do Collier and Gunning place any importance on the nonconvexities inherent in the metabolic processes that govern an individual's ability to function.

Easterly (2002) and Sen (1999) focus on different measures of human well-being for judging economic success (GNP and freedoms, respectively), but they both show a lack of interest in what makes for sustainable well-being. Neither author takes nature's role in our lives seriously, nor does either offer any analysis of the causes and consequences of rapid population growth in the poorest regions of the world.[9]

Contemporary models of economic growth are equally dismissive of the importance of nature. In their extreme form, growth models contain an assumed positive link between the creation of ideas (technological progress) and population growth in a world where the natural resource base comprises a fixed, indestructible factor of production (Kremer 1993 develops such a model to account for 1 million years of world economic history). The problem with the latter assumption is that it is wrong: the natural environment consists of degradable resources (agricultural soil, watersheds, fisheries, and sources of freshwater, and ecological services more generally). It may be sensible to make that wrong assumption for studying a period when natural resource constraints did not bite, but it is not sensible when studying development possibilities open to today's poor regions. The latter move is especially suspect when no grounds are offered for supposing that technological progress can be depended upon indefinitely to more than substitute for an ever increasing loss of the natural resource base.

In any event, we should be skeptical of a theory that places such an enormous burden on an experience not much more than 200 years old. Extrapolation into the past is a sobering exercise: over the long haul of history (a 5,000-year stretch, say, until about 200 years ago), economic growth even in the currently rich countries was for most of the time not much above zero (see Fogel 1994, 1999; Johnson 2000; and especially Maddison 2001).[10]

The decision concerning whether or not to pay heed to nature in economic analysis cannot now be left to the discretion of development economists and growth analysts. An enormous body of empirical work in anthropology and ecology not only emphasizes the role natural capital plays in our lives, but also points to the interplay of reproductive behavior and the way communities manage local resources. Moreover, a substantial literature on environmental and resource economics has exposed the

inability of commonplace institutions to price natural resources in ways that reflect their scarcity values. Furthermore, agricultural scientists have drawn attention to the fact that future prospects of food being available to the world's poorest inhabitants depend critically on our ability to manage human numbers and natural capital (International Food Policy Research Institute 1995). The theory of poverty traps I discuss here is built on these findings.

Plan of the Paper

The plan of the paper is as follows: The first part offers macro-level evidence on contemporary poverty. It begins by collating a few well-known regional statistics on poverty, measured in terms of income, but income is a flow and, as customarily measured, is unable to reflect long-run possibilities. At a conference on economic development, we should be interested in the long run. Attention is then drawn to the weaknesses of indicators of short-run well-being, such as income and GNP per head, and argues that movements in an all-inclusive index of wealth should be used instead to determine whether long-run well-being is sustainable. The argument is based on Pearce and Atkinson (1995); Dasgupta and Mäler (2000); Dasgupta (2001a); and Arrow, Dasgupta, and Mäler (2003a,b), who prove the result at progressively greater levels of generality. By manipulating rough estimates of changes in wealth in a number of countries, I then put the theory to work and show that over the past quarter of a century or so, average long-run well-being has declined in the two poorest regions of the world, namely, the Indian subcontinent and Sub-Saharan Africa. The estimates on which my calculations are based are extremely crude, but there is nothing I can do about that: applied development economics has not kept pace with the part of economic theory that takes the environment and sustainable development seriously. Nevertheless, the findings, such as they are, imply that the poorest countries have de-cumulated their wealth relative to their population sizes.

In the second part of the paper I offer a microeconomic theory that goes some way toward explaining the macro-level statistics of the first part. The account I offer is qualitative and is based on the presence of positive feedback between (a) nutritional status and the capacity to work, and (b) poverty, population growth, and degradation of the local natural resource base. The latter types of feedback involve reproductive and environmental externalities, but environmental externalities amount to property rights failures over natural resources. In this connection, I distinguish between two broad categories of resources: (a) micro-organisms, such as pathogens, and (b) macro resources, such as threshing grounds, sources of water, agricultural soil, woodlands, and forests.

The paper identifies four types of reproductive and environmental externalities. Three ensure that the private returns to pro-natalist behavior are higher than the social returns. They also ensure that private returns to resource conservation are lower than the social returns. The effect of the fourth (because of conformist behavior) can go either way, but the evidence suggests that it has been pro-natalist in the Indian subcontinent and Sub-Saharan Africa. To put it simply, the analysis points to

the presence of significant economic distortions, implying that the inhabitants of the poorest countries have also suffered from too little consumption. Taken together, the analysis implies that, over the past few decades, people in the poorest countries both consumed and invested too little.

Economists have often responded to the concerns people express about population growth and resource degradation by pointing to contemporary history's winners (see, for example, Johnson 2001). After all, as they rightly observe, world output of grain has more than kept pace with population growth and people on average live longer today than they did in the past, eat better, are better educated, and (except in Sub-Saharan Africa) earn more. However, village-level studies in the poorest regions of the world, being more discriminatory, have frequently revealed something else also: they have uncovered enormous additional hardship that people have experienced during the process of recent economic changes. One purpose of the paper is to show how a revision of national accounts can enable macro-level statistics to better reflect such micro-level facts. It seems to me that a better connection between the two is necessary if we are to identify policies that can be expected to generate economic progress rather than economic growth. The theoretical framework developed in the second part of the paper is able to point to contemporary history's many losers.

I make no attempt to forecast the future, nor do I try to review how societies that are currently affluent grew in population even while accumulating wealth by substituting knowledge, skills, and manufactured capital for natural resources.[11] My aim here is to use economic theory and the recent historical experience in poor regions to suggest a way of thinking about economic stagnation in the contemporary world. I do not suggest that the experience I summarize here had anything inevitable about it. Public choices could have been made (for example, establishing safety nets, providing education and health care services, improving the collective management of the local natural resource base, and removing absurd production and trade restrictions) that would have resulted in superior collective outcomes. Implicitly though, I argue that such choices were ignored in part because of faulty economic analysis.

Formally speaking, systems that are characterized by positive feedback often possess multiple basins of attraction. Suppose that an economy initially faces good prospects, which is to say that it sits in a good basin of attraction. Suppose, however, that over time, ill-advised policies and deplorable institutions push the economy into a bad basin of attraction, one involving, say, a poverty trap. Getting the economy out of the trap may now require external help even if the institutions were to improve and the economic policies now chosen were sound.

Other than drawing a few broad recommendations in the concluding section, I shall resist discussing policy. Some of the policies that commend themselves emerge directly from the analysis presented here: they need no special pleading. In any case, the empirical evidence I put together to illustrate the theory is tentative, and I am loath to draw firm conclusions from it. The one point on which I am not at all tentative, though, is that development economists have neglected certain key features of the environments in which people in rural parts of poor countries make decisions. This paper is largely an attempt to redress the balance.

Macro-Level Evidence

This section considers the general background pertinent to a review of the macro-level evidence and the connection between wealth and sustainable—as opposed to short-run—well-being.

Aggregate Background

Visitors to the Indian subcontinent routinely observe emaciated beggars on the streets of large cities. They are the economically disenfranchised. However, for the average person in India, village life is even worse than urban life. Table 1 summarizes evidence on poverty and population growth. Poverty (sometimes referred to as extreme poverty) is taken to be the condition of a person living on less than US$1 a day. There were 1.2 billion poor people in the world at the turn of the century, thus the poor account for about one-fifth of the world's population. They are concentrated in China, South Asia, and Sub-Saharan Africa, but there are differences in the incidence of poverty even among those three regions: proportionately, South Asia and Sub-Saharan Africa are home to the largest numbers of poor people. It would appear that the prevailing social institutions in China offer safety nets to the most vulnerable, keeping them from destitution (see Jalan and Ravallion 2003).[12] Other than a few observations I make later, I ignore China and instead explore the conditions of persistent poverty in rural South Asia and Sub-Saharan Africa only.

The International Food Policy Research Institute (1995) estimates that 800 million people (of whom approximately 530 million were in South and East Asia and

TABLE 1.
Poverty and Population Growth, by Region, 1998

Region	Number of people living on less than US$1/day, 1998 (millions)	Headcount index[a]
East Asia and the Pacific (excluding China)	65	11
China	213	18
Europe and Central Asia	24	5
Latin America and the Caribbean	78	16
South Asia (Bangladesh, India, Pakistan)	522 (495)	40
Sub-Saharan Africa	290 (242)	46
Total	1,192 (1,270)	24

Note: Figures in parentheses are the number of people living on less than US$1 a day in 1990.

a. The headcount index is the proportion of people that are poor.

Source: World Bank (2000a, table 2.1; 2000b, table 1.1).

170 million were in Sub-Saharan Africa) suffer from food insecurity. Of course, the idea of a poverty line, whether it is based on income or on food insecurity, can be criticized, but the practical advantages of thinking in terms of a line that divides the poor from those who are not poor are considerable, so the concept is used widely.

Deficiencies in micronutrients are pervasive in the poor world. Some 1.2 billion people (and more than half the number of pregnant women in poor countries) suffer from anemia, 600 million suffer from iodine deficiency disorders, and 125 million preschool children suffer from vitamin A deficiency. At the same time, more than 2 billion people in poor countries have no access to sanitation facilities, and some 1.3 billion people lack access to potable water.

Eradicating micronutrient deficiencies would not demand many resources. Rough calculations indicate that less than 0.3 percent of world income is all that would be required on an annual basis. A problem of far greater magnitude is the availability of dietary energy. The general consensus among nutritionists is that, barring diets that build on root and tuber crops, those containing adequate energy are also adequate in their protein content. Among the world's poor, cereals (namely, wheat, rice, maize, and barley) are the main sources of nutrition, accounting for more than 50 percent of their energy intake. So when people worry about food prospects in, say, the year 2020 or 2050, they typically worry about the availability of cereals.[13]

The poor live in unhealthy surroundings, a fact that is both a cause and an effect of their poverty. Nearly 2 million women and children die annually in poor countries from exposure to indoor pollution (cooking can be a lethal activity among the poor). More than 70 percent of freshwater sources are contaminated or degraded. Moreover, groundwater withdrawal in poor countries exceeds natural recharge rates by a phenomenal 160 billion cubic meters per year. The World Bank (2001) suggests that 5 million to 12 million hectares of land are lost annually to severe degradation and that soil degradation affects 65 percent of African croplands and 40 percent of Asian croplands, in part because of nitrogen and phosphorus losses.

The poorest countries are in great measure agriculture-based, subsistence economies. The agricultural labor force as a proportion of the total labor force is some 60 to 70 percent. The share of agricultural value added in GNP is approximately 25 to 30 percent.

The connection between rural poverty and the state of the local natural resource base should be self-evident. When wetlands, inland and coastal fisheries, woodlands, ponds and lakes, and grazing fields are damaged, say because of agricultural encroachment, urban extensions, the construction of large dams, or collective failure at the village level, traditional dwellers suffer. For them—and they are among the poorest in society—there are frequently no alternative sources of livelihood. In contrast, for rich ecotourists or importers of primary products there is something else, often somewhere else, which means that alternatives are available. So whether or not there are substitutes for a particular resource is not entirely a technological matter, nor a mere matter of consumer taste. Often the poor cannot move and are thus caught in a trap. But even if they were to migrate, it could be that they are unable to find employment. The poorest of the poor experience a lack of substitution possibilities

in ways the rich do not. Similarly, they experience nonconvexities in a way the rich do not. Even the range between a need and a luxury is enormous and context-ridden. A pond in one village is a different asset from a pond in another village, not only because the ponds' ecological characteristics are likely to be different, but also because the communities making use of them are likely to face different economic circumstances. Macroeconomic reasoning glosses over the heterogeneity of the Earth's resources and the diverse uses to which they are put, both by people residing at the site and by those elsewhere. National income accounts reflect that reasoning by failing to record a wide array of our transactions with nature.

The reason why changes in the size and composition of natural capital are in large measure missing from national accounts is that nature's services most often do not come with a price tag. The reason for that is that property rights to natural capital are often difficult to establish, let alone enforce, and the reason for that is that natural capital is frequently mobile. At the broadest level, soil, water, and the atmosphere (which are capital assets themselves) are media in which capital assets connect among themselves and flourish, meaning that a disturbance to any one asset can be expected to reverberate on many others at distances away, sometimes at far distances. Under current practice though, the consequences of the connectedness of natural capital can easily go unnoted in economic transactions. Thus those who destroy mangroves to create shrimp farms, or who cut down forests in the uplands of watersheds to export timber, may not be required to compensate fishermen dependent on the mangroves or farmers and fishermen in the lowlands whose fields and fisheries are protected by the upland forests. Economic development in the guise of growth in per capita GNP can come in tandem with a decline in the wealth of some of the poorest members of society. Moreover, being very heterogeneous, aggregate indexes of natural resources are hard to find.

That the populations of both South Asia and Sub-Saharan Africa have grown in excess of 2 percent per year for several decades is well known. Table 2 offers a picture of population growth in terms of crude birth and death rates. The table shows that increases in population size were due to a decline in mortality rates (a remarkably good thing), which were unmatched by reductions in fertility rates. Population increase in the poorest regions brought in its wake additional pressure on local resource bases (a not so good thing). The overall phenomenon requires explanation. The models discussed later offer a language in which to make an inquiry.

Wealth and Sustainable Well-Being

In this following section I show why we need to broaden the study of economic performance from a near-exclusive attention on short-run measures of well-being, such as GNP per head and the United Nations' human development index (HDI), by including in our assessments movements in an inclusive measure of wealth.

Why Wealth?

The United Nations Development Programme (1994) castigates those who regard GNP to be an index of an economy's well-being on the ground that it is a measure of

TABLE 2.
Crude Birth and Death Rates Per 1,000 People, Selected Countries and Regions, 1980 and 1996

Region or country	Birth rate		Death rate		Birth rate – death rate	
	1980	1996	1980	1996	1980	1996
China	18	17	6	7	12	10
Bangladesh	44	28	18	10	26	18
India	35	25	13	9	22	16
Pakistan	47	37	15	8	32	29
Sub-Saharan Africa	47	41	18	14	29	27
Nigeria	50	41	18	13	32	28
World	27	22	10	9	17	13

Source: World Bank (1998, table 2.2).

a country's opulence. The criticism is faulty in two ways. First, opulence is a stock concept, and GNP is not a return on any index of opulence that I am aware of.[14] Second, and more important, the connection welfare economics has drawn between what one may call the constituents and determinants of well-being (Dasgupta 1993, 2001a) tells us that it is not a mistake to seek to measure a society's well-being in terms of an index of opulence. The point is not that opulence misleads, but rather that we should search for the right measure of opulence.

An economy's prospects are shaped by its institutions and by the size and distribution of its capital assets. Taken together they are its productive base. However, institutions are different from capital assets in that the former comprise a social infrastructure for guiding the allocation of resources (for instance, laws and property rights), among which are the capital assets themselves.

We have a name for the overall worth of an economy's capital assets: wealth. Even though economic statisticians have customarily interpreted wealth narrowly, the measure is in fact an inclusive one. Wealth is based on a comprehensive list of assets, one that includes not only manufactured capital (roads and buildings, machinery and equipment, cables and ports) and human capital (health, knowledge, and skills), but also natural capital (oil and minerals, fisheries, forests, grazing land and aquifers, and ecosystems more broadly). Even though wealth is an aggregate measure, it can include distributional concerns if we are prepared to weight the wealth of different people differently before adding them.

To say that an economy's wealth has increased is to say that in terms of its worth, there has been an overall accumulation of capital assets. By the same token, to say that wealth has declined is to say that there has been an overall de-cumulation. Of course, even if some assets have de-cumulated, wealth would increase if there were a compensatory accumulation of other assets in the economy. I shall use the term inclusive

investment to mean a change in wealth at constant prices, regardless of whether the change is a decline or an increase.[15] Inclusive investment should be contrasted with recorded investment. Because a wide range of services obtained from natural capital are missing from standard economic accounts, recorded investment could be positive even if inclusive investment were negative. This would happen if the economy accumulated manufactured and human capital, but destroyed or degraded natural capital at a rapid rate, a possibility I explore later. However, current accounting practice does not recognize that nutrition, health care, and potable water are not merely consumption goods, but are simultaneously investment goods, so there is a corresponding undercount in recorded investment.

An asset's worth is measured in terms of the flow of benefits it is able to generate over time. Being the aggregate worth of all capital assets, wealth therefore reflects something like an economy's capacity to sustain human well-being, today and in the future. In fact one can say more: subject to certain qualifications, a rise in wealth per person, measured at constant shadow prices, corresponds to an increase in the average well-being of present and future generations, taken together. (For the most general theorem on this, see Arrow, Dasgupta, and Mäler 2003a; Dasgupta 2001a.) This is the sense in which wealth is a measure of intergenerational well-being. It is also the sense in which accumulation of wealth corresponds to sustained development. Inclusive investment is thus a key to economic progress.

The notion of inclusive investment I am advocating here is not only inclusive of various types of capital assets, but is also sensitive to individual and locational differences. Earlier I noted that a pond in one location is a different asset from a pond in another, because the ponds' ecological characteristics are likely to be different, and because the communities making use of them are likely to face different economic circumstances. Therefore seemingly identical ponds should have different shadow prices attributed to them. Of course, in practice such refinements may not be attainable, but it is always salutary to be reminded that macroeconomic reasoning glosses over the heterogeneity of the Earth's resources and the diverse uses to which they are put, both by people residing at the site and by those elsewhere. Shadow prices depend not only on technology and consumer preferences, but also on institutions and on their combined effect on people's lives.

Now consider in contrast GNP, which is the sum of consumption and (gross) investment in manufactured and human capital. GNP misleads not only because it ignores changes in the size and composition of much natural capital, but also because the index does not acknowledge that capital assets depreciate. So it is possible for GNP to increase over a period of time even while the economy's wealth declines. The output-wealth ratio merely increases when this happens. It would happen if increases in GNP were brought about by mining capital assets—for example, degrading ecosystems and depleting oil and mineral deposits—without investing appropriate amounts of output in the accumulation of other forms of capital, such as health and education. There is then little reason to expect movements in GNP to parallel movements in wealth. Of course, a situation where GNP grows and wealth declines cannot last forever. If wealth de-cumulates sufficiently, GNP will eventually have to decline also.

But the moral is telling: GNP (or for that matter, the HDI) is not a measure of long-run human well-being, meaning that movements in GNP (or the HDI) are a poor basis for judging economic progress.

What about the Residual?

The aggregate output of an economy is produced by various factors of production. We can therefore decompose observed changes in output over time into their sources: how much can be attributed to changes in labor force participation, how much to accumulation of manufactured capital and human capital, how much to the accumulation of knowledge brought about by expenditure on research and development, how much to changes in the use of natural resources, and so on. If a portion of the observed change in output cannot be credited to any of the foregoing factors of production, that portion is called the change in total factor productivity (TFP). Growth in TFP is also known as the residual to indicate that it is that bit of growth in output that cannot be explained.

Should wealth decline, could growth in TFP not compensate for the decline and ensure that long-run well-being is sustained? Traditionally, labor force participation, manufactured capital, and marketed natural resources have been the recorded factors of production. In recent years partial measures of human capital have been added. Attempts have also been made to correct for changes in the quality of manufactured capital brought about by research and development. But national accounts mostly still do not include the use of nonmarketed natural resources (nor, for that matter, of nonmarketed labor effort), for the understandable reason that shadow prices of nonmarketed natural resources are extremely hard to estimate. Moreover, how do you estimate unrecorded labor effort? Now imagine that over a period of time the economy makes increasing use of the natural resource base or of unrecorded labor effort. The residual would be overestimated. In fact a simple way to increase the residual would be to "mine" the natural resource base at an increasing rate, but this would be a perverse thing to do if we seek to measure economic prospects.

What if it were possible to decompose the growth of an economy's aggregate output in a comprehensive manner by tracing the growth to the sources originating in all the factors of production? To assume that over the long run the residual could still be positive is to imagine that the country enjoys a "free lunch." Is the latter a possibility? One way to enjoy a free lunch, for poor countries at least, is to use technological advances made in other countries without paying for them. The residual would then reflect increases in freely available knowledge. Note, however, that adaptation is not without cost. To meet local conditions, adjustments need to be made to product design and to the processes involved in production, all of which require appropriate local institutions, which are frequently missing in poor countries.

Of course, TFP can have short bursts in imperfect economies. Imagine that a government reduces economic inefficiencies by improving the enforcement of property rights or by reducing centralized regulations (import quotas, price controls, and so forth). We would expect the factors of production to find better uses. As factors realign in more productive fashions, TFP would increase.

TABLE 3.
Sources of Economic Growth, the United States and Selected Regions,
1960–94

Region or country	(A) g(Y/L)	(B) g(K)	(C) g(H)	A − (B + C) g(A)[a]
East Asia	4.2	2.5	0.6	1.1
South Asia	2.3	1.1	0.3	0.8
Africa	0.3	0.8	0.2	−0.6
Middle East	1.6	1.5	0.5	−0.3
Latin America	1.5	0.9	0.4	0.2
Industrial countries (excluding the United States)	2.9	1.5	0.4	1.1
United States	1.1	0.4	0.4	0.4

g(Y/L)   Annual percentage rate of change in GNP per head.

g(K)   Share of GNP attributable to manufactured capital multiplied by the annual percentage rate of change in manufactured capital.

g(H)   Share of GNP attributable to human capital multiplied by the annual percentage rate of change in human capital.

g(A)   Percentage rate of change in total factor productivity (residual).

[a]Rounding errors are possible.

Source: Collins and Bosworth (1996).

In the opposite vein, TFP could decline over a period. Increased government corruption could be a cause, or civil strife, which not only destroys capital assets, but also damages a country's public and civic institutions. When institutions deteriorate, assets are used even more inefficiently than previously: TFP declines.

Table 3, taken from Collins and Bosworth (1996), gives estimates of the annual rate of growth of GNP per head and its breakdown among two factors of production (manufactured and human capital) in various regions from 1960 to 1994. The estimates are given in the first three columns. The fourth column represents the residual in each region. Collins and Bosworth did not include nature's services as factors of production. If the use of those services has grown during the period in question (a most likely possibility), the residual is an overestimate. Even so, the residual in Africa was negative (−0.6 percent annually). The true residual was in all probability even lower. The residual in South Asia, the other really poor region of the world, was 0.8 percent annually, but as this is an overestimate, whether there has been any growth in TFP in that part of the world remains unclear.

Wealth Movements in Poor Countries

Even though there are no markets for many natural assets—and therefore no observable prices that reflect their worth—it is possible to estimate the shadow prices of nature's services if we are prepared to put in the effort and apply some low cunning.[16] Shadow prices measure the social worth of goods and services in an economy and are the ones to use in determining movements in wealth.

By estimating shadow prices and then adding net investment in natural capital to recorded investment, Hamilton and Clemens (1999) calculate inclusive investment in a large number of countries. There is a certain awkwardness in the steps the investigators took to arrive at their figures. Their accounts are also incomplete. For example, among the resources making up natural capital, only commercial forests, oil and minerals, and the atmosphere as a sink for carbon dioxide were included. Not included were water resources, forests as agents of carbon sequestration, fisheries, air and water pollutants, soil, and biodiversity. So there is an undercount, possibly a serious one. Moreover, some of the methods they deployed to estimate shadow prices are dubious. Nevertheless, if we are to ascertain the true macroeconomic character of the recent economic history of poor countries, we have to start somewhere.

The first column of figures in table 4 contains estimates of inclusive investment as a percentage of GNP. Notice that both Bangladesh and Nepal disinvested: aggregate capital assets declined there during the period in question. In contrast, inclusive investment was positive in China, India, Pakistan, and Sub-Saharan Africa. So the figures could suggest that the latter were wealthier at the end of the period than at the beginning, but when population growth is taken into account, the picture changes.

The second column of figures in table 4 provides the annual rate of growth of population over the period 1965–96. All but China experienced rates of growth in excess of 2 percent a year, with the rate in Sub-Saharan Africa and Pakistan having grown at nearly 3 percent a year.

TABLE 4
Genuine Investment and Wealth Accumulation, Selected Countries and Sub-Saharan Africa, Various Years

Region or country	I/Y (%)	g(L)	g(W/L)	g(Y/L)	Δ(HDI)
Bangladesh	−0.3	2.3	−2.40	1.0	Positive
India	10.7	2.1	−0.50	2.3	Positive
Nepal	−1.5	2.4	−2.60	1.0	Positive
Pakistan	8.2	2.9	−1.70	2.7	Positive
China	14.4	1.7	1.09	6.7	Positive
Sub-Saharan Africa	4.7	2.7	−2.00	−0.2	Positive

I/Y   Inclusive investment as a percentage of GNP. Inclusive investment includes total health expenditure (public plus private), estimated as an average during 1983–93.

g(L)   Average annual percentage rate of growth of population, 1965–96.

g(W/L)   Average annual percentage rate of change in per capita wealth at constant prices, 1970–93.

g(Y/L)   Average annual percentage rate of change in per capita GNP, 1965–96.

Δ(HDI)   Sign of change in the HDI, 1987–97.

Note: Assumed output-wealth ratio is 0.15 per year.

Sources: I/Y: Hamilton and Clemens (1999, tables 3 and 4); personal communication from Katie Bolt, World Bank; World Health Organization data; g(L): World Bank (1998, table 1.4); g(W/L): Author's calculations; g(Y/L): World Bank (1998, table 1.4); Δ(HDI): United Nations Development Programme (1990, 2000).

The third column of table 4 contains my estimates of the annual rate of change in per capita wealth at constant prices. The procedure I followed in arriving at the figures was to multiply inclusive investment as a proportion of GNP by the output-wealth ratio and to subtract the population growth rate from that product. This is a crude way to adjust for population change, but more accurate adjustments would involve greater computation.[17] Because a wide variety of capital assets (for example, human capital and various forms of natural capital) are unaccounted for in national accounts, there is a bias in published estimates of output-wealth ratios, which traditionally have been taken to be something like 0.25 per year. In arriving at the figures reported in the third column, I have used 0.15 per year as a check against the bias in traditional estimates for poor countries. Even this is almost certainly too high.

The striking message of the third column is that there was capital de-cumulation on a per capita basis during the period in question in all the places in the table other than China. This may not be a surprise in the case of Sub-Saharan Africa, which is widely known to have regressed in terms of most socioeconomic indicators, but the figures for Bangladesh, India, Nepal, and Pakistan should cause surprise. Even China, so greatly praised for its economic policies, has just about managed to accumulate wealth in advance of population growth. In any event, a more accurate figure for the output-wealth ratio would almost surely be considerably lower than 0.15. Using a lower figure would reduce China's accumulation rate. Moreover, the estimates of inclusive investment do not include soil erosion or urban pollution, both of which are thought by experts to be especially problematic in China.

How do changes in per capita wealth compare with changes in conventional measures of well-being? The fourth column of table 4 contains figures for the rate at which GNP per head changed during 1965–96, and the fifth column records whether the change in the United Nations' HDI over the period 1987–97 was positive or negative.

Judged in terms of movements in wealth per capita, notice how misleading our assessment of long-term economic development in the Indian subcontinent would be if we were to look at growth rates in per capita GNP. Pakistan, for example, would be seen as a country where GNP per head grew at a healthy 2.7 percent a year, implying that the index doubled in value between 1965 and 1993. The corresponding figure in the third column implies, however, that the average Pakistani became poorer by a factor of about 1.5 during that period. Bangladesh is recorded as having grown in terms of per capita GNP at 1 percent a year during 1965–96, but the figure in the third column of table 4 says that at the end of the period the average Bangladeshi was only about half as wealthy as he or she was at the beginning.

The case of Sub-Saharan Africa is especially depressing. At an annual rate of decline of 2 percent in per capita wealth, the average person in the region became poorer by nearly a factor of two over the period. The ills of Sub-Saharan Africa are routine reading in today's newspapers and magazines, but the ills are not depicted in terms of a decline in wealth. Table 4 suggests that Sub-Saharan Africa has experienced an enormous decline in its capital assets over the past three decades.

What of the HDI? As the third and fifth columns of table 4 show, HDI offers a picture that is very different from the one based on wealth per head. Only China's

performance does not mislead: the HDI increased there. However, for Sub-Saharan Africa the index grew. Moreover, Bangladesh and Nepal have been exemplary in terms of the HDI. The HDI misleads even more than GNP per head.

The figures in table 4 for movements in per capita wealth are rough and ready, and we should not regard them with anything like the certitude we have developed over the years for international statistics on GNP and the demographic and morbidity statistics of poor countries. My estimates are a first cut at what is an enormously difficult set of exercises. But the figures, such as they are, show how accounting for natural capital can make for substantial differences in our conception of the processes of economic development. We would by now have been far ahead in our understanding of what really has happened in poor countries if development economists had taken nature's services seriously in the past.

Micro-Level

Having presented an overview of recent experiences in the world's poorest regions, I now turn to an analysis of the various pathways that may account for them.

Nutrition, Health, and Human Productivity

One component of inclusive wealth is human health. Here we regard health as a constituent of human capital. Nutrition, potable water, and health care should therefore be thought of as investment goods, not merely consumption goods.

The energy requirement for maintaining human life is substantial: 60 to 75 percent of the energy intake of someone in nutrition balance goes toward maintenance, with the remaining 25 to 40 percent spent on "discretionary" (work and leisure) activities. Maintenance requirements are like fixed costs. They lead to positive feedback and are the sources of the nonconvexity inherent in human metabolic pathways.

Why can the market mechanism not be relied upon to eliminate undernutrition? The reason is the large energy maintenance requirement for human functioning. Suppose that in a poor country a large fraction of people possess nothing other than their potential labor power. It can be shown that the market mechanism cannot ensure that everyone earns the nutrition intake they need to be able to fulfill their potential. To be precise, it can be shown that if the economy is not wealthy on a per capita basis, a fraction of the asset-less would be shut out of the market: the market for raw labor would not clear. Initial poverty in such an economy would be the source of emerging inequality. In a poor economy, asset-less people operate on the boundary of the nonconvex region of their nutrition-productivity possibilities, whereas people who possess sufficient assets are able to get onto the boundary of the convex region (Dasgupta and Ray 1986). The point is that the undernourished are at a severe disadvantage in their ability to obtain food: the quality of work they are able to offer is inadequate for obtaining the food they require if they are to improve their nutritional status. Thus, over time,

undernourishment can be both a cause and a consequence of someone falling into a poverty trap. Because undernourishment displays hysteresis, such poverty can be dynastic: once a household falls into a poverty trap, it can prove especially hard for descendents to emerge out of it.

It can also be shown that if the distribution of nonhuman assets were made less unequal in the foregoing economy, the market for labor would function better. Of course, the key issue is access to nutrition and health care, not so much the distribution of assets. Safety nets would provide that access to people if all else fails for them (Dasgupta and Ray 1987).

Much international attention has been given to saving lives in times of crises in poor countries. This is as it should be. International agencies have also paid attention to keeping children alive in normal times through public health measures, such as family planning counseling, immunization, and oral rehydration. This too is as it should be. That many poor countries fail to do either is not evidence of the problems' being especially hard to solve. In fact they are among the easier social problems: they can be addressed even while no major modification is made to the prevailing resource allocation mechanism. Much the harder problem, in intellectual design, political commitment, and administration, is to ensure that those who remain alive are healthy. It is also a problem whose solution brings no easily visible benefit. But the persistence of large-scale undernourishment caused by inadequate diet and lack of sanitation and potable water is a sure sign of economic backwardness. For example, the stunting of both cognitive and motor capacity is a prime hidden cost of energy deficiency and anemia among children and, at one step removed, among mothers. It affects learning and skill formation, and thereby future productivity. The price is paid in later years, but it is paid.

Reproductive and Environmental Externalities

If in recent decades poor countries have experienced both a decline in wealth per head and unprecedented population growth, what accounts for this? Contemporary writings on population growth in poor countries stress that there is a negative link between education (especially female education) and fertility. So it is now a commonly held view that female education triggers fertility reduction. The problem with this viewpoint is that the education elasticity of fertility would appear to differ substantially across space and time.[18] Moreover, Cochrane, who is responsible for the first clear studies showing the links between female education and fertility reduction, was herself reluctant to attribute causality to her findings (Cochrane 1979, 1983), as were investigators studying more recent data (Cohen 1993; Jolly and Gribble 1993), for the reason that it is extremely difficult to establish causality. Women's education may well reduce fertility. However, the initiation of childbearing may be a factor in the termination of education. Moreover, even when education is made available by the state, households may choose not to take up the opportunity: the ability (or willingness) of governments in poor countries to enforce school attendance or make good education facilities available is frequently extremely limited. Economic costs

and benefits of education and the mores of the community to which people belong influence their decisions. It could be that the very characteristics of a community that are reflected in low educational attainment for women are also those giving rise to high fertility (for example, absence of associational activities among women, lack of communication with the outside world, or inheritance rules that place women at a disadvantage).

Demographic theories striving for generality would regard both women's education and fertility to be endogenous variables. The negative relationship between education and fertility in such theories would be an association, not a causal relationship. The two variables would be interpreted as moving together in samples, nothing more. The models underlying the following discussion are based on institutional and ecological fundamentals. They stand in contrast to the prevailing view that female education is the driver of fertility reduction.

Children are both ends in themselves and a means to economic betterment—even survival. What causes the private and social costs (and benefits) of reproduction to differ? One likely source of the differences has to do with the finiteness of space (see, for example, Harford 1998). A larger population means greater crowding, and households acting on their own would not be expected to "internalize" crowding externalities. The human epidemiological environment becomes more and more precarious as population densities rise. Crowded centers of population provide a fertile ground for the spread of pathogens, and there are always new strains in the making. Conversely, the spread of infections, such as HIV, would be expected to affect demographic behavior, though in ways that are not yet obvious (Ezzell 2000).

Large-scale migrations of populations occasioned by crop failure, war, or other disturbances are an obvious form of externality, but by their very nature they are not of the persistent variety. Of those that are persistent, four types come to mind. In what follows, I discuss them.

Cost Sharing

Fertility behavior is influenced by the structure of property rights, for instance, rules of inheritance. In his influential analysis of fertility differences between preindustrial 17th and 18th century northwest Europe on the one hand, and Asiatic preindustrial societies on the other, Hajnal (1982) distinguishes between nuclear and joint household systems. He observes that in northwest Europe, marriage normally meant establishing a new household, which implied that the couple had to have, by saving or transfer, sufficient resources to establish and equip the new residence. This requirement in turn led to late marriages. It also meant that parents bore the cost of rearing their children. Indeed, fertility rates in England were a low 4 in 1650–1710, long before modern family planning techniques became available and long before women became literate (Coale 1969; Wrigley and Schofield 1981). Hajnal contrasted this with the Asiatic pattern of household formation, which he saw as joint units consisting of more than one couple and their children.

Parental costs of procreation are also lower when the cost of rearing the child is shared among the kinship. In Sub-Saharan Africa, fosterage within the kinship is

commonplace. Children are not raised solely by their parents: the responsibility is more diffuse within the kinship group (Bledsoe 1990; Caldwell and Caldwell 1990). Fosterage in the African context is not adoption. It is not intended to, nor does it in fact, break ties between parents and children. The institution affords a form of mutual insurance protection in semi-arid regions. It is possible that, because opportunities for saving are few in the low-productivity agricultural regions of Sub-Saharan Africa, fosterage also enables households to smooth their consumption across time.[19] In parts of West Africa, up to half the children have been found to be living with kin at any given time. Nephews and nieces have the same rights of accommodation and support as do biological offspring. There is a sense in which children are seen as a common responsibility. However, the arrangement creates a free-rider problem if the parents' share of the benefits from having children exceeds their share of the costs. (For a proof of the proposition, see Dasgupta 1993.) From the point of view of parents, taken as a collective, too many children would be produced in those circumstances.

In Sub-Saharan Africa, communal land tenure within the lineage social structure has in the past offered further inducement for men to procreate. Moreover, conjugal bonds are frequently weak, so fathers often do not bear the costs of siring children. Anthropologists have observed that the unit of African society is a woman and her children rather than parents and their children. Frequently there is no common budget for the man and woman. Descent in Sub-Saharan Africa is for the most part patrilineal and residence is patrilocal (the Akan people of Ghana are an exception). Patrilineality, weak conjugal bonds, communal land tenure, and a strong kinship support system of children, taken together, have been a broad characteristic of the region (see Bledsoe and Pison 1994; Caldwell 1991; Caldwell and Caldwell 1990). They are a source of reproductive externalities that stimulate fertility. Admittedly, patrilineality and patrilocality are features of the northern parts of the Indian subcontinent also,[20] but conjugal bonds are substantially greater there. Moreover, because agricultural land is not communally held in India, large family size leads to fragmentation of landholdings. In contrast, large families in Sub-Saharan Africa are (or at least were until recently) rewarded by a greater share of land belonging to the lineage or clan.

A desire to pool risks means, more generally, that material gains from good fortune are shared among one's kinship. But it has been observed by social scientists that in African societies the fruits of hard work and thrift are not distinguished greatly from good fortune (Platteau and Hayami 1998). This dulls private incentives.

Conformity and "Contagion"

That children are seen as an end in themselves provides another mechanism by which reasoned fertility decisions at the level of every household can lead to an unsatisfactory outcome from the perspectives of all households. The mechanism arises from the possibility that traditional practice is perpetuated by conformity. Procreation in closely knit communities is not only a private matter, it is also affected by social norms, influenced by both family experiences and the cultural milieu. Formally

speaking, behavior is conformist if, other things being equal, the greater every household's most desired family size, the larger the average family size in the community (Dasgupta 1993). This formulation of conformism is a reduced form of the concept, and the source of a desire to conform could lie in reasons other than an intrinsic desire to be like others. For example, similar choices made by households might generate mutual positive externalities, say, because people care about their status, and a household's choice of actions signals its predispositions (for instance, its willingness to belong), and so affects its status (Bongaarts and Watkins 1996). In a world where people conform, the desire for children is endogenous.[21]

Whatever the basis of conformism, there would be practices encouraging high fertility rates that no household would unilaterally desire to break. Such practices could well have had a rationale in the past, when mortality rates were high, rural population densities were low, threats of extermination from outside attack were large, and mobility was restricted, but practices can survive even when their original purposes have disappeared. Thus, as long as all others follow the practice and aim at large family size, no household on its own may wish to deviate from the practice; however, if all other households were to restrict their fertility rates, each would desire to restrict its fertility rate as well. In short, conformism can be a reason for the existence of multiple reproductive equilibriums (Dasgupta 1993). The multiple equilibriums may even be Pareto rankable, in which case a community could get stuck at an equilibrium mode of behavior even though another equilibrium mode of behavior would be better for all.

These are theoretical possibilities. Testing for multiple equilibriums is difficult. As matters stand, it is only analytical reasoning that tells us that a society could, in principle, get stuck at a self-sustaining mode of behavior characterized by high fertility (and low educational attainment), even when there is another, potentially self-sustaining, mode of behavior characterized by low fertility (and high educational attainment).

This does not mean that the hypothetical society would be stuck with high fertility rates forever. External events could lead households to "coordinate" at a low fertility equilibrium even if they had earlier coordinated at a high fertility equilibrium. The external events could, for example, take the form of public exhortations aimed at altering household expectations about one another's behavior (for example, family planning campaigns run by women). This is a case where the community "tips" from one mode of behavior to another, even though there has been no underlying change in household attitudes to trigger the change in behavior.

In addition to being a response to external events, the tipping phenomenon can occur because of changes in the peer group on whose behavior households base their own behavior. Inevitably, there are those who experiment, take risks, and refrain from joining the crowd. They subsequently influence others. These are the tradition breakers, who often lead the way. It has been observed that educated women are among the first to make the move toward smaller families (see Farooq, Ekanem, and Ojelade 1987 for a commentary on West Africa). Members of the middle classes can also be the trigger, becoming role models for others.

A possibly even stronger pathway is the influence that newspapers, radio, television, and now the Internet exert in transmitting information about other lifestyles (Bongaarts and Watkins 1996; Iyer 2002). The analytical point here is that the media may be a vehicle through which conformism increasingly becomes based on the behavior of a wider population than the local community: the peer group widens. Such pathways can give rise to demographic transitions, in that fertility rates display little to no trend over extended periods, only to cascade downward over a relatively short interval of time, giving rise to the classic logistic curve of diffusion processes.

In a pioneering article, Adelman and Morris (1965) found the openness of a society to outside ideas to be a powerful stimulus to economic growth. It is possible that the fertility reductions that have been experienced in Bangladesh and India in recent years were the result of the wider influence people have been subjected to via the media or to attitudinal differences arising from improvements in family planning programs. To be sure, fertility reductions have differed widely across the Indian subcontinent (not much reduction in Pakistan so far, a great deal in southern India), but we should not seek a single explanation for so complex a phenomenon as fertility transition.[22]

Demographers have made few attempts to discover evidence of behavior that is guided in part by an attention to others. Two exceptions are Easterlin, Pollak, and Wachter (1980) and Watkins (1990).[23] The former studied intergenerational influence in a sample of families in the United States, and they report a positive link between the number of children with whom someone had been raised and the number of children they themselves had. The latter studied demographic change in Western Europe over the period 1870–1960 and shows that regional differences in fertility and nuptiality within each country declined. In 1870, before the large-scale declines in marital fertility had begun in most areas of Western Europe, demographic behavior differed greatly within countries: provinces (for instance, counties and cantons) differed considerably, even while differences within provinces were low. There were thus spatial clumps within each country, suggesting the importance of the influence of local communities on behavior. By 1960, differences within each country were less than they had been in 1870. Watkins explains this convergence in behavior in terms of increases in the geographical reach national governments enjoyed over the 90 years in question. The growth of national languages could have been the medium through which reproductive behavior spread.

One recent finding could also point to contagious behavior. Starting in 1977, when the total fertility rate in Bangladesh exceeded 6, 70 so-called treatment villages in Matlab Thana were served by a massive program of birth control, while 79 control villages were offered no such special service. The prevalence of contraceptive use in the treatment villages increased from 7 to 33 percent within 18 months, and then rose more gradually to a level of 45 percent by 1985. The prevalence also increased in the control villages, but only to 16 percent by 1985. Fertility rates in both sets of villages declined, but at different speeds, with the difference in fertility rates reaching 1.5 births per woman, even though there had been no difference to begin with (Hill 1992). If we assume that, even though influence travels,

geographical proximity matters, we could explain why the control villages followed the example of treatment villages but did not follow them all the way. Contagion did not spread completely.[24]

Interactions among Institutions

Externalities are prevalent when market and nonmarket institutions coexist. Arnott and Stiglitz (1991) have developed a formal account of the externalities that prevail when market and nonmarket institutions supply what is, in effect, the same commodity (for example, insurance). Their purpose was to show that the existence of nonmarket institutions can prevent people from transacting as much in the market as they should for their collective good. The argument can be extended to include the reluctance people have to engage in the public sphere of life when communitarian institutions are entrenched at the local level. In such circumstances the state is also able to be predatory, which in turn can erode communitarian institutions. The message of this analysis is that if nonmarket institutions are entrenched, they can prevent potentially more efficient institutions from emerging. In what follows, I present a case offering the opposite moral, namely, that an expansion of markets can destroy nonmarket institutions and make certain vulnerable groups worse off. How and why might such externalities affect fertility behavior? A number of pathways suggest themselves (Dasgupta 1993, 1999).

Long-term cooperation is frequently sustained by social norms, for example, norms of reciprocity. Social norms can be reliably observed only among people who expect to encounter one another in recurring situations.[25] Consider a community of "far-sighted" people who know one another and expect to interact with one another for a long time. By far-sighted, I mean someone who applies a low rate to discount future costs and benefits of alternative courses of action. Assume that the parties in question are not individually mobile (although they could be collectively mobile, as in the case of nomadic societies); otherwise the chance of future encounters with one another would be low, and people would discount heavily the future benefits of the current costs they incur for the purposes of cooperation.

Simply stated, if people are far-sighted and are not individually mobile, a credible threat by all that they would impose stiff sanctions on anyone who broke an agreement would deter everyone from breaking it, but the threat of sanctions would cease to have bite if opportunistic behavior were to become personally more profitable. The latter would happen if formal markets develop nearby. As opportunities outside the village improve, people with lesser ties, such as young men, are more likely to take advantage of them and make a break with those customary obligations that are enshrined in the prevailing social norms. People with greater attachments would perceive this and infer that the expected benefits from complying with agreements are now lower. Norms of reciprocity would break down, making certain groups of people (for instance, women, the old, and the very young) worse off. This is a case where improved institutional performance elsewhere, such as the growth of markets in the economy at large, has an adverse effect on the functioning of a local, nonmarket institution: it is a reflection of an externality.

When long-term relationships break down, people build new ones to further their economic opportunities. Those who face particularly stressful circumstances resort to draconian measures to build new economic channels. Guyer (1994) observes that in the face of deteriorating economic circumstances, some women in the Yaruba area of Nigeria have borne children by different men so as to create immediate lateral links with them. Polyandrous motherhood enables women to have access to more than one resource network.

Cain (1981, 1983) shows that where capital markets are nonexistent and public or community support for the elderly is weak, children provide security in old age. The converse is that if community-based support systems decline, children become more valuable. But we have just noted that community-based support systems in rural areas may degrade with the growth of markets in cities and towns. So there is a curious causal chain here: the growth of markets in towns and cities can lead to an increase in fertility in poor villages, other things being the same. There is evidence of this. In her work on Sarawak, Malaysia, Heyzer (1996) observes that half the forest area there has now been lost and that this has disrupted the lives of indigenous people in different ways. Communities that lived in the heart of the forest were most severely affected, while others living near towns were able to turn from swidden agriculture to wage labor. This transformation, however, involved male migration, leaving women behind to cope with a decreasing resource base. As subsistence alternatives declined, children became one of the few remaining resources that women could control. There was thus a new motivation for having children: to help their mothers with an increased workload. The process involved the creation of new patterns of wealth and poverty, where wealth is based on resource extraction and poverty results from the loss of a community's resource base.

Of course, by making children less reliable as an investment for old age, the growth of markets in towns and cities can lead to a reduction in fertility. Here we have identified an influence of the growth of markets on fertility that runs in the opposite direction. Only formal modeling of the process would enable us to determine which influence dominates under what conditions.

Household Labor Needs and the Local Commons

Among poor households in rural communities, much labor is needed even for simple tasks.[26] Moreover, many households lack access to the sources of domestic energy available to households in advanced industrial countries. Nor do they have water on tap. In semi-arid and arid regions, water supply is often not even close at hand, nor is fuelwood nearby when the forests recede. This means that the relative prices of alternative sources of energy and water faced by rural households in poor countries are quite different from those faced by households elsewhere. In addition to cultivating crops; caring for livestock; cooking food; and producing simple, marketable products, household members may have to spend several hours a day fetching water and collecting fodder and wood. These complementary activities have to be undertaken on a daily basis if households are to survive. Labor productivity is low, because both manufactured capital and environmental resources are scarce. From an early age (as early as six), children in poor households in the poorest countries take care

of their siblings and domestic animals; fetch water; and collect fuelwood, dung (in the Indian subcontinent), and fodder. Generally they do not go to school. Not only are education facilities in the typical rural school woefully inadequate, but parents need their children's labor. Children between 10 and 15 years old have been routinely observed to work at least as many hours as adult males (Filmer and Pritchett 2002).

The need for many hands can, in principle, lead to a destructive situation when parents do not have to pay the full price of rearing their children but share such costs with their community. In recent years, social norms that once regulated local resources have changed. Since time immemorial, rural assets such as village ponds and water holes, threshing grounds, grazing fields, swidden fallows, and local forests and woodlands have typically been owned communally. As a proportion of total assets, the presence of such assets ranges widely across ecological zones. In India, the local commons are most prominent in arid regions, mountain regions, and unirrigated areas; they are least prominent in humid regions and river valleys (Agarwal and Narain 1989). There is a rationale for this, based on the human desire to reduce risks. Community ownership and control enabled households in semi-arid regions to pool their risks. An almost immediate empirical corollary is that income inequalities are less where common property resources are more prominent. Aggregate income is a different matter however, and the arid and mountain regions and unirrigated areas are the poorest. As would be expected, dependence on common property resources even within dry regions would appear to decline with increasing wealth across households.

Jodha (1986, 1995), studying evidence from more than 80 villages in 21 dry districts in India, concluded that among poor families, the proportion of income based directly on their local commons is for the most part in the range of 15 to 25 percent. A number of resources, such as fuelwood and water, berries and nuts, medicinal herbs, and resin and gum, are the responsibility of women and children. In a study of 29 villages in southeastern Zimbabwe, Cavendish (2000) arrives at even larger estimates: the proportion of income based directly on the local commons is 35 percent, with the figure for the poorest quintile reaching 40 percent. Such evidence does not, of course, prove that the local commons are well managed, but it suggests that rural households have strong incentives to devise arrangements whereby they would be well managed.

A number of investigators have shown that many communities have traditionally protected their local commons from overexploitation by relying on social norms, by imposing fines for deviant behavior, and by other means (see, for example, Baland and Platteau 1996; Chopra, Kadekodi, and Murty 1990; Howe 1986, Ostrom 1990, 1992; Wade 1988).[27] I argued earlier that the very process of economic development, as exemplified by urbanization and mobility, can erode traditional methods of control. Social norms are also endangered by civil strife and by the usurpation of resources by landowners or the state. For example, resource allocation rules practiced at the local level have frequently been overturned by central fiat. A number of states in the Sahel imposed rules that in effect destroyed community management practices in the forests. Villages ceased to have authority to enforce sanctions against

those who violated locally instituted rules of use. State authority turned the local commons into free-access resources (see Baland and Platteau 1996; Thomson, Feeny, and Oakerson 1986).[28] As social norms degrade, whatever the cause, parents pass some of the costs of children on to the community by overexploiting the commons. This is another instance of a demographic free-rider problem.

The perception of an increase in the net benefits of having children induces households to have too many. This is predicted by the standard theory of the imperfectly managed commons. It is also true that when households are further impoverished because of the erosion of the commons, the net cost of children increases (of course, household size continues to remain above what is desirable from the collective point of view). Loughran and Pritchett (1998), for example, find that in Nepal, increasing environmental scarcity lowered the demand for children, implying that the households in question perceived resource scarcity as raising the cost of children. Apparently, increasing firewood and water scarcity in the villages in the sample did not have a strong enough effect on the relative productivity of child labor to induce higher demand for children, given the effects that work in the opposite direction. Environmental scarcity there acted as a check on population growth.

However, theoretical considerations suggest that, *in certain circumstances*, increased resource scarcity induces further population growth.[29] As the community's natural resources are depleted, households find themselves needing more hands. No doubt additional hands could be obtained if the adults worked even harder, but in many cultures it would not do for the men to gather fuelwood and fetch water for household use. No doubt, too, additional hands could be obtained if children at school were withdrawn and put to work, but, as we have seen, mostly the children do not go to school anyway. In short, when all other sources of additional labor become too costly, more children are produced, thereby further damaging the local resource base and, in turn, providing the household with an incentive to enlarge still more. This does not necessarily mean that the fertility rate will increase. If the infant mortality rate were to decline, there would be no need for more births for a household to acquire more hands. However, along this pathway, poverty, household size, and environmental degradation could reinforce one another. By the time some countervailing set of factors diminished the benefits of having more children, many people could have experienced a worsening of poverty.

Cleaver and Schreiber (1994) provide rough, aggregative evidence of a positive link between population increase and environmental degradation in the context of rural Sub-Saharan Africa; Batliwala and Reddy (1994) do so for villages in Karnataka, India; and Heyzer (1996) does so for Sarawak, Malaysia. In a statistical analysis of evidence from villages in South Africa, Aggarwal, Netanyahu, and Romano (2001) find a positive link between fertility increase and environmental degradation, while Filmer and Pritchett (2002) report a weak positive link in the Sindh region of Pakistan.

None of these investigations quite captures what the theory I am sketching here tells us to study, namely, the link between desired household size and the state of the local natural resource base, but they come close enough. Limitations in existing data

prevent investigators from getting closer to the theory.[30] In any event, these studies cannot reveal causal connections, but, except for the study by Loughran and Pritchett (1998), they are not inconsistent with the idea of a positive feedback mechanism such as I have described. Over time, the feedback would be expected to have political effects, as manifested by battles for scarce resources, for example, among competing ethnic groups (Collier and others 2003; Durham 1979; Homer-Dixon 1994, 1999). The latter connection deserves greater investigation than it has elicited so far.

To be sure, families with greater access to resources would be in a position to limit their size and propel themselves into still higher income levels. Admittedly too, people from the poorest of backgrounds have been known to improve their circumstances. Nevertheless, there are forces at work that pull households away from one another in terms of their living standards. Such forces enable extreme poverty to persist despite growth in well-being for the rest of society.

Conclusions

The dependence of human well-being on the natural environment is central to the research concerns of geographers. Elementary textbooks on human geography invariably contain chapters explaining why people eat what they eat, wear what they wear, use the materials they use, and, more generally, live the way they live by reference to where they live. Models of ecosystem dynamics, for example, explain the emergence of population heterogeneity and modularity. They tell us that patchyness is the order of things in a landscape, not uniformity (see, in particular, Levin 1999).

Economists, in contrast, have moved steadily away from viewing location as a determinant of the human experience.[31] Economic progress itself is seen to be a release from location's grip on life. We economists emphasize that over the centuries, investment has reduced transportation costs. We observe too the role of industrialization in ironing out the effects of geographical differences on societies, such as differences in climate, soil quality, distance from navigable water, and, concomitantly, local ecosystems. Modern growth theories have tended to dismiss geography, and thus ecology, as being no more than a negligible factor in economic progress (see, for example, Jones 1998). The term globalization is used to signal that location per se does not matter.

Justifying Remarks

In this paper I have explored mechanisms that would seem to be responsible for the persistence of acute poverty in Sub-Saharan Africa and large parts of the Indian subcontinent. Both the character of human metabolic pathways and weaknesses in human–nature interactions were shown to play significant roles. The latter seems an obvious enough thought, because the rural poor in the world's poorest regions either practice subsistence agriculture or live around it. The fact that declines in mortality

in recent decades were not matched by fertility decline (table 2) seemed to me to be allied to that problem, and therefore also in need of explanation.

Toward that end, externalities in human-nature interactions were the natural starting point, reflecting as they do institutional failure, including the failure of communities to come to grips with local resource allocation problems. But institutional failure includes government and market failure as well. Nonconvexities in the underlying ecological and economic processes imply that small initial differences in the performance of institutions across space (arising, say, from small differences in the costs of monitoring one another's actions) can lead to large differences in their economic consequences over time, other things being the same. Nonconvexities also mean that small differences in the local ecology can lead to growing differences in the economic prospects facing societies, other things being the same.

Policy

Inclusive wealth is a measure of social well-being. By this I mean that an increase in wealth, at constant prices, reflects an improvement in the well-being of generations, taken as a whole. I noted earlier that wealth can be used to judge whether the long-run, average well-being of a community is being sustained. It transpires that the present discounted value of the flow of social profits of a policy reform, for example, an investment project, is the contribution the reform makes to wealth (Arrow, Dasgupta, and Mäler 2003a; Dasgupta 2001a). So, wealth can also serve in policy evaluation.

Because wealth comprises manufactured, human, and natural capital, the public policies that recommend themselves are those that would lead to the accumulation of an appropriate mix of those capital assets. Identifying an appropriate mix requires knowledge of shadow prices. Estimating the shadow prices of local environmental resources is now of the greatest importance for countries.

Development economists frequently observe that estimating shadow prices is fraught with difficulties. They say it involves subjective elements and so is best avoided. The current practice in national accounting reflects this reasoning when it treats nonmarketed environmental resources as valueless, but this cannot be right. It is like concluding that a problem does not exist from the fact that it is difficult. In any event, there is usually some information on which one can draw to estimate shadow prices as a first approximation. It would also do no harm to cultivate the practice of offering a range of values. Sensitivity analysis is a useful exercise.

Nutrition, potable water, sanitation, and health care, on the one hand, and education, on the other, are inputs into the production of human capital. Externalities are inherent in the provision and use of a number of these commodities. Moreover, some involve large fixed costs. This means that markets on their own cannot be relied upon to select the right projects. Markets need to be augmented by public investment and engagement.

Traditional welfare economics recommends public involvement in the production and maintenance of infrastructure because its provision also involves large fixed

costs, but whether a fixed cost is large is a relative matter. What may be a trivial amount of money and effort for a rich community would be large for a poor community. Some of the infrastructure of importance in a rural setting comprises the local natural resource base. The fact that it is local means that public engagement in its protection and promotion does not necessarily mean government engagement. I noted earlier that the state has, in many instances, helped to destroy communitarian institutions. In such cases the state itself has been the problem. But an enlightened state can offer aid to build communitarian institutions if they have been destroyed or help to improve their workings if changing circumstances have created unexpected problems. Thus the local community is the right institution for managing local externalities and local public goods, while the state is appropriate for those that have a wider reach, such as enforcing the rule of law.

An absence of efficient risk markets means that safety nets are of vital importance to the poorest of the poor. Some risks are personal, while others involve the entire community. The state should be the right institution for providing security against the latter, while security from some of the former types of risk are probably best provided by appropriately designed communitarian institutions. As safety nets are frequently advocated today, this may seem to be a banal conclusion to reach, but the trick is to know who require safety nets, why they need them, and how much they need them if they are to propel themselves out of the mire. Only the application of good theory can tell us that.

Each such engagement requires resources. But as the pathways we have examined in this paper involve thresholds (the basis of several of the positive feedback mechanisms we have studied), the required resources need to come in lumps: small amounts at a time are pretty much useless. The currently rich countries managed to mobilize such resources in earlier times (they too were once poor). There are a number of countries, until recently poor, that also managed to do so. But there are many others that simply cannot afford such resources now. I noted earlier that systems characterized by positive feedback often possess multiple basins of attraction. It is thus possible that the countries now caught in a poverty trap faced good economic prospects once but are where they are now because of ill-judged policies and bad institutions. Getting those economies out of the trap may require external help, even if the institutions were to improve and the economic policies now chosen were sound. It seems to me that the theory of poverty traps I have outlined here provides an intellectual reconciliation between those who insist that the poorest countries of the world are where they are now because of their own fault and those who plead that those same countries today require external help if they are to lift themselves out of poverty.

Notes

1. Spurr (1988, 1990), World Health Organization (1983, 1985), and Waterlow (1992) are classic publications about the metabolic pathways themselves.
2. For evidence on the synergies, see Harrison and Waterlow (1990) and Scrimshaw and others (1968).

3. During the 1974 famine in Bangladesh, deaths in excess of those that would have occurred under previous nutritional conditions numbered around 1.5 million. The stock was replenished within a year (Bongaarts and Cain 1981). Of course, undernourishment can still have an effect on sexual reproduction through its implications for the frequency of stillbirths, age at menarche, failure to ovulate, maternal and infant mortality, and the frequency of sexual intercourse.

4. Gallup, Sachs, and Mellinger (1999) observe that location per se, and not merely the local ecology, can matter. They note, for example, that to be landlocked and surrounded by poor neighbors further reduces a country's economic options.

5. For example, the financial estimates of resources required to meet the Millennium Development Goals.

6. Which is, of course, a misnomer. The original Kuznets curve, which was an inverted U, related income inequality to real national income per head on the basis of country-level time series.

7. Ecologists use the more general term nonlinear to highlight what economists call nonconvex. The term linear (and its negative, nonlinear) is widely understood. The same is probably not true for the term nonconvex. The definition is as follows. A commodity bundle, say z, is a convex combination of commodity bundles x and y if z is a weighted average of x and y, where the weights are positive and sum to unity. A set of commodity bundles is said to be convex if every convex combination of every pair of commodity bundles in the set is in the set. A set is nonconvex if it is not convex. Ecological and socioeconomic pathways involving positive feedback create nonconvex transformation possibility sets among goods and services. Even though it may seem to be an esoteric notion, whether or not transformation possibility sets are nonconvex has serious implications for the efficacy (or its lack) of the market mechanism. On why, see Debreu (1959) and Koopmans (1957).

8. For further discussions of the environmental Kuznets curve, see Arrow and others (1995) and the responses it elicited in symposia built around the article in *Ecological Economics* 1995, 15(1); *Ecological Applications* 1996, 6(1); and *Environment and Development Economics* 1996, 1(1). See also the special issue of *Environment and Development Economics* 1997, 2(4); and Dasgupta, Levin, and Lubchenco (2000). In a fine recent paper, Anderies (2003) develops a model of an economic process involving positive feedback where economic growth and technological progress are destabilizing. He shows that if the rate at which natural resources regenerate is slow in comparison with economic growth and technological progress, population overshoots and the natural resource base collapses (see also Brander and Taylor 1998). For the mathematics of nonlinear ecological processes, see Murray (1993).

9. Sen (1994) is even contemptuous of those who argue that high population growth has been a hindrance to economic betterment in Sub-Saharan Africa and the Indian subcontinent.

10. The claim holds even if the past 200 years were to be included. The rough calculation is simple enough. World per capita output today is about US$5,000. The World Bank regards US$1 a day to be about as bad as it can be: people would not be able to survive on anything much less than that. It would then be reasonable to suppose that 2,000 years ago per capita income was not less than US$1 a day, so let us assume that it was US$1 a day. This would mean that per capita income 2,000 years ago was about US$350 a year. Rounding off numbers, this means, very roughly speaking, that per capita income has risen about 16 times since then. This in turn means that world income per head has doubled every 500 years, which in its turn means that the average annual rate of growth has been about 0.14 percent per year, a figure not much in excess of zero.

11. Landes (1969, 1998) provides compelling accounts of that experience. Fogel (1994, 1999) offers a highly original account of how improvements in nutrition intake helped propel European economies in earlier centuries.

12. Estimates of the poverty gaps in South Asia and Sub-Saharan Africa suggest that, even in those regions, a mere 4 percent growth in income, if it were distributed efficiently among the poor, would eliminate the extreme poverty reported in table 1.

13. I have gone into these issues in greater detail in Dasgupta (1998a). Estimates of poverty based on anthropometric indicators remain sketchy. James and others (1992) is an exception.

14. One can even argue that because it does not take note of capital depreciation, GNP cannot be a measure of opulence.

15. Inclusive investment is called genuine saving by Hamilton and Clemens (1999).

16. The search for ways to estimate shadow prices of natural capital is an active field of research today. The hard part of the work lies in determining the connectedness of natural capital from a study of the ecological processes at work.

17. Arrow, Dasgupta, and Mäler (2003b) develop precise formulas for how the conversion ought to be done under a variety of circumstances.

18. There are also places in Africa where the elasticity has been found to have the "wrong" sign (Jolly and Gribble 1993).

19. This hypothesis could be tested by comparing the age structure of households that foster out and those that foster in.

20. Among the prominent Nayyars of the southern state of Kerala, India, descent is matrilineal. Kerala is noteworthy today for being among the poorer of Indian states even while attaining a total fertility rate of less than 2.

21. Household preferences embodying such interactions are often called social preferences.

22. In this connection, the Indian state of Andhra Pradesh offers an interesting example. Female illiteracy there is a high 55 percent and some 75 percent of the population have access to radio or television. The fertility rate there is now 2.3, whereas it was in excess of 5.0 three decades ago.

23. A recent exception is Krishnan (2001) on data from India.

24. I am grateful to Lincoln Chen for helpful 1996 correspondence on this point. For a formal account of contagion models, see Blume and Durlauf (2000).

25. This is the setting studied in the theory of repeated games. Maintaining reputation could also be a reason why people keep trust, but such motives for cooperative behavior are not based on social norms. See Fudenberg and Tirole (1991).

26. See Dasgupta (2000, appendix) for a formal model that captures the ideas developed in this section.

27. See Dasgupta and Heal (1979, chapter 3) for the theory of the commons, both when managed cooperatively and when not.

28. Eicher (1999) traces Sub-Saharan Africa's current inability to feed itself to state policies in the 1960s and 1970s that amounted to the emasculation of agriculture.

29. I emphasize the qualification, because the theory has been misunderstood by a number of colleagues to be saying that the negative link between local resource availability and fertility is unconditional. Bearing and raising children involve costs. In the text I refer to circumstances where those costs are outweighed by the benefits of further procreation.

30. Deon Filmer has informed me that his colleagues at the World Bank have found a positive relationship between primary school attendance and the availability of local natural resources in a sample of Nepalese villages.

31. Landes (1998) begins his book with an account of the decline of the role of geography in the social sciences. He does not defend the decline; he notes it.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Adelman, Irma, and Cynthia Taft Morris. 1965. "A Factor Analysis of the Interrelationship between Social and Political Variables and Per Capita Gross National Product." *Quarterly Journal of Economics* 79(2): 555–78.

Agarwal, Anil, and Sunita Narain. 1989. *Towards Green Villages: A Strategy for Environmentally Sound and Participatory Rural Development*. New Delhi: Centre for Science and Development.

Aggarwal, Rimjhim, Sinaia Netanyahu, and Claudia Romano. 2001. "Access to Natural Resources and the Fertility Decision of Women: The Case of South Africa." *Environment and Development Economics* 6(2): 209–36.

Anderies, J. M. 2003. "Economic Development, Demographics, and Renewable Resources: A Dynamical Systems Approach." *Environment and Development Economics* 8(2): 219–46.

Arnott, Richard, and Joseph E. Stiglitz. 1991. "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out or Peer Review?" *American Economic Review* 81(1): 179–90.

Arrow, Kenneth, Partha Dasgupta, and Karl-Goran Mäler. 2003a. "Evaluating Projects and Assessing Sustainable Development in Imperfect Economies." *Environmental and Resource Economics* 26(4): 647–85.

_____. 2003b. "The Genuine Saving Criterion and the Value of Population." *Economic Theory* 21(2): 217–25.

Arrow, Kenneth, Bert Bolin, Robert Costanza, Partha Dasgupta, Carl Folke, Crawford Holling, Bengt-Owe Jansson, Simon Levin, Karl-Goran Mäler, Charles Perrings, and David Pimentel. 1995. "Economic Growth, Carrying Capacity, and the Environment." *Science* 268(5210): 520–21.

Baland, Jean-Marie, and Jean-Philippe Platteau. 1996. *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* Oxford, U.K.: Oxford University Press.

Batliwala, Srilatha, and Amulya K. N. Reddy. 1994. "Energy Consumption and Population." In Francis Graham-Smith, ed., *Population: The Complex Reality*. London: The Royal Society.

Bledsoe, Caroline. 1990. "The Politics of Children: Fosterage and the Social Management of Fertility among the Mende of Sierra Leone." In W. Penn Handwerker, ed., *Births and Power: Social Change and the Politics of Reproduction*. London: Westview Press.

Bledsoe, Caroline, and G. Pison, eds. 1994. *Nuptiality in Sub-Saharan Africa: Contemporary Anthropological and Demographic Perspectives*. Oxford, U.K.: Clarendon Press.

Blume, Lawrence, and Stephen N. Durlauf. 2000. "The Interactions-Based Approach to Socioeconomic Behavior." In Stephen N. Durlauf and H. Peyton Young, eds., *Social Dynamics*. Cambridge, Mass.: MIT Press.

Bongaarts, John. 1980. "Does Malnutrition Affect Fecundity? A Summary of the Evidence." *Science* 208: 564–69.

Bongaarts, John, and Mead Cain. 1981. *Demographic Responses to Famine*. New York: Population Council.

Bongaarts, John, and Susan Cotts Watkins. 1996. "Social Interactions and Contemporary Fertility Transitions." *Population and Development Review* 22(4): 639–82.

Brander, James A., and M. Scott Taylor. 1998. "The Simple Economics of Easter Island: A Ricardo-Malthus Model of Renewable Resource Use." *American Economic Review* 88(1): 119–38.

Cain, Mead. 1981. "Risk and Insurance: Perspectives on Fertility and Agrarian Change in India and Bangladesh." *Population and Development Review* 7(3): 435–74.

_____. 1983. "Fertility as an Adjustment to Risk." *Population and Development Review* 9(4): 688–702.

Caldwell, John C. 1991. "The Soft Underbelly of Development: Demographic Transition in Conditions of Limited Economic Change." *Proceedings of the Annual Bank Conference on Development Economics 1990*. Washington, D.C.: World Bank.

Caldwell, John C., and Patricia Caldwell. 1990. "High Fertility in Sub-Saharan Africa." *Scientific American* 262(5): 82–89.

Cavendish, William. 2000. "Empirical Regularities in the Poverty-Environment Relationships of Rural Households: Evidence from Zimbabwe." *World Development* 28: 1979–2003.

Chopra, Kanchan, Gopal Kadekodi, and M. N. Murty. 1990. *Participatory Development: People and Common Property Resources*. New Delhi: Sage Publications.

Cleaver, Kevin M., and Gotz A. Schreiber. 1994. *Reversing the Spiral: The Population, Agriculture, and Environment Nexus in Sub-Saharan Africa*. Washington, D.C.: World Bank.

Coale, Ansley J. 1969. "The Decline of Fertility in Europe from the French Revolution to World War II." In Jere Behrman, L. Corsa, and Ronald Freedman, eds., *Fertility and Family Planning: A World View*. Ann Arbor, Mich.: University of Michigan Press.

Cochrane, Susan. 1979. *Fertility and Education: What Do We Really Know?* Baltimore, MD: The Johns Hopkins University Press.

_____. 1983. "Effects of Education and Urbanization on Fertility." In R. Bulatao and Ronald Lee, eds., *Determinants of Fertility in Developing Countries,* vol. II. New York: Academic Press.

Cohen, Barney. 1993. "Fertility Levels, Differentials and Trends." In Karen A. Foote, Kenneth H. Hill, and Linda G. Martin, eds., *Demographic Change in Sub-Saharan Africa*. Washington, D.C.: National Academy Press.

Collier, Paul, V. L. Elliott, H. D. vard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. Policy Research Report. Washington, D.C.: World Bank.

Collier, Paul, and James W. Gunning. 1999. "Explaining African Economic Performance." *Journal of Economic Literature* 37(1): 64–111.

Collins, S., and Barry Bosworth. 1996. "Economic Growth in East Asia: Accumulation Versus Assimilation." *Brookings Papers on Economic Activity* 2: 135–91.

Cropper, Maurine L., and C. Griffiths. 1994. "The Interaction of Population Growth and Environmental Quality." *American Economic Review* 84 (Papers & Proceedings): 250–54.

Dasgupta, Partha. 1982. *The Control of Resources*. Cambridge, Mass.: Harvard University Press.

_____. 1990. "Well-Being and the Extent of its Realization in Poor Countries." *Economic Journal* 100 (Supplement): 1–32.

_____. 1993. *An Inquiry into Well-Being and Destitution.* Oxford, U.K.: Clarendon Press.

_____. 1996. "The Economics of the Environment." *Environment and Development Economics* 1(4): 387–428.

_____. 1997. "Nutritional Status, the Capacity for Work, and Poverty Traps." *Journal of Econometrics* 77(1): 5–38.

_____. 1998a. "The Economics of Food." In John C. Waterlow, D. G. Armstrong, Leslie Fowden, and Ralph Riley, eds., *Feeding a World Population of More than Eight Billion People: A Challenge to Science.* New York: Oxford University Press.

_____. 1998b. "The Economics of Poverty in Poor Countries." *Scandinavian Journal of Economics* 100(1): 41–68.

_____. 1999. "Economic Progress and the Idea of Social Capital." In Partha Dasgupta and Ismail Serageldin, eds., *Social Capital: A Multifaceted Perspective.* Washington, D.C.: World Bank.

_____. 2000. "Population, Resources, and Poverty: An Exploration of Reproductive and Environmental Externalities." *Population and Development Review* 26(4): 643–89.

_____. 2001a. *Human Well-Being and the Natural Environment.* Oxford, U.K.: Oxford University Press.

_____. 2001b. "On Population and Resources: Reply." *Population and Development Review* 26(4): 748–54.

_____. 2003. "Population, Poverty, and the Natural Environment." In Karl-Goran Mäler and Jeffrey Vincent, eds., *Handbook of Environmental Economics,* vol. 1. Amsterdam: Elsevier Science.

Dasgupta, Partha, and Geoffrey M. Heal. 1979. *Economic Theory and Exhaustible Resources.* Cambridge, U.K.: Cambridge University Press.

Dasgupta, Partha, and Karl-Goran Mäler. 1991. "The Environment and Emerging Development Issues." *Proceedings of the Annual World Bank Conference on Development Economics 1990.* Washington, D.C.: World Bank.

_____. 2000. "Net National Product, Wealth, and Social Well-Being." *Environment and Development Economics* 5(1): 69–93.

Dasgupta, Partha, and Debraj Ray. 1986. "Inequality as a Determinant of Malnutrition and Unemployment, 1: Theory." *Economic Journal* 96(4): 1011–34.

_____. 1987. "Inequality as a Determinant of Malnutrition and Unemployment, 2: Policy." *Economic Journal* 97(1): 177–188.

Dasgupta, Partha, Simon A. Levin, and Jane Lubchenco. 2000. "Economic Pathways to Ecological Sustainability." *BioScience* 50(4): 339–45.

Debreu, Gerard. 1959. *Theory of Value.* New York: John Wiley.

Durham, William. 1979. *Scarcity and Survival in Central America: Ecological Origins of the Soccer War.* Stanford, Calif.: Stanford University Press.

Easterlin, Richard, Robert Pollak, and Michael Wachter. 1980. "Toward a More General Model of Fertility Determination: Endogenous Preferences and Natural Fertility." In Richard Easterlin, ed. *Population and Economic Change in Developing Countries.* Chicago: University of Chicago Press.

Easterly, William. 2002. *The Elusive Quest for Growth.* Cambridge, Mass.: MIT Press.

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

Eicher, C. K. 1999. "Institutions and the African Farmer." Distinguished Economist Lecture, CIMMYT (International Maize and Wheat Improvement Center), Mexico City.

Ezzell, C. 2000. "Care for a Dying Continent." *Scientific American* 282(5): 72–81.

Farooq, Ghazi, Ita I. Ekanem, and Sina Ojelade. 1987. "Family Size Preferences and Fertility in South-Western Nigeria." In Christine Oppong, ed., *Sex Roles, Population, and Development in West Africa*. London: James Currey.

Filmer, Deon, and Lant Pritchett. 2002. "Environmental Degradation and the Demand for Children: Searching for the Vicious Circle in Pakistan." *Environment and Development Economics* 7(1): 123–46.

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.

_____. 1999. "Catching up with the Economy." *American Economic Review* 89(1): 1–19.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, Mass.: MIT Press.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger. 1999. "Geography and Economic Development." *Proceedings of the Annual World Bank Conference on Development Economics 1998*. Washington, D.C.: World Bank.

Grossman, Gene M., and Alan B. Krueger. 1995. "Economic Growth and the Environment." *Quarterly Journal of Economics* 110(2): 353–77.

Guyer, Jane L. 1994. "Lineal Identities and Lateral Networks: The Logic of Polyandrous Motherhood." In Caroline Bledsoe and G. Pison, eds., *Nuptiality in Sub-Saharan Africa: Contemporary Anthropological and Demographic Perspectives*. Oxford, U.K.: Clarendon Press.

Hajnal, John. 1982. "Two Kinds of Preindustrial Household Formation Systems." *Population and Development Review* 8(3): 449–94.

Hamilton, K., and M. Clemens. 1999. "Genuine Savings in Developing Countries." *World Bank Economic Review* 13(4): 333–56.

Harford, Jon D. 1998. "The Ultimate Externality." *American Economic Review* 88(1): 260–65.

Harrison, G. A., and John C. Waterlow, eds. 1990. *Diet and Disease in Traditional and Developing Countries*. Cambridge, U.K.: Cambridge University Press.

Heyzer, N. 1996. *Gender, Population, and Environment in the Context of Deforestation: A Malaysian Case Study*. Geneva: United Nations Research Institute for Social Development.

Hill, Kenneth. 1992. "Fertility and Mortality Trends in the Developing World." *Ambio* 21(1): 79–83.

Homer-Dixon, T. E. 1994. "Environmental Scarcities and Violent Conflict: Evidence from Cases." *International Security* 19(1): 5–40.

_____. 1999. *Environment, Scarcity, and Violence*. Princeton, N.J. Princeton University Press.

Howe, J. 1986. *The Kuna Gathering: Contemporary Village Politics in Panama*. Austin, Texas: University of Texas Press.

International Food Policy Research Institute. 1995. *A 20/20 Vision for Food, Agriculture, and the Environment*. Washington, D.C.

Iyer, Sriya. 2002. *Demography and Religion in India*. New Delhi: Oxford University Press.

Jalan, J., and Martin Ravallion. 2003. "Household Income Dynamics in Rural China." World Bank, Washington, D.C. Processed.

James, W. F. T. and others. 1992. *Body Mass Index: An Objective Measure of Chronic Energy Deficiency in Adults*. Rome: Food and Agriculture Organization of the United Nations.

Jodha, Narpat S. 1986. "Common Property Resources and the Rural Poor." *Economic and Political Weekly* 21: 1169–81.

_____. 1995. "Common Property Resources and the Environmental Context: Role of Bio-physical Versus Social Stress." *Economic and Political Weekly* 30: 3278–83.

Johnson, D. Gale. 2000. "Population, Food, and Knowledge." *American Economic Review* 90(1): 1–14.

_____. 2001. "On Population and Resources: A Comment." *Population and Development Review* 27(4): 739–47.

Jolly, Carole L., and James N. Gribble. 1993. "The Proximate Determinants of Fertility." In Karen A. Foote, Kenneth H. Hill, and Linda G. Martin, eds., *Demographic Change in Sub-Saharan Africa*. Washington, D.C.: National Academy Press.

Jones, Charles I. 1998. *Introduction to Economic Growth*. New York: W. W. Norton.

Koopmans, Tjalling C. 1957. "The Price System and the Allocation of Resources." In Tjalling C. Koopmans, *Three Essays on the State of Economic Science*. New York: McGraw Hill.

Kremer, Michael. 1993. "Population Growth and Technological Change: One Million B.C. to 1990." *Quarterly Journal of Economics* 108(3): 681–716.

Krishnan, Pramila. 2001. "Cultural Norms, Social Interactions, and the Fertility Transition in India." University of Cambridge, Faculty of Economics, Cambridge, U.K. Processed.

Landes, David. 1969. *The Unbound Prometheus*. Cambridge, U.K.: Cambridge University Press.

_____. 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. New York: W. W. Norton.

Levin, Simon A. 1999. *Fragile Dominion: Complexity and the Commons*. Reading, Mass.: Addison Wesley Longman.

Loughran, D., and Lant Pritchett. 1998. "Environmental Scarcity, Resource Collection, and the Demand for Children in Nepal." World Bank, Washington, D.C. Processed.

Maddison, Angus. 2001. *The World Economy: A Millennial Perspective*. Paris: Organisation for Economic Co-operation and Development, Development Research Centre.

Murray, James D. 1993. *Mathematical Biology*. Berlin: Springer-Verlag.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, U.K.: Cambridge University Press.

_____. 1992. *Crafting Institutions for Self-Governing Irrigation Systems*. San Francisco: ISC Press for the Institute for Contemporary Studies.

Pearce, David, and Giles Atkinson. 1995. "Measuring Sustainable Development." In Daniel W. Bromley, ed., *Handbook of Environmental Economics*. Oxford: Basil Blackwell.

Platteau, Jean-Philippe, and Y. Hayami. 1998. "Resource Endowments and Agricultural Development: Africa Versus Asia." In Y. Hayami and Masahico Aoki, eds., *The Institutional Foundations of East Asian Economic Development*. London: MacMillan.

Pritchett, Lant. 1997. "Divergence, Big Time." *Journal of Economic Perspectives* 11(1): 3–17.

Scrimshaw, Nevin C., and others. 1968. *Interactions of Nutrition and Infection*. Geneva: World Health Organization.

Sen, Amartya. 1994. "Population: Delusion and Reality." *New York Review of Books,* September 22, pp. 62–71.

_____. 1999. *Development as Freedom*. Oxford, U.K.: Oxford University Press.

Spurr, George B. 1988. "Marginal Malnutrition in Childhood: Implications for Adult Work Capacity and Productivity." In K. J. Collins and D. F. Roberts, eds., *Capacity for Work in the Tropics*. Cambridge, U.K.: Cambridge University Press.

_____. 1990. "The Impact of Chronic Undernutrition on Physical Work Capacity and Daily Energy Expenditure." In G. A. Harrison and J. C. Waterlow, eds., *Diet and Disease in Traditional and Developing Countries.* Cambridge, U.K.: Cambridge University Press.

Thomson, J. T., D. H. Feeny, and R. J. Oakerson. 1986. "Institutional Dynamics: The Evolution and Dissolution of Common Property Resource Management." In National Research Council, ed., *Proceedings of a Conference on Common Property Resource Management.* Washington, D.C.: U.S. National Academy of Science Press.

Turner, B. L., and A. M. S. Ali. 1996. "Induced Intensification: Agricultural Change in Bangladesh with Implications for Malthus and Boserup." *Proceedings of the National Academy of Sciences* 93: 14984–91.

United Nations Development Programme. 1990. *Human Development Report.* New York: Oxford University Press.

_____. 1994. *Human Development Report.* New York: Oxford University Press.

_____. 2000. *Human Development Report.* New York: Oxford University Press.

Wade, Robert. 1988. *Village Republics: Economic Conditions for Collective Action in South India.* Cambridge, U.K.: Cambridge University Press.

Waterlow, John C. (with contributions by A. M. Tomkins and S. M. Grantham-McGregor). 1992. *Protein Energy Malnutrition.* Sevenoaks, Kent, U.K.: Edward Arnold.

Watkins, Susan C. 1990. "From Local to National Communities: The Transformation of Demographic Regions in Western Europe 1870–1960." *Population and Development Review* 16(2): 241–72.

World Bank. 1992. *World Development Report 1992: Development and the Environment.* New York: Oxford University Press.

_____. 1998. *World Development Indicators.* Washington, D.C.

_____. 2000a. *World Development Indicators.* Washington, D.C.

_____. 2000b. *World Development Report, 2000/2001: Attacking Poverty.* New York: Oxford University Press.

_____. 2001. *World Development Indicators.* Washington, D.C.

World Health Organization. 1983. *Measuring Change in Nutritional Status.* Geneva.

_____. 1985. *Energy and Protein Requirements.* Technical Report Series 724. Geneva.

Wrigley, E. Anthony, and Roger S. Schofield. 1981. *The Population History of England 1541–1871: A Reconstruction.* Cambridge, U.K.: Edward Arnold.

Development Strategy: Transition and Challenges of Development in Lagging Regions

JUSTIN YIFU LIN AND MINGXING LIU

Following World War II, many developing countries adopted various measures to industrialize their economies, hoping to accelerate the process of catching up with the industrial countries. However, except for a small number of economies in East Asia, the gap between the developing countries and the industrial countries has increased. Recent studies find that the poor development performance in the developing countries is largely attributable to their problems in relation to institutions, including all kinds of market distortions and government interventions. Most developing countries, both socialist and nonsocialist, started to reform their institutions in the early 1980s. Nevertheless, the growth performance of most developing countries has deteriorated further.

We argue that the poor growth performance and many institutional distortions in the developing countries following World War II can be largely explained by their adoption of an inappropriate development strategy. Motivated by nation building, most developing countries, including the socialist countries, adopted a comparative advantage defying (CAD) strategy to accelerate the growth of capital-intensive, advanced sectors in their countries. Many firms in those priority sectors were non-viable in open, competitive markets because of the violation of their economies' comparative advantages. To implement a CAD strategy, governments in the developing countries adopted a series of distortions in input and output markets to subsidize and/or protect the nonviable firms or both, resulting in rent-seeking, soft-budget constraints, macroeconomic instability, and income disparities. Economic stagnation, or even sudden collapse, became unavoidable, prompting the developing countries, voluntarily or involuntarily, to embark on market-oriented reforms. The implementation of market-oriented reforms without first addressing the issue of firms' viability

might result in widespread bankruptcies, unemployment, and social and political instability. For fear of such consequences, many governments found other disguised ways to protect and subsidize nonviable firms after implementing reforms. As a result, not only could they not achieve the transition to a well-functioning market economy, but also their economic performance became poorer than that before the reforms.

When reform starts, governments should liberalize entry into the labor-intensive sectors, which were previously repressed, and address the viability issue of those firms in sectors targeted by the CAD strategy. In this way, the economy can achieve dynamic growth during its transition.

This paper tests several hypotheses about the impacts of development strategy and the transition approach on economic performance using macroeconomic data from 103 countries for 1962–99. The results are consistent with the hypotheses: a CAD strategy is detrimental to long-term growth and causes economic growth to be volatile, and the growth of labor-intensive sectors is conducive to overall economic growth during the transition.

Introduction

Following World War II, many governments in the developing countries adopted various policy measures to industrialize their economies. Per capita gross domestic product (GDP), measured in comparable prices, has increased in almost all nations since World War II; however, as figure 1 shows, the gap in per capita GDP between the developed regions and those lagging behind widened substantially between 1950 and 1992. Only a small number of economies in East Asia have actually narrowed the gap and converged to the level of per capita income in the industrial countries.

The rapid increase in wealth and national power in the industrial countries after the industrial revolution in the 18th century was rooted in their speedy upgrading of industries and technologies. Therefore the early development attempts in the developing countries focused primarily on how to accelerate the development and/or adoption of the dominant industries and technologies of the industrial countries (Krueger 1992).

Recent studies find that the poor development performance of the developing countries is largely attributable to their problems in relation to institutions, including market distortions, government interventions, macroeconomic instability, inequality in income distribution, colonial heritage, and so forth, which impeded the functioning of markets (Acemoglu 2002; Acemoglu and Robinson 2002; Acemoglu, Johnson, and Robinson 2001a, b; Djankov and others 2003; Rodrik 1998, 2003; Shleifer and Vishny 1998). To improve their economic performance, since the late 1970s many developing countries, including the formerly socialist countries, have embarked on market-oriented reforms. Nevertheless, with few exceptions, such as China and Vietnam, the growth performance of many developing countries has deteriorated further despite various measures to improve factors that are considered to be determinants of growth (Barro 1997). Easterly (2001, p. 2) notes that "in 1980–98,

FIGURE 1.
GDP Per Capita, 56 Selected Countries, 1950–92

1900 Geary-Khamis dollars

Legend: Western Europe | Western offshoots | Southern Europe | Eastern Europe | Latin America | Asia | Africa
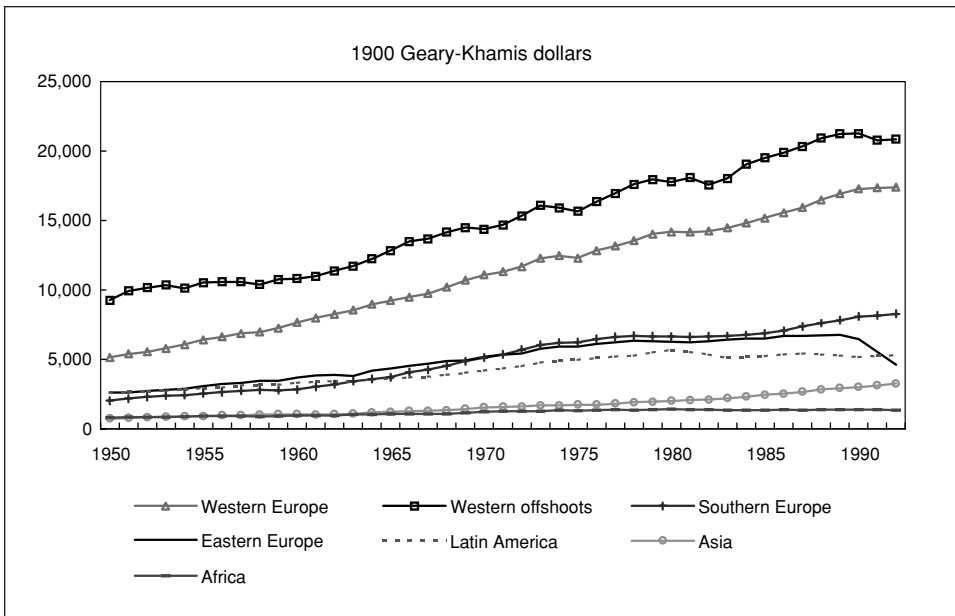
Source: Maddison (1995).

Western offshoots refer to Australia, Canada, New Zealand, the United States, and other countries whose populations are descendants of Western European countries.

median per capita income growth in developing countries was 0.0 percent, as compared to 2.5 percent in 1960–79."

We argue that many distortions and interventions in the markets of poorly performing developing countries were endogenous to their governments' attempts to pursue a development strategy that was CAD, capital intensive and oriented toward heavy industries following World War II. This strategy caused many firms in the governments' priority sectors to be nonviable in open, competitive markets. Consequently, the governments adopted a series of price distortions and administrative interventions in the markets to support and protect the nonviable firms. As a result of such measures, developing countries were able to establish advanced sectors and initially saw investment-led growth. However, their economies became inefficient because of the misallocation of resources, the distortion of incentives, the presence of rampant rent-seeking, and so forth. When they had depleted their domestic resources and exhausted options for further resource mobilization from international sources to support the nonviable firms, their economies stagnated, and in some cases were hit by crises.

Reforms were essential for revitalizing the developing countries' economies. However, many distortions and interventions in the markets were endogenous to the need to protect the nonviable firms in the priority sectors. If the reforms did not start out by solving the firms' viability problems, either the nonviable firms collapsed immediately,

resulting in widespread unemployment and social and economic disruptions, or the government needed to protect and subsidize those nonviable firms through various disguised means that entailed distortion and protection. Therefore not only could reforms toward a fully fledged market economy not be completed, but also the countries' economic performance could become poorer than it had been before the reforms had started. This paper suggests that governments should adopt an approach that liberalizes entry into the labor-intensive sectors, which were previously repressed, and create conditions to address the viability issue of firms in the priority sectors targeted by the CAD strategy, so that their economies can achieve stable, dynamic growth during the course of transition.

The paper is organized as follows: The next section reviews the literature on the determinants of economic growth and discusses the impacts of the development strategy of a developing country government on the viability of firms and economic institutions and the strategy's economic and social consequences. The following section explores the different approaches to economic reform and transition in the developing countries and their consequences. The paper then empirically tests hypotheses derived from the previous two sections. The paper finishes with concluding remarks.

Determinants of Economic Growth and the Impact of Development Strategy on Institutions and Growth Performance in Developing Countries

Why the developing countries cannot catch up with the industrial countries has been a challenging question and puzzling phenomenon to economists. The neoclassical growth theory (Solow 1956), with its assumption of exogenously given technology, predicts that the developing countries would grow faster than the industrial countries, that per capita income in the developing countries would converge to the level of that in the industrial countries, and that the GDP growth rate in any country would eventually be the same as its population growth rate. However, with a few exceptions in East Asia, most developing countries' per capita income failed to converge to the level in the industrial countries (Romer 1994), and the economic growth rates in the industrial countries continue to exceed their population growth rates.

Unsatisfied with the neoclassical growth theory, Romer (1986) and Lucas (1988) pioneered the new growth theory, which treats technological innovation as endogenously determined by the accumulation of human capital, research and development, learning-by-doing, and so on, and argue that the failure of the developing countries to converge with the industrial countries is due to their lack of investment in those factors that are important for technological innovations. This argument is insightful in relation to the continuous growth of per capita income in the industrial countries. However, the new growth theory fails to provide a satisfactory explanation for the extraordinary growth and convergence of the newly industrialized economies in East Asia, including Hong Kong (China), the Republic of Korea, Singapore, Taiwan (China), and recently China, during the last three decades of the 20th century

(Grossman and Helpman 1994; Pack 1994). During the catching-up process, these newly industrialized economies' investments in research and development, human capital, and learning-by-doing were much lower than those of the industrial countries.

Many economists now believe that the developing countries failed to catch up with the industrial countries because of bad institutions caused by government intervention and regulations, including widespread corruption, weak protection for investors, and extensive social conflicts. As Rodrik (2003, p. 7) states:

> Institutions have received increasing attention in the growth literature as it has become clear that property rights, appropriate regulatory structure, quality and independence of the judiciary, and bureaucratic capacity could not be taken for granted in many settings and that they were of utmost important to initiating and sustaining economic growth.

Economists have emphasized the important impact on economic development of a country's legal origin (La Porta and others 1998, 1999) and institutional inheritance (Acemoglu, Johnson, and Robinson 2001a,b; Engerman and Sokoloff 1997).

Generally speaking, the government is the most important institution in a developing country. Its economic policies shape the incentive structure that firms in the country face. Both policy reformers and researchers have tried to understand how government intervention and regulation occur and how and whether they can subsequently be sustained (Rodrik 1996). The classical theory of the role of government (Pigou 1938) has been called the helping hand view. An alternative strand, the grabbing hand view (Shleifer and Vishny 1998), holds that the government pursues interventions for the benefit of politicians and bureaucrats. Politicians use regulation to favor friendly firms and other political constituencies, and in exchange obtain campaign contributions and votes. In addition, "[A]n important reason why many of these permits and regulations exist is probably to give officials the power to deny them and to collect bribes in return for providing the permits" (Shleifer and Vishny 1993, p. 601). Djankov and others (2002) provide an empirical test of the theory of the grabbing hand to ascertain whether the barrier to business entry might arise from corruption among bureaucrats.

Supposing that government regulations in the developing countries were a result of the grabbing hand of government or of political elites, the unsolved question in the literature is how to understand the evolution of institutional structure under government intervention. In the developing countries, the institutional structure shaped by government intervention is extremely complicated. What are the incentives for political leaders to design such complicated systems, given that the increase in the costs of expropriation and political control resulting from the complexity of institutions diminishes the gains of grabbing? Corruption induced by special interest groups might not be a good answer to this question either, because the groups that would benefit are often taxed or suppressed in tandem with the protection and/or subsidies. Moreover, many interventions do not have obvious beneficiary groups.

 Lin (2003) proposes an alternative hypothesis for government intervention and regulations in a developing country based on conflicts between the government's development strategy and the country's endowment structure (the discussion in this

section draws heavily on Lin 2003). Many of the early generation of political leaders in both socialist and nonsocialist developing countries, such as Mao Zedong in China, Nasser in Egypt, Nehru in India, Sukarno in Indonesia, and Ho Chi Minh in Vietnam, were elites participating in independent movements or revolutions for the purpose of nation building. According to Lin (2003), the institutions created by this early generation of political leaders were endogenously shaped by the conflicts between the elites' ambitious drives to industrialize and modernize and their nations' economic realities. The key to the argument is the viability of firms in the sectors accorded priority by the government during the course of industrialization.[1]

Lin (2003, p. 280) defines the term viability as follows: "If, without any external subsidies or protections, a *normally managed* firm is expected to earn socially acceptable profits in a free, open, and competitive market, then the firm is viable. Otherwise the firm is nonviable." Obviously no one will invest in a firm if they do not expect it to earn a socially acceptable, normal profit. Such a firm will exist only if the government gives it support. According to Lin, the viability of a normally managed firm depends on whether the industry the firm enters and the technology the firm adopts in relation to production are consistent with the comparative advantages determined by the economy's endowment structure. Other things being equal, in an economy characterized by a relative abundance of labor (or capital), a firm will be viable only if it enters a relatively labor-intensive (capital-intensive) industry and uses relatively labor-intensive (capital-intensive) technology for production.

In a free, open, competitive market without government intervention, only viable firms will exist, therefore the structure of industry and technology in the economy will be endogenously determined by the economy's endowment structure. Most developing countries are characterized by a relative abundance of labor and scarcity of capital.[2]

As such, in a free, open, and competitive market, the structure of industry and technology in a developing country will be relatively labor intensive. However, inspired by the dream of nation building, developing country political leaders, economists, and social elites often strive to develop capital-intensive industries and to adopt advanced technologies similar to those in the most advanced industrial countries within the shortest period of time. Lin refers to this type of approach toward development in a developing country as the CAD strategy, because the government attempts to encourage firms to ignore the country's existing comparative advantages in relation to their entry and choice of industry or technology.[3] Most firms in sectors given priority by a CAD strategy are not viable in free, open, and competitive markets. Therefore the developing country government has to subsidize and protect these firms by means of various interventions.

If a government adopts a CAD strategy and the deviation of firms' choices of technology and industry from the optimal ones determined by the economy's endowment structure is small, and if the number of nonviable firms that the government attempts to support is limited, the government may subsidize the firms directly by means of tax transfers as in the case of agricultural protection in many countries of the Organisation for Economic Co-operation and Development. However, when a developing

country government adopts a CAD strategy, the extent of deviation is often large, the number of nonviable firms tends to be numerous, and the government's taxation capacity is frequently extremely weak. Therefore the government often turns toward implicit subsidy measures through price distortions, limitations on market competition, direct administrative allocation of resources, and so on.[4] Indeed, traditional planning systems that existed before economic transitions in the socialist economies were the typical institutional arrangements for supporting and protecting nonviable heavy industrial firms (Lin, Cai, and Li 2003, chapter 2).

Moreover when a developing country government adopts a CAD strategy, it cannot know exactly how large sufficient subsidies will be because of information asymmetry. Firms in the priority sectors will have incentives to use their viability problem as an excuse and to use resources to lobby government officials not only for more ex ante policy favors, such as access to low interest rate loans, tax reductions, tariff protection, and legal monopolies, but also for ex post administrative assistance, such as being given access to greater numbers of preferential loans or being allowed to run tax arrears. The economy will be rife with rent-seeking activities and corruption. Because the firms can use the viability problem as an excuse to bargain for more government support, and because it is hard for the government to shun such responsibility, the firms' budget constraints become soft (Lin and Tan 1999).[5] When a soft budget constraint exists, firms face no pressure to improve productivity and their efficiency will be low. Moreover, given the subsidies and protection and the soft budget constraints applicable to firms in the priority sectors, investing in those sectors becomes a privilege. Thus political leaders in a nonsocialist developing country may select their close friends or political supporters to invest in those priority sectors, resulting in the phenomenon of crony capitalism.

In addition, if the government in a developing country adopts a CAD strategy, the economy will become more inward-oriented than if the government adopts a comparative advantage following strategy (CAF). This is because a CAD strategy attempts to substitute for the importation of capital-intensive manufactured goods by domestic production, causing a reduction in imports. Exports will also be suppressed because of the inevitable transfer of resources away from the industries for which the economy has comparative advantages toward the priority sectors targeted by the CAD strategy; plus, exchange rates are likely to be overvalued to facilitate the importation of technology and equipment for the priority industries, effectively hampering export opportunities. In addition, under a CAD strategy, the government's development strategy is normally focused on large firms. To support the financial needs of nonviable, large firms, the government often nationalizes the firms and uses direct fiscal appropriation, skipping financial intermediation, to support them. Such was the case in the former socialist planned economies and continues to be the case in many developing countries.

Even if the government relies on private firms to carry out its CAD strategy, the financial needs of large firms will be significant and will only be met by a heavily regulated, oligopolistic banking system or a stock market in which the government has administratively intervened. In either case, the country's financial system will

be inefficient. The development of the nonviable firms relies heavily on external financial support. The government first mobilizes domestic resources to support these firms through interventions in the financial system. Once it has depleted domestic financial resources, the government often turns to international financial markets to support further development of these firms. This exacerbates fiscal deficits, bad loans, external debts, and financial fragility, resulting in macroeconomic instability, and possibly leading to financial crises, which may also trigger serious social conflicts and political instability (Caselli and Coleman 2002; Rodrik 1998).

Thus to summarize, according to Lin (2003), the poor growth performance and observed distortions and interventions in developing countries are largely attributable to their governments' adoption of a CAD strategy.

Viability and Economic Reform and Transition

Empirical evidence shows that the adoption of a CAD strategy is detrimental to a developing country's growth (Lin 2003). However, a CAD strategy, with its resulting government intervention, is initially good at mobilizing scarce resources for investing in a few clear, well-defined, priority sectors (Ericson 1991). The countries that adopt a CAD strategy can also enjoy a period of investment-led growth so long as they can mobilize resources administratively from domestic or international sources for investing in the priority sectors. Therefore because political leaders in developing countries lacked knowledge about the long-term consequences of a CAD strategy, had aspirations for quick nation building, or had concerns about short-term performance during their tenure in office, they found a CAD strategy attractive, and almost all governments in the developing countries adopted such a strategy after World War II (Chenery 1961). However, once financial resources from domestic and international sources had been depleted, the economies stagnated and the problems inherent in a CAD strategy started to appear.[6] The economies encountered all kinds of difficulties, and since the late 1970s, voluntarily or involuntarily, market-oriented reforms have become an unavoidable choice for the developing countries, both socialist and nonsocialist alike (Krueger 1992).

When they embarked on reform, most developing countries focused their attention on distortions and government interventions and attempted to establish institutions that were considered essential for markets to function efficiently (Kolodko 2000). However, except for China, Vietnam, and a few other countries, the growth performance of socialist economies during their transition has been miserable. The disappointing performance of transition in the former Soviet Union and Eastern Europe (FSUEE) is especially striking. When the transition started in the FSUEE in the late 1980s and early 1990s, most economists were optimistic about its expected outcomes based on the countries' adoption of "shock therapy," which attempted to restore market institutions as soon as possible.[7] Ten years have elapsed since the transition started; however, contrary to the early optimism, the FSUEE countries that implemented shock therapy experienced a prolonged

period of rampant inflation, output collapses, sharp widening of inequality, and worsening of other social indicators (Blejer and Skreb 2001; Roland 2000; World Bank 2002a). The cumulative output declines were much more serious in all countries of the Commonwealth of Independent States and in most countries in Central and Southern Europe and the Baltics than the decline in the United States during the Great Depression (World Bank 2002a).

The transition to a market economy was disappointing not only in the FSUEE, but also in other developing countries. The poor performance is puzzling because, as Easterly's (2001) study shows, variables that are considered important in growth regressions (such as financial deepening, trade and exchange rate liberalization, health, education, fertility, and infrastructure) generally improved compared with the situation before the transition and the adoption of reform in 1960–79. Easterly speculates that global factors (like the increase in world interest rates, the increased debt burden of developing countries, the growth slowdown in the industrial world, and the skill-biased technological changes) contributed to the developing countries' stagnation in the lost decades of the 1980s and the 1990s. However, Easterly's hypothesis is inconsistent with China's remarkable annual GDP growth rate of 9.6 percent during 1980–2000 and Vietnam's of 6.5 percent during 1985–2000.

As argued in the previous section, many distortions and government regulations in the socialist economies and developing countries that are considered detrimental to economic performance from the viewpoint of neoclassical economics are endogenous to the viability problem of firms in the priority sectors targeted by a CAD strategy. Without first appropriately addressing the viability problems of firms in the priority sectors, eliminating or liberalizing the endogenous distortions or regulations may result in changing institutions from second-best to third-best, causing economic performance to deteriorate after the reform. This is because if all the distortions and regulations are removed, the viability problems of firms in the priority sectors will turn from implicit to explicit. The nonviable firms will become bankrupt immediately if they do not receive any subsidy or protection. If the number of nonviable firms and the number of workers employed are both small and the political determination for giving up the CAD strategy is strong, shock therapy can succeed. The elimination of distortions and protection may cause a few firms to become bankrupt, but viable firms in previously depressed sectors may grow rapidly after liberalization and overcompensate for the loss of outputs and employment resulting from the bankruptcy of the nonviable firms. However, if the number of nonviable firms and the number of workers employed are large, the forceful elimination of the distortions and protection will lead to widespread unemployment, resulting in economic collapse instead of recovery, as occurred in the FSUEE and recently in Indonesia. Social and political stability would be difficult to maintain. To prevent such consequences, after the initial attempt to forcefully implement reforms, the government often finds other ways to subsidize or protect nonviable firms, resulting in incomplete reform and a worse economic performance than that before the transition. This is why some FSUEE countries have failed, and the outcomes deserve further analysis.

When they embark on reforms, most developing countries are saddled with a large number of nonviable firms set up under their previous CAD strategy. The direct implementation of many policy reforms based on existing neoclassical economics, which implicitly assumes that firms are viable, may not be appropriate (Lin 2002). Governments need to find a way to revive a country's economic dynamism while solving the viability issues of firms in the priority sectors and allowing the economy to gradually move toward a well-functioning market system. China's experience may provide a useful example for other economies already undergoing transition or about to undertake it.

China started its transition in 1979 with a piecemeal, gradual, dual-track approach. At the start of reforms, the Chinese government gave partial autonomy to managers of state-owned enterprises (SOEs) and de-collectivized farms to improve incentives, but the government still provided protection and support to nonviable SOEs in the traditional sectors to buffer them from the threat of bankruptcy. The improvement in incentives resulted in productivity increases in both the agriculture and industrial sectors (Lin 1992; World Bank 1992). At the same time, the government relaxed its strict control of entry into sectors that were consistent with China's comparative advantages and had been depressed under the previous CAD strategy, resulting in the rapid growth of labor-intensive, small and medium firms not owned by the state, such as township and village enterprises (TVEs), joint ventures, and private enterprises. The rapid development of TVEs is illustrative. During 1978–96, the number of TVEs increased from 1.52 million to 23.36 million, and the number of workers employed by TVEs increased from 28.27 million to 135.08 million, or from 9.5 to 29.8 percent of the total rural labor force. Equally remarkably, TVEs became one of the major forces behind China's overall sustained growth. The industrial output value of the TVEs increased from 9.1 percent of the national total in 1978 to 57.9 percent in 1997. Rural industry is no longer merely a supplement to agricultural production, but has become an indispensable source of growth nationwide. It is widely acknowledged that exports have been one of the leading factors contributing to China's recent success, and TVEs' share of exports increased from 9.2 percent of total exports in 1986 to 45.8 percent in 1997 (Lin and Yao 2001).

Vietnam is another country that has achieved dynamic growth after its transition. As in China, Vietnam started its transition by de-collectivizing agriculture, enhancing SOEs' autonomy, and promoting small and medium enterprises not owned by the state in sectors that had previously been repressed. SOEs maintained their dominant role in the industrial sector.[8] Through this cautious and gradual approach, the growth rate of average annual GDP reached 6.5 percent in 1985–2000.

In the FSUEE, after the initial transition recession, recovery also came mainly from the entry of small and medium enterprises into the previously repressed labor-intensive sectors. In 1998, new small enterprises that employed fewer than 50 workers contributed about 50 percent of employment in the leading reformers, such as the Czech Republic, Hungary, Latvia, Lithuania, and Poland, whereas in the poorly performing countries, such as Belarus, Kazakhstan, Russia, and Ukraine, the employment share of small enterprises was only between 10 and 20 percent (World Bank 2002a).

The entry of small and medium firms into sectors that had been repressed under the previous CAD strategy allows China and a few other economies to enjoy dynamic growth during the course of transition. However, a country's reform and transition to a market economy will not be complete until the viability issue of firms in the priority sectors under a CAD strategy has been solved; otherwise the government will need to maintain its interventions in markets to protect and/or subsidize the nonviable firms, with the resulting inevitable distortions. For example, alongside the rapid economic growth during China's transition, the share of nonperforming loans looms large and corruption is widespread (Lardy 1998). These problems have their roots in the viability problem of the SOEs. After 1983, the approach the Chinese government adopted to support the SOEs changed from direct fiscal appropriation to offering low interest rate loans from the state-owned commercial banks. Currently SOEs account for more than 70 percent of bank loans, but because of their poor performance, many SOEs have been unable to repay the loans, and thus the banks have accumulated large numbers of nonperforming loans. To support the SOEs, the government also limits market entry to certain sectors so that the SOEs can enjoy monopolistic rents. Many SOEs (and non-SOEs) lobby the government to acquire more loans at low interest rates or licenses for market entry to the regulated sectors, thereby adding fuel to the widespread corruption.

Whether or not SOEs can earn acceptable profits in a competitive market becomes their managers' own business only after the viability problems have been solved.[9] The government will no longer need to find ways to intervene in the markets to protect or subsidize the firms. Only then can the elimination of distortions and government interventions be carried out successfully. However, whether the government will follow the required policy reforms for a well-functioning market system also depends on whether the government has the wisdom and determination to give up its CAD strategy.[10]

Development Strategy, Transition, and Growth Performance: Empirical Testing

The discussion in the previous sections gives rise to several testable hypotheses about economic performance in developing countries before and after economic reform in the 1980s. This section uses cross-country, time-series data to carry out empirical tests on the following hypotheses:

- *Hypothesis I*. Over an extended period of time, growth performance in a country that adopts a CAD strategy will be poor.

- *Hypothesis II*. Over an extended period of time, the economy in a country that adopts a CAD strategy will be volatile.

- *Hypothesis III*. In the transition to a market economy from an economy that had followed a CAD strategy, the greater the entry of labor-intensive, small and medium enterprises, the better the overall economic growth.

To test these hypotheses we use macroeconomic data from 103 countries for 1962–99. The definitions of the variables, their means and standard deviations, and the sources of the data are reported in table 1.

To test the foregoing hypotheses, we need to have a proxy for a country's development strategy. We construct a simple statistical indicator, the technology choice index (TCI), of the manufacturing sector in a country, which is defined as follows:

$$TCI_{it} = \frac{AVM_{it}/GDP_{it}}{LM_{it}/L_{it}}, \tag{1}$$

where AVM_{it} is the added value of manufacturing industries of the ith country in year t, GDP_{it} is the GDP of the ith country in year t, LM_{it} is the labor in the manufacturing industry, and L_{it} is the total labor of ith country in year t.[11]

Based on the earlier discussion, the TCI can be a proxy for a government's adoption of a CAD strategy. If a government adopts a CAD strategy to promote its capital-intensive industries, the country's TCI is expected to be larger than it would otherwise have been. This is because a country that adopts a CAD strategy for its manufacturing industries will be more capital intensive and absorb less labor, other things being equal. Therefore the denominator in equation (1) will be smaller for a country that adopts a CAD strategy. Meanwhile, to overcome the viability issue, a government that adopts a CAD strategy may provide subsidized credit and price-suppressed inputs to firms in the priority sectors and at the same time give the firms a monopoly position so that they can charge higher prices for their products. These policy measures will result in a larger AVM than would otherwise be the case. Therefore, the nominator in equation (1) will be larger for a country that adopts a CAD strategy. As such, given income levels and other conditions, the magnitude of the TCI can be used as a proxy to the extent that a country follows a CAD strategy. The appendix reports the mean and standard deviation in the TCI for each of the 103 countries during 1962–99.

In testing the hypotheses, we include initial per capita GDP to control for the effect of the stage of development and initial population size to control for the effect of market size. We also include an indicator of the rule of law to reflect institutional quality; a trade dependency ratio to reflect a country's openness; the population size to measure market size; and variables to reflect geographic characteristics, one being the distance from the equator and the other whether or not a country is landlocked. To control for the endogeneity of institutional quality to income, we use as instruments the share of the population that speaks English and the share that speaks a major European language, which are intended to capture the long-run impacts of colonial origin on current institutional quality. Similarly, we also use the fitted values of trade predicted by a gravity model as the instrument for openness. This approach was first proposed by Frankel and Romer (1999) and revised by Dollar and Kraay (2003). In the regressions that use panel data, we instrument openness by the one-period lagged value of openness.

TABLE 1.
Definitions and Sources of Variables

Variable	Definition	Mean	Standard deviation	Source
LnGDP60	Log of real GDP per capita in 1960	7.33	0.80	World Bank (2002b)
LnGDP80	Log of real GDP per capita in 1980	7.91	1.05	World Bank (2002b)
LnGDP	Log of real GDP per capita in 1960, 1970, 1980, 1990	7.73	1.02	World Bank (2002b)
LnTCI1	Log of the average technology choice index from 1963 to 1999	0.96	0.90	World Bank (2002b); UNIDO (2002)
LnTCI2	Log of the average technology choice index per decade in the 1960s, 1970s, 1980s, 1990s	0.85	0.84	World Bank (2002b); UNIDO (2002)
LnTCI70	Log of the average technology choice index from 1970 to 1979; if not available, then log of the average technology choice index from 1980 to 1985	0.91	0.92	World Bank (2002b); UNIDO (2002)
DELTCI	Log of the average technology choice index from 1999 to 1990 minus LnTCI70	0.07	0.38	World Bank (2002b); UNIDO (2002)
RL01	Rule of law in 2000–01	0.003	0.95	Kaufmann, Kraay, and Zoido-Lobatón (2002)
LnOPEN1	Log of the average (exports + imports)/GDP from 1960 to 1999	−1.11	0.81	Dollar and Kraay (2003)
LnOPEN2	Log of the decadal average (exports + imports)/GDP in the 1960s, 1970s, 1980s, 1990s	−1.30	0.84	Dollar and Kraay (2003)
LnPOP1	Log of the total midyear population from 1960 to 1999	15.2	2.11	World Bank (2002b)
LnPOP2	Log of the total initial year population in the 1960s, 1970s, 1980s, 1990s	14.93	2.12	World Bank (2002b)
LANDLOCK	Dummy variable taking a value of 1 if a is country is landlocked, 0 otherwise	0.18	0.39	Dollar and Kraay (2003)
LnDIST	Log (DISTEQ+1), where DISTEQ is the distance from the equator, measured as the absolute value of the latitude of the capital city	2.96	0.88	Dollar and Kraay (2003)
ENGFRAC	Fraction of the population speaking English	0.07	0.24	Hall and Jones (1999), taken from Dollar and Kraay (2003)

(continued)

TABLE 1.
Definitions and Sources of Variables (*continued*)

Variable	Definition	Mean	Standard deviation	Source
EURFRAC	Fraction of the population speaking a major European language	0.22	0.38	Hall and Jones (1999), taken from Dollar and Kraay (2003)
LnFRINST	Instrument variable for LnOPEN	−2.83	0.64	Dollar and Kraay (2003)
INST	Predicted value of RL01 in the cross-section estimation (using ENGFRAC and EURFRAC as the instruments)	0.003	0.34	Authors
TRADE1	Predicted value of LnOPEN1 in the cross-section estimation (using LnFRINST as the instrument)	−1.11	0.38	Authors
TRADE2	Predicted value of LnOPEN2 in the panel data estimation (using the lagged value as the instrument)	−1.27	0.79	Authors

Testing of Hypothesis I

We use two constructed data sets to test hypothesis I. The first data set consists of cross-sectional data where the dependent variable is the average annual growth rate of per capita GDP for the period 1962–99, and the second data set consists of panel data where the growth rate is constructed as the average annual growth rate of per capita GDP during the 1960s, 1970s, 1980s, and 1990s.

Table 2 reports the estimates from the first data set. Regression models 1.1 and 1.2 use the ordinary least squares (OLS) approach to obtain the estimates. The explanatory variables in model 1.1 include only the proxy for the development strategy, LnTCI1, and the initial GDP per capita, LnGDP60, whereas model 1.2 includes other explanatory variables that capture institutional quality, openness, geographic location, and market size. Model 1.3 has the same explanatory variables, but the model uses the two-stage least squares (2SLS) approach in which INST and TRADE1 are predicted values for RL01 and LnOPEN1 from the first-stage least squares. The instrument variables for INST are EMGFRAC and EURFRAC, whereas the instrument variable for TRADE1 is LnFRINST. The purpose of using the 2SLS approach is to control for the endogeneity of institutional quality and openness.

The results show that the TCI has the expected negative effect and is highly significant in all three regressions. This finding supports hypothesis I that the further a country pursued a CAD strategy, the worse its growth performance during 1962–99. The estimated coefficients of LnTCI1 range from –0.66 to –1.25. From the estimates, we can infer that a 10 percent increase in the TCI from the mean may result in

TABLE 2.
Estimates for Hypothesis 1, First Data Set
(dependent variable is the average annual growth rate of per capita GDP for the period 1962–99)

Variable	Model 1.1 (OLS)	Model 1.2 (OLS)	Model 1.3 (2SLS)
Constant	7.32***	4.66**	3.26
	(1.60)	(1.87)	(2.15)
LnTCI1	−1.25***	−.66***	−.92***
	(.20)	(.18)	(.19)
LnGDP60	−.54***	−.99***	−.59***
	(.20)	(.18)	(.21)
RL01		.58***	
		(.21)	
INST			.22
			(.41)
LnOPEN1		.70***	
		(.22)	
TRADE1			.93**
			(.43)
LnDIST	.	20	.47***
		(.16)	(.16)
LnPOP1		.33***	.22**
		(.09)	(.09)
LANDLOCK		.07	.46
		(.32)	(.38)
Adjusted-R^2	.36	.56	.44
Number of observations	85	83	83

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: The figures in parentheses are standard errors.

Source: Authors' calculations.

approximately a 0.1 percentage point reduction in the country's average annual growth rate of per capita GDP for the entire period 1962–99.

The regression results also show that initial per capita income and population size have the expected signs and significant effects on the growth rate. The rule of law, openness, and distance to the equator also have the expected signs. However, the rule of law is not significant in the 2SLS regression and the distance to the equator is not significant in the OLS regression. Whether or not a country is landlocked is not significant in all three regressions.

Table 3 reports the results from the second data set, where the dependent variable is the average annual growth rate of per capita GDP during the 1960s, 1970s, 1980s, and 1990s. The regression approaches to fit the estimates are respectively OLS for models 2.1 and 2.2, one-way fixed effect for model 2.3, 2SLS for model 2.4, and

2SLS and one-way fixed effect for model 2.5. In the fixed effect models, time dummies are added to control for the time effects, whereas the 2SLS models are used to control for the endogeneity of institutional quality and openness. The instrumental variables for INST are again ENGFRAC and EURFRAC and the instrumental variable for TRADE2 is the lagged value LnOPEN2.

As for the first data set, the estimates for the TCI have the expected negative sign and are highly significant in all the regressions. The finding is once again consistent with the prediction of hypothesis I that a country's development strategy is a prime determinant of its long-run economic growth performance. The results for other explanatory variables are similar to those in table 2.

TABLE 3.
Estimates for Hypothesis 1, Second Data Set
(dependent variable is the average annual growth rate of per capita GDP during the 1960s, 1970s, 1980s, and 1990s)

Variable	Model 2.1 (OLS)	Model 2.2 (OLS)	Model 2.3 (fixed effect)	Model 2.4 (2SLS)	Model 2.5 (2SLS, fixed effect)
Constant	7.15***	8.36***	3.83*	−.74	−2.70
	(1.61)	(2.16)	(2.11)	(2.56)	(2.37)
LnTCI2	−1.10***	−.69***	−.40**	−.69***	−.47**
	(.21)	(.20)	(.19)	(.24)	(.22)
LnGDP	−.54***	−1.39***	−.86***	−.17	.17
	(.18)	(.23)	(.23)	(.27)	(.25)
RL01		1.45***	1.12***		
		(.23)	(.22)		
INST				−.38	−.67*
				(.42)	(.38)
LnOPEN2		.24	.35		
		(.23)	(.22)		
TRADE2				.01	−.06
				(.29)	(.27)
LnDIST		−.04	−.10	.27	.17
		(.18)	(.17)	(.20)	(.18)
LnPOP2		.32***	.41***	.22*	.27**
		(.10)	(.09)	(.12)	(.12)
LANDLOCK		−.31	.08	−.23	.02
		(.39)	(.36)	(.46)	(.43)
Adjusted-R^2	.08	.23	.36	.08	.24
Number of observations	315	278	278	213	213

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: The figures in parentheses are standard errors. Models 2.3 and 2.5 include a time dummy.

Source: Authors' calculations.

Testing of Hypothesis II

Hypothesis II pertains to the effects of a CAD strategy on the volatility of a country's economic growth rate. If a country follows a CAD strategy, it may see a period of investment-led growth, but the growth is not sustainable and an economic crisis is likely. Therefore a country that follows a CAD strategy is likely to be more volatile than would otherwise be the case. In our empirical testing of this hypothesis we use the following to measure the volatility of a country's per capita GDP growth rate during 1962–99:

$$
V_i = \left[(1/38) \sum_{t=1962}^{T=1999} \left(\frac{g_{it}}{\left(\sum_{t=1962}^{T=1999} g_{it} \right) \Big/ 38} - 1 \right)^2 \right],
\tag{2}
$$

where g_{it} is the growth rate of GDP per capita of ith country in year t.

In testing hypothesis II, the dependent variable is the log of the measurement of volatility, V_i, from equation (2), and the explanatory variables are the same as those used to test hypothesis I. The approaches to fitting the regression equation are also similar to those used previously. Table 4 reports the results from fitting the regression models.

As expected, the estimates of the TCI are positive and highly significant in all three regressions. The results support hypothesis II and indicate that the more deeply a country followed a CAD strategy, the more volatile the economic growth rate in the country. From the estimates we can infer that a 10 percent increase in the TCI may cause volatility to increase about 4 to 6 percent.

The estimates for other explanatory variables show that quality of institutions, degree of openness, whether or not a country is landlocked, and population size all have negative effects on economic volatility. However, except for population size, which is a proxy for the size of an economy and whose coefficients are significant in both the OLS and 2SLS models, the estimated coefficients for other variables are insignificant in either the OLS model or the 2SLS model. The estimates for initial per capita income in 1960 and distance to the equator are not significant in all three regressions.

Testing of Hypothesis III

Hypothesis III attempts to provide a partial explanation to a puzzling phenomenon identified by Easterly (2001). Most developing countries started their reforms in the 1980s; however, the median growth rate of per capita GDP in the developing countries in the 1980s and 1990s was 0 percent, whereas the median growth rate in the 1960s and 1970s was 2.5 percent. As argued earlier, the development of labor-intensive sectors, in which the developing countries had comparative advantages, was repressed, and their adoption of a CAD strategy led to many

TABLE 4.
Estimates for Hypothesis II
(dependent variable is the log of the volatility of the growth rate of per capita income in 1962–99)

Variable	Model 3.1 (OLS)	Model 3.2 (OLS)	Model 3.3 (2SLS)
Constant	.49	3.03**	3.63**
	(1.06)	(1.44)	(1.56)
LnTCI1	.64***	.41***	.56***
	(.13)	(.14)	(.14)
LnGPP60	−.04	.17	−.07
	(.13)	(.14)	(.15)
RL01		−.33**	
		(.16)	
INST			−.20
			(.29)
LnOPEN1		−.46***	
		(.17)	
TRADE1			−.53
			(.33)
LnDIST		−.003	−.15
		(.11)	(.11)
LANDLOCK		−.31	−.53*
		(.24)	(.28)
LnPOP1		−.26***	−.18**
		(.06)	(.07)
Adjusted-R^2	.29	.47	.37
Number of observations	103	93	93

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: The figures in parentheses are standard errors.

Source: Authors' calculations.

distortions and interventions before they embarked on reforms. Therefore the growth performance of a developing country depends on its ability to create an enabling environment for the development of labor-intensive sectors and at the same time find a way to solve the viability issue for firms inherited from the previous development strategy. A CAD strategy is associated with a high TCI. If, following reform, a developing country successfully develops its labor-intensive sectors, its resource allocation and growth performance will improve and the TCI will decline. Therefore a successful transition from a CAD strategy is expected to result in a negative change in the TCI. The larger the negative change, the higher the expected growth rate.

We create a variable, DELTCI, the difference between the log of the average TCI during 1990–99 and the log of the average TCI during 1970–79, to measure the growth of labor-intensive sectors after transition.

The dependent variable in the regressions is the log of the average annual growth rate of per capita GDP during 1980–99. In addition to DELTCI, the explanatory variables include the log of the average TCI in the 1970s; the initial per capita GDP in 1980; and the other explanatory variables representing institutional quality, openness, and population size, which are similar to those used in the previous regressions.

TABLE 5.
Estimates for Hypothesis III
(dependent variable is the average growth rate of per capita GDP during 1980–99)

Variable	Model 4.1 (OLS)	Model 4.2 (OLS)	Model 4.3 (2SLS)	Model 4.4 (OLS)	Model 4.5 (OLS)	Model 4.6 (2SLS)
Constant	2.53	3.79	−2.94	4.28	-4.50	−9.03
	(3.17)	(3.63)	(3.97)	(4.24)	(5.01)	(6.43)
DELTCI	−1.25**	−.91**	−1.12**	−1.16*	−1.02*	−1.30**
	(.55)	(.45)	(.51)	(.66)	(.52)	(.60)
LnTCI70	−.84**	−.38	−.52	−.61	−.26	−.31
	(.41)	(.34)	(.38)	(.48)	(.38)	(.45)
LnGDP80	−.04	−1.32***	−.31	−.34	−.78*	−.12
	(.35)	(.37)	(.38)	(.50)	(.45)	(.57)
RL01		1.31***			1.78***	
		(.37)			(.47)	
INST			.44			.96
		(.60)				(1.18)
LnOPEN1		.71*			.54	
		(.36)			(.49)	
TRADE1			1.50**			2.23*
			(.70)			(1.26)
LnDIST		.16	.57*		−.06	.34
		(.28)	(.29)		(.33)	(.36)
LnPOP1		.52***	.44***		.79***	.78**
		(.17)	(.16)		(.19)	(.29)
LANDLOCK		−.87	−.06		−.55	.54
		(.57)	(.68)		(.73)	(1.15)
Adjusted-R^2	.13	.43	.27	.03	.45	.24
Number of observations	76	72	72	50	49	49

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: The figures in parentheses are standard errors. The data samples in the regressions of models 4.4–4.6 only include the developing countries defined by Easterly and Sewadeh (2002).

Source: Authors' calculations.

We use two data sets to test the hypothesis. The first one includes observations from all countries in the dataset, whereas the second one includes only the developing countries as defined by Easterly and Sewadeh (2002). We carry out three regressions for each data set, two OLS regressions and one 2SLS regression, as in the testing of previous hypotheses. Table 5 reports the results of the regressions.

As expected, the sign of DELTCI is negative and the estimates are significantly different from zero in all six regressions. The results support the hypothesis that the reduction in the TCI from its level in the 1970s to its level in the 1990s had a positive effect on the average per capita GDP growth rate during 1980–99. Therefore for a country that adopted a CAD strategy, its growth performance will improve if the government allows labor-intensive sectors to grow. From the estimates, we can infer that a 10 percent reduction in the TCI level from the 1970s to the 1990s may have resulted in a 0.10 to 0.13 percentage point increase in the average annual growth rate of per capita GDP during 1980–99.

The other explanatory variables all have the expected signs. However, except for population size, which is positive and highly significant in all six regressions, the other variables are either not significant at all or significant in some regressions but not in others.

Concluding Remarks

Most developing countries are confronting two challenges: how to close their development gaps with the industrial countries and how to complete their transition from the existing highly regulated and distorted systems to well-functioning market systems. A continuous upgrading of technology and industry is essential for sustained growth in an economy. Potentially, a developing country could benefit from the gap between its level of technology and industry and that in the industrial countries and achieve a convergence of development to the industrial countries. However, most developing countries, socialist and nonsocialist alike, adopted a CAD strategy, which caused all kinds of distortions and did not bring about a convergence. Eventually economic reform became an unavoidable choice for a developing country that had adopted a CAD strategy. Because the distortions and poor performance before the reform were endogenous to the CAD strategy, a government's commitment to replacing its old CAD strategy with a new CAF strategy is essential.

The empirical results from a panel dataset covering 103 countries for the period 1962–99 are consistent with the main arguments of this paper: the adoption of a CAD strategy is detrimental to a developing country's economic growth and stability, and a developing country undertaking reform will grow faster if the government creates conditions for the growth of previously suppressed labor-intensive sectors.

Appendix: Means and Standard Deviations of the TCI and Growth Rate of GDP Per Capita, Selected Economies and Years

Country	TCI (1963–99)		Growth rate of GDP per capita (%) (1962–99)	
	Mean	Standard deviation	Mean	Standard deviation
Algeria	2.157	0.979	1.108	8.789
Argentina	2.564	0.588	1.147	5.567
Australia	1.073	0.162	2.205	2.088
Austria	1.083	0.071	2.864	1.879
Bahamas	1.929	0.845	1.543	7.757
Bangladesh	4.302	0.902	1.384	4.564
Barbados	1.283	0.521	2.716	4.784
Belgium	1.017	0.122	2.709	2.059
Belize	1.067	0.072	2.816	3.989
Bolivia	7.341	2.905	0.437	3.753
Botswana	1.791	0.801	6.586	5.177
Brazil	5.373	1.195	2.644	4.263
Bulgaria	1.372	0.089	0.214	5.603
Burkina Faso	38.845	0.078	1.177	3.325
Burundi	44.402	4.318	0.445	6.483
Cameroon	7.018	1.626	0.798	6.443
Canada	1.531	0.199	2.215	2.168
Central African Republic	9.830	2.221	−0.698	4.032
Chile	4.307	1.223	2.655	5.219
China	4.165	1.327	5.338	7.419
Colombia	4.466	0.701	1.968	2.147
Congo, Rep.	3.353	0.809	1.581	6.313
Costa Rica	2.190	0.683	1.711	3.427
Côte d'Ivoire	6.370	0.499	1.084	5.525
Croatia	1.581	0.637	−1.380	12.185
Cyprus	1.308	0.310	5.756	4.896
Denmark	1.178	0.079	2.442	2.241
Dominican Republic	2.532	0.368	2.767	5.516
Ecuador	3.878	1.238	1.781	4.888
Egypt, Arab Rep.	2.012	0.238	3.177	3.084
El Salvador	4.229	1.569	0.801	4.233
Ethiopia	17.921	2.621	0.463	8.455
Finland	1.237	0.116	2.896	3.248
France	1.106	0.096	2.585	1.751
Gabon	2.352	0.351	3.189	10.953
Gambia, The	7.214	2.361	0.578	3.360
Ghana	5.962	2.075	−0.186	4.389
Greece	1.337	0.087	3.460	3.553
Guatemala	3.303	0.279	1.360	2.634
Honduras	3.183	0.790	0.841	3.194
Hong Kong, China	0.713	0.071	5.671	4.156
Hungary	1.151	0.183	3.222	4.532

Appendix: Means and Standard Deviations of the TCI and Growth Rate of GDP Per Capita, Selected Economies and Years (*continued*)

Country	TCI (1963–99)		Growth rate of GDP per capita (%) (1962–99)	
	Mean	Standard deviation	Mean	Standard deviation
Iceland	0.802	0.134	2.963	3.959
India	3.635	0.421	2.431	3.238
Indonesia	3.073	0.408	4.106	3.060
Iran, Islamic Rep.	1.750	0.326	−0.528	7.842
Iraq	1.646	0.577	−2.633	18.672
Ireland	1.853	0.507	4.141	2.858
Israel	1.287	0.232	3.121	3.662
Italy	1.292	0.134	2.907	2.263
Jamaica	3.248	0.621	0.614	4.636
Japan	1.680	0.083	4.577	3.599
Jordan	1.936	0.492	2.134	8.071
Kenya	0.335	0.030	1.519	5.094
Korea, Rep.	2.816	0.493	6.428	3.303
Kuwait	1.090	0.477	−3.114	11.198
Latvia	1.638	0.010	2.063	8.162
Lesotho	8.719	2.037	3.427	7.023
Luxembourg	0.914	0.101	3.029	3.201
Macao, China	0.384	0.060	2.799	3.962
Madagascar	5.373	0.498	−1.128	3.388
Malawi	8.631	2.923	1.502	5.549
Malaysia	1.854	0.191	4.135	2.952
Malta	1.143	0.091	5.866	4.117
Mauritius	1.121	0.447	3.609	6.368
Mexico	2.926	0.257	2.083	3.494
Mongolia	3.697	0.860	−0.009	5.134
Morocco	3.201	0.383	1.804	4.771
Nepal	4.174	0.342	1.123	2.972
Netherlands	1.158	0.204	2.377	1.912
New Zealand	1.061	0.188	1.328	2.979
Nigeria	9.338	6.549	0.650	7.819
Norway	0.914	0.072	3.149	1.714
Oman	1.036	0.151	7.381	17.523
Pakistan	6.114	1.221	2.855	2.429
Papua New Guinea	7.250	1.541	1.526	4.639
Peru	5.128	1.162	0.896	5.346
Philippines	4.571	1.143	1.312	3.174
Poland	1.704	0.327	1.651	5.091
Portugal	1.265	0.257	4.015	3.779
Puerto Rico	3.814	0.718	3.507	3.135
Romania	1.086	0.046	0.698	6.058
Russian Federation	0.999	0.108	2.590	9.405
Rwanda	33.545	15.987	0.087	10.208

Appendix: Means and Standard Deviations of the TCI and Growth Rate of GDP Per Capita, Selected Economies and Years (*continued*)

Country	TCI (1963–99)		Growth rate of GDP per capita (%) (1962–99)	
	Mean	Standard deviation	Mean	Standard deviation
Senegal	11.296	2.968	−0.287	4.359
Sierra Leone	9.914	4.146	−0.769	8.142
Singapore	1.406	0.203	6.561	3.842
Slovenia	1.071	0.112	1.324	5.325
South Africa	1.853	0.162	0.976	3.789
Spain	1.267	0.199	3.418	2.877
Sri Lanka	2.728	0.341	2.821	1.782
Suriname	2.409	0.532	0.433	6.480
Swaziland	3.817	0.733	2.263	4.437
Sweden	1.206	0.124	2.066	2.054
Syrian Arab Republic	2.058	0.755	2.890	8.519
Tanzania	3.233	0.370	0.430	5.053
Thailand	7.201	2.613	5.072	2.809
Togo	10.466	1.003	1.322	6.702
Trinidad and Tobago	1.475	0.446	2.538	5.604
Tunisia	2.891	1.243	3.146	3.835
Turkey	4.586	0.968	2.318	3.575
United Arab Emirates	0.365	0.013	−2.727	8.734
United Kingdom	1.358	0.154	2.019	1.924
United States	1.588	0.108	2.014	2.018
Uruguay	2.036	0.430	1.217	4.177
Venezuela, R.B. de	2.826	0.843	−0.179	4.147
Zambia	5.909	1.694	−0.953	4.689
Zimbabwe	5.118	1.358	1.233	5.585

Source: Authors.

Notes

1. Bureaucrats in lower levels of government in a developing country may use intervention and regulations rooted in the nation building attempt for their personal grabbing hand purpose; however, the grabbing hand of bureaucrats should be viewed as a consequence rather than a cause of the distortions and regulations.

2. The other possibility is a relative abundance of natural resources and relatively scarce capital and labor. The discussions and conclusions in this paper can easily be extended to cover this case.

3. The CAD strategy includes the development strategy oriented toward heavy industry in the socialist countries and in such developing countries as India, and the secondary import substitution strategies of many African and Latin America countries. The CAD strategy also includes the protection of certain industries that have lost their comparative advantage

because of the development of the economy, such as the protection of agriculture in many countries of the Organisation for Economic Co-operation and Development.

4. From this perspective, the root of interventions in a developing country is not the grabbing hands of government officials or manipulation by interest groups, but the dream of nation building among political elites. Corruption may be an endogenous phenomenon of the distortions and interventions arising from the conflict between the economy's endowment structure and the political leaders' ambitious and unrealistic development attempts. In this case the political target should be separated from the corruption view of the grabbing hand approach.

5. The soft budget constraint is a term Kornai (1986) coined to explain the problem in the socialist countries. According to Kornai, the soft budget constraint arises from the paternalistic nature of the socialist government toward the state-owned firm. His argument cannot explain why the soft budget constraint exists in nonsocialist economies and why the soft budget constraint still exists 10 years after privatization in Russia and the Eastern European transition economies (World Bank 2002a). Dewatripont and Maskin (1995) argue that the soft budget constraint arises from banks' imperfect information about investment projects and the time inconsistence problem of projects. However, this argument cannot explain the prevalence and persistence of the soft budget constraint phenomenon in the socialist and in many developing countries, where investment decisions are made by the governments and the banks are owned by the states.

6. How long a country can sustain a CAD strategy depends on how rich the natural resources per capita in the country are (Ranis and Syed 1992). In addition, the length of time may also depend on the size of the country's population. In the early 1950s, East Asian economies such as Korea, Singapore, and Taiwan (China) also followed a CAD strategy; however, because of their poor natural resource endowments and small population sizes, their economies immediately encountered huge fiscal deficits, high inflation, and external imbalances. They were therefore forced to give up their CAD strategy. As a result, comparative advantage following became the de facto strategy, which may explain the successful development experiences of these economies.

7. The key elements of shock therapy include price liberalization, privatization, and macroeconomic stabilization.

8. Indeed, in Vietnam the share of SOEs in GDP increased from 33 percent in 1989 to 39 percent in 1996 (Sun 1997).

9. Lin, Cai, and Li (1998) recommend four reform measures to address the issue of an SOE's viability. First, if the SOE's output is essential for national defense, the government should use fiscal appropriation continuously to support the operations of the firm. Second, if the SOE's output has a large domestic market, the SOE can get access to external capital either by forming a joint venture with a multinational company or by being listed in the international equity market. Third, if the SOE's output does not have a large domestic market but the SOE has good engineering and management capacity, the SOE can rely on its strength in human capital and shift its operations to products that are consistent with the nation's comparative advantage and that have a large domestic market. Fourth, if the SOE's output does not have a large domestic market and the SOE does not have engineering and managerial strength either, the SOE should be allowed to go bankrupt; however, the number of firms in the last category tends to be small. As long as the economy can maintain dynamic growth, the economy should be able to create enough jobs to absorb workers released from the bankrupt firms and enough resources to compensate for losers in the transition process.

10. The traditional heavy industries may not be attractive to the developing countries any more; however, in many societies the desire for heavy industries has been replaced by a desire for information, biotechnology, and other high-tech industries. If a developing country

government wants to accelerate the growth of these new industries in its economy, firms in these industries will be nonviable and will require government subsidies and protection as was case with the traditional CAD strategy, and the institutional reforms needed for a well-functioning market economy cannot be completed successfully.

11. In Lin (2003), TCI is defined as $TCI = (K_m/L_m)/(K/L)$, where K_m/L_m is the capital-labor ratio of the manufacturing sector and K/L is the capital-labor ratio of the whole economy.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Acemoglu, Daron. 2002. "Why Not a Political Coase Theorem? Social Conflict, Commitment and Politics." Working Paper no. 9377. National Bureau of Economic Research, Cambridge, Mass.

Acemoglu, Daron, and James A. Robinson. 2002. "Economic Backwardness in Political Perspective." Working Paper no. 8831. National Bureau of Economic Research, Cambridge, Mass.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001a. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369–1401.

_____. 2001b. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." Working Paper no. 8460. National Bureau of Economic Research, Cambridge, Mass.

Barro, Robert J. 1997. *Determinants of Economic Growth: A Cross-Country Empirical Study*. Cambridge, Mass.: MIT Press.

Blejer, Mario I., and Marko Skreb, eds. 2001. *Transition: The First Decade*. Cambridge, Mass.: MIT Press.

Caselli, Francesco, and Wilbur John Coleman II. 2002. "On the Theory of Ethnic Conflict." Harvard University, Cambridge, Mass. Processed.

Chenery, Hollis B. 1961. "Comparative Advantage and Development Policy." *American Economic Review* 51(1): 18–51.

Dewatripont, Mathias, and Eric Maskin. 1995. "Credit and Efficiency in Centralized and Decentralized Economies." *Review of Economic Studies* 62(4): 541–56.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2002. "Regulation of Entry." *Quarterly Journal of Economics* 117(1): 1–37.

Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "The New Comparative Economics." *Journal of Comparative Economics* 31(4): 595–619.

Dollar, David, and Aart Kraay. 2003. "Institutions, Trade, and Growth." *Journal of Monetary Economics* 50(1): 133–62.

Easterly, William. 2001. "The Lost Decades: Explaining Developing Countries' Stagnation in Spite of Policy Reform 1980–1998." *Journal of Economic Growth* 6(2): 135–57.

Easterly, William, and Mirvat Sewadeh. 2002. *Global Development Network Growth Database*. Washington, D.C.: World Bank. http://www.worldbank.org/research/growth/GDNdata.htm.

Engerman, Stanley, and Kenneth Sokoloff. 1997. "Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economics: A View from Economic Historians of the United States." In Stephen Haber, ed., *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico 1800–1914*. Stanford, Calif.: Stanford University Press.

Ericson, Richard E. 1991. "The Classical Soviet-Type Economy: Nature of the System and Implications for Reform." *Journal of Economic Perspectives* 5(4): 11–27.

Frankel, Jeffrey, and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(June): 379-99.

Grossman, Gene M., and Elhanan Helpman. 1994. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives* 8(Winter): 23–44.

Hall, Robert E., and Charles Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics* 114(1): 83–116.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 2002. "Governance Matters II: Updated Indicators for 2000/01. Policy Research Department Working Paper no. 2772. World Bank, Washington, D.C.

Kolodko, Grzegorz W. 2000. *From Shock to Therapy: The Political Economy of Post-Socialist Transformation*. New York: Oxford University Press.

Kornai, Janos. 1986. "The Soft Budget Constraint." *Kyklos* 39(1): 3–30.

Krueger, A. O. 1992. *Economic Policy Reform in Developing Countries*. Oxford, U.K.: Basil Blackwell.

La Porta, Rafael, Florencio Lopez-di-Silanes, Andrei Shleifer, and Robert Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106(December): 1113–55.

_____. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization* 15(1): 222–79.

Lardy, Nicholas R. 1998. *China's Unfinished Economic Revolution*. Washington, D.C.: Brookings Institution Press.

Lin, Justin Yifu. 1992. "Rural Reforms and Agricultural Growth in China." *American Economic Review* 82(March): 34–51.

_____. 2002. "Viability, Economic Transition, and Reflections on Neoclassical Economics." Working Paper no. E2002007. Peking University, China Center for Economic Research.

_____. 2003. "Development Strategy, Viability, and Economic Convergence." *Economic Development and Cultural Change* 53(2): 277–308.

Lin, Justin Yifu, and Guofu Tan. 1999. "Policy Burdens, Accountability, and the Soft Budget Constraint." *American Economic Review: Papers and Proceedings* 89(2): 426–31.

Lin, Justin Yifu, and Yang Yao. 2001. "Chinese Rural Industrialization in the Context of the East Asian Miracle." In Joseph E. Stigilitz and Shahid Yusuf, eds., *Rethinking the East Asian Miracle*. Oxford, U.K., and New York: Oxford University Press.

Lin, Justin Yifu, Fang Cai, and Zhou Li. 1998. "Competition, Policy Burdens, and State-Owned Enterprise Reform." *American Economic Review Papers and Proceedings* 88(2): 422–27.

_____. 2003. *China's Miracle: Development Strategy and Economic Reform*, 2nd ed. (English-language version). Hong Kong, China: Chinese University of Hong Kong Press. Chinese-language version published in 1994 by the Shanghai Sanlian Press, Shanghai.

Lucas, Robert E. 1988. "On the Mechanism of Economic Development." *Journal of Monetary Economics* 22(March): 3–42.

Maddison, Angus. 1995 *Monitoring the World Economy, 1820–1992*. Paris: Organisation for Economic Co-operation and Development.

Pack, Howard. 1994. "Endogenous Growth Theory: Intellectual Appeal and Empirical Short-comings." *Journal of Economic Perspectives* 8(Winter): 55–72.

Pigou, Arthur C. 1938. *The Economics of Welfare*, 4th ed. London: Macmillan.

Ranis, Gustav, and Mahmood Syed. 1992. *The Political Economy of Development Policy Change.* Cambridge, Mass.: Blackwell.

Rodrik, Dani. 1996. "Understanding Economic Policy Reform." *Journal of Economic Literature* 34(March): 9–41.

_____. 1998. "Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapse." Harvard University, Cambridge, Mass. Processed.

_____. 2003. "Institution, Integration, and Geography: In Search of the Deep Determinants of Economic Growth." In D. Rodrik, ed., *In Search of Prosperity: Analytic Country Studies on Growth*. Princeton, N.J.: Princeton University Press.

Roland, Gerard. 2000. *Transition and Economics: Politics, Markets, and Firms.* Cambridge, Mass.: MIT Press.

Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94(5): 1002–37.

_____. 1994. "The Origins of Endogenous Growth." *Journal of Economic Perspectives* 5(Winter): 3–22.

Shleifer, A., and R. Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108(2): 599–617.

_____. 1998. *The Grabbing Hand: Government Pathologies and their Cures*. Cambridge, Mass.: Harvard University Press.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70(February): 65–94.

Sun, Laixing. 1997. *Emergence of Unorthodox Ownership and Governance Structure in East Asia: An Alternative Transition Path.* Helsinki: United Nations University, World Institute for Development Economics Research.

UNIDO (United Nations Industrial Development Organization). 2002. *Industrial Statistics Database 3-Digit Level of ISIC Code 1963–1999, 2002.* CD-Rom. Vienna.

World Bank. 1992. *Reform and the Role of Planning in the 1990s.* Washington, D.C.

_____. 2002a. *Transition: The First Ten Years, Analysis and Lessons for Eastern Europe and the Former Soviet Union.* Washington, D.C.

_____. 2002b. *World Development Indicators.* Washington, D.C.

Comment on "World Poverty: Causes and Pathways" by Partha Dasgupta and "Development Strategy: Transition and Challenges of Development in Lagging Regions" by Justin Yifu Lin and Mingxing Liu

M. GOVINDA RAO

The development of lagging regions has been a matter of considerable concern to both policymakers and development economists. The persistence of wide differences in standards of living, whether between countries or between different regions within a country, is undesirable, and strategies to redress such differences have preoccupied economists and policymakers.

The developing countries need to catch up with the industrial countries, because the vast majority of the poor live in the developing countries and need to be pulled out of the poverty trap. Balanced regional development within a country is considered important not only as a way to banish poverty by taking opportunities to the people rather than moving people in search of opportunities, but also to ensure proper exploitation of the developmental potential of different regions. Redressing regional imbalance is considered necessary also to avoid social and political tensions. However, despite considerable research and understanding gained, the development of lagging regions continues to be a major challenge.

The two papers presented on the subject offer significant insights into our understanding of the growth process and the causes of divergence in different countries. Both the papers deal with why the developing regions have failed to catch up with the industrial countries in terms of well-being, despite predictions to the contrary by neoclassical growth theory. Also the two papers complement each other well. Partha Dasgupta brings in the missing element—the role of natural resources in ensuring human well-being—and adds a third dimension, local ecology, to the analysis of growth processes to the prevailing two, policies and institutions. By contrast, Justin Yifu Lin and Mingxing Liu take us back in history to review the growth strategies the developing countries adopted, and argue that their adoption of a comparative advantage defying (CAD) rather than a comparative advantage following (CAF)

strategy has been the principal reason for the divergence. In terms of policies, while the first paper points to the importance of social safety nets, of the development of human capital, and of measures to protect local ecology, the second paper argues for the careful calibration of policies to reorient them toward a CAF strategy.

Both the papers are extremely important, as they add significant insights into the process of growth and poverty reduction in developing countries. Nevertheless, gaps persist in our understanding of why some countries and regions within countries have lagged in exploiting their developmental potential.

Comments on "World Poverty: Causes and Pathways"

In his insightful paper Dasgupta emphasizes the missing element: the role of natural resources in ensuring human well-being. He argues, "Policies matter, as do institutions, but the local ecology matters too." At the level of the individual, the paper seeks to understand the link between nutritional status and work capacity or labor productivity, and at the local and regional levels, it attempts to unravel the mysteries surrounding ecological and socioeconomic pathways and reproductive and environmental externalities. It examines the nexus between poverty, population growth, and degradation of the local natural resource base and concludes that over time, each of the foregoing influences and is influenced by the others.

The first part of the paper deals mainly with the measurement of an inclusive index of wealth. The paper argues that the sum of consumption and investment in physical and human capital indicated by gross national product is an inappropriate measure of well-being, because it does not take depreciation into account and, more important, does not account for the depletion of natural resources. The adoption of an inclusive measure of well-being also has significant implications for economic analysis. The total factor productivity measured from the residual, for example, is an overestimate, because it includes the depletion of natural resources (which is negative). Dasgupta estimates investment in natural resources based on shadow prices and adds them to the Hamilton-Clemens estimate of recorded investment to arrive at an estimate of inclusive investment. Assuming the output-wealth ratio to be 0.15, he computes the value of the annual growth rate in per capita wealth. The results show that during 1973–93, Sub-Saharan Africa, as well the South Asian countries, experienced wealth de-cumulation.

The second part of the paper tries to explain this phenomenon in terms of the relationship between nutritional status and capacity to work on the one hand, and the relationship between population growth, poverty, and degradation of natural resources on the other. The paper also examines a number of microeconomic issues, such as female education and fertility, structure of property rights, incentive structure in relation to local commons, and cost of childbearing in families. The important conclusions are that a significant portion of the labor force is simply shut out of the market, because at market wages some people cannot obtain the nutrition they need to perform at their potential. Equally important is the institutional failure (including market and government failure) caused by human–nature interactions. The existence

of thresholds caused by nonconvexities results in a situation wherein small differences in local ecology can cause large differences in outcomes.

While the paper is not explicit about remedial policies, some of them are implicit. There cannot be any quarrel about the need to consider inclusive wealth as a measure of well-being. Natural capital does not have market prices, and imputing shadow prices for natural capital assumes immense importance in the valuation of wealth. As regards public policies, the role of the government now extends to some new areas of market failure not recommended by traditional welfare economics. This includes ensuring that the poor have access to adequate nutrition, potable water, sanitation, and health care to realize their potential earning power. Equally important is investment in human capital. Government intervention in providing these is necessary because of the externalities inherent in them.

This paper is an important contribution and will improve our understanding of the dynamics of the growth process. The paper also shows how little we actually know about the dynamics of growth, which affects our ability to recommend policies. The addition of considerations about local ecology to policies and institutions in relation to devising reform strategies to improve well-being is important and necessary. However, we still know little about the nature and consequences of interactions between the three sets of factors, namely, policies, institutions, and local ecology. Future research will have to unravel the mysteries of these interactions.

The policy implications that follow from the paper can help to accelerate development only partially. Most socialist countries ensured nutritional security and access to health care for their populations, yet the countries failed to perform to their potential because of other factors (the theme of the Lin and Liu paper). Thus providing nutritional security, health care, and education by themselves will not ensure convergence.

Another important issue is the valuation of natural capital. Natural resources are not traded in the market, and determining their prices is not going to be easy. The imputation of shadow prices involves judgments and can be arbitrary. Much work needs to be done to ensure the standardization of shadow prices. Indeed, with a different set of shadow prices and assumptions, the conclusion that wealth decumulated during 1970–93 could be reversed.

How much does the new theory help us understand differences in the performance of different regions within a country? The explanations about the causes of and pathways to poverty have some generality, but there are other institutional and policy dimensions in societies, and large regional differences in endowments and institutions can explain divergence. How important the local ecology is in relation to other factors is a matter of conjecture.

Comments on "Development Strategy: Transition and Challenges of Development in Lagging Regions"

The paper by Lin and Liu deals with the challenge of arresting the increasing gap in per capita gross domestic product between developing countries and industrial

countries. The paper attempts to provide an explanation for the lack of convergence observed since World War II. It also tries to ascertain why the measures labeled as the Washington consensus failed to bring about corrections. Drawing lessons from the rapidly growing East Asian economies, the paper demonstrates how the developing countries could calibrate their policy reforms to accelerate economic growth and reduce poverty.

The central theme of the paper is that the failure to achieve dynamism in developing countries is mainly attributable to their CAD strategy of industrialization. A CAF strategy would have resulted in resource allocation according to endowments, resulting in a competitive economy with large surpluses and high saving propensities, and would have continuously upgraded the endowments. In such economies, the unemployment rate would be low, the society would be more egalitarian, and the country would face fewer social tensions. In contrast, when an economy follows a CAD strategy, the industrial structure is unviable in open, competitive markets.

The authors argue that in their anxiety to accelerate the pace of industrialization, the early generation of leaders in independent socialist and nonsocialist developing countries unwittingly adopted a CAD strategy. Low levels of saving and investment, shortages of foreign exchange, obsolete technology, and export pessimism led them to adopt import-substituting industrialization based on heavy industry and dominated by the public sector, which is essentially a CAD strategy. The strategy is basically unviable, and firms have to be sustained by (a) distorted macroeconomic policies, such as financial repression, overvalued exchange rates, and other price and quantity interventions; (b) administered allocation of credit, foreign exchange, and scarce commodities at nonmarket prices; (c) public ownership and control of production and distribution enterprises; and (d) measures to reduce inequalities and control monopolies. The consequence of this has been to constrain the economy from growing to its potential, to create an unviable economic structure, and to create a high degree of economic inequality and a rent-seeking environment. In contrast, the countries in East Asia that followed a CAF strategy experienced high growth, faster poverty reduction, and improvements in their populations' well-being.

Even though the problem is well understood, the solution is not easy. Most developing countries follow a package of measures referred to as the Washington consensus. The measures include strengthening fiscal discipline, liberalizing trade, devaluing their currencies, privatizing public enterprises, and removing entry barriers. They argue that their experience with implementing these measures has not been successful because of the inherent unviability of their domestic industries, which developed in a highly protected environment. The remedial measures require a large number of firms to go bankrupt and a large proportion of the workforce to become unemployed. The effective implementation of such measures could result in serious social and political unrest. Applying shock therapy to reform could result in renewed demands for subsidies and protection, and eventually could stall the entire reform process.

Lin and Liu draw lessons from the successful reform experience of China, where the market-oriented reforms employed a gradualist approach. The viability of each sector was carefully addressed by looking at the structure of incentives to improve productivity, and liberalization was CAF-oriented to ensure the rapid growth of

labor-intensive township and village enterprises not owned by the state. The gradualist approach has avoided social and economic instability while ensuring high growth rates.

The problems with import-substituting industrialization strategy, referred to as a CAD strategy in this paper, is well known (Bhagwati 1993; Srinivasan 2000; World Bank 1993). However, at the beginning of the 1950s, when countries like India launched their industrialization drive and met with some measure of success in accelerating economic growth and building their industrial base, the strategy was hailed as a success story. Note also that the strategy adopted in Japan, the Republic of Korea, and other East Asian economies cannot be termed an unadulterated CAF strategy. Both Japan and other East Asian economies developed within walls of heavy protection, but they provided a heavy export thrust using a variety of means. Their emphasis on physical infrastructure and, more important, on human development; their capacity to pick winners; their repressed interest rates; and their incentives to exporters have all been widely discussed (World Bank 1993). Korea followed a strategy of heavy and chemical industries until the late 1970s, when macroeconomic stability forced it to abandon the strategy. What is important is that these countries were flexible enough to change their strategies when they realized that an import-substituting strategy was inherently unviable.

It is also important to understand that it was possible to change the strategy without serious social and economic instability. Compressing public expenditure by 5 to 6 percentage points of gross domestic product might have been extremely difficult in a democratic polity; nevertheless, Korea succeeded in undertaking such an adjustment during 1979–84 in the wake of a serious macroeconomic crisis (Rao 1998). Vietnam also undertook a similar compression during 1989–92. What is important is that in Vietnam, the changing structure of incentives allowed the absorption of displaced labor in the rapidly expanding agriculture sector. Lin and Liu agree that Washington consensus measures are desirable for the long-term development of the developing countries, but want to solve the unviability problems of domestic manufacturers during the transition period. The speed at which the measures can be implemented should therefore depend on the ability of institutions to absorb changes.

The CAD versus CAF strategy explanation explains per capita income trends among industrial and developing countries, but does not explain why richer regions within a country have continued to perform better than poorer regions. Any theory that attempts to explain the poor performance of lagging regions should also explain this in the context of different regions within a country, not just between different countries. This takes us back to other explanations in relation to policies and institutions, and from the Dasgupta paper, to local ecology.

References

Bhagwati, Jagdish. 1993. *India in Transition: Freeing the Economy.* Oxford, U.K.: Clarendon Press.

Rao, M. Govinda. 1998. "Accommodating Public Expenditure Policies: The Case of Fast Growing Asian Economies." *World Development* 26(4): 673–94.

Srinivasan, T. N. 2000. *Eight Lectures on Indian Economic Reforms*. New Delhi: Oxford University Press

World Bank. 1993. *The East Asian Miracle.* New York: Oxford University Press

Comment on "World Poverty: Causes and Pathways" by Partha Dasgupta and "Development Strategy: Transition and Challenges of Development in Lagging Regions" by Justin Yifu Lin and Mingxing Liu

KAREN R. POLENSKE

I have the pleasure of commenting on the papers by Partha Dasgupta, in which he details pathways to world poverty, and by Justin Yifu Lin and Mingxing Liu, in which they argue against the so-called comparative advantage defying (CAD) strategy as a strategy for development in poor regions. Even though the authors discuss fundamentally different issues, both sets of authors deal with issues of poverty and poverty alleviation and how to achieve sustainable development. In addition, both agree that the growth of per capita gross domestic product (GDP) in developing countries is diverging from rather than converging toward that of the industrial countries, but they disagree about the critical issues that are causing the poverty and the differences in growth rates.

Comments on "World Poverty: Causes and Pathways"

Dasgupta conducts a powerful, extensive, and detailed examination of world poverty based upon his more than 20 years of experience analyzing resource allocation, poverty, and related issues (see, for example, Dasgupta 1982, 2000, 2001). I especially commend him for incorporating ecological considerations into his most recent analyses, including this one. Dasgupta provides a stark, haunting picture of world poverty, showing that many of the growth theories and strategies development economists have proposed are not working. As he emphasizes, despite the many poverty alleviation efforts in various parts of the world, 24 percent of the 1998 world population—40 percent of Asia's and 46 percent of Sub-Saharan Africa's—continue to live in poverty (International Food Policy Research Institute 1995).

Karen R. Polenske is professor of regional political economy and planning at the Massachusetts Institute of Technology in Cambridge, Massachusetts.

Dasgupta examines why poverty traps have allowed "hunger and poverty to be a persistent experience for large groups of people in poor regions even while others there and elsewhere are able to prosper." He claims that people caught in these poverty traps live on the brink of subsistence, in a world where "relatively low mortality rates can coexist alongside a high incidence of undernutrition and morbidity." In the spring and summer of 2003, while television kept highlighting the fewer than 1,000 deaths from severe acute respiratory syndrome (SARS) in countries around the world, hundreds of thousands of women and children died from indoor air pollution in developing countries. Using many similar depressing examples, Dasgupta shows us the lack of attention given to the plight of the poor. These harsh facts present a challenge to those of us in development economics and planning to find ways to make economics less of a dismal science than the profession in which we have been trained by doing something to rectify the dire situation. The question is, what can we do?

Reviewing the work of others who have attempted to answer this question, Dasgupta rightly criticizes Collier and Gunning (1999) for not considering population growth and natural resource degradation adequately in their survey of Sub-Saharan Africa and Sen (1999) and Easterly (2002) for failing not only to account for causes of population growth in the poorest countries, but also for overlooking the critical role of the natural resource base in achieving sustainable development.

To study the issue and to work toward a theoretical framework of analysis, Dasgupta divides the poverty issue into two types of pathways:

- Individual metabolic pathways, which show the effects of nutrition on individuals. The pathways are specified separately for adults and children:
  - Nutrition affects the work capacity of adults.
  - Nutrition affects the physical and mental development of children.

- Site-specific ecological and socioeconomic pathways, which describe two ways in which economic options differ depending on local ecology:

  - Undernourished people are vulnerable to infectious diseases.
  - Poverty, population growth, and degradation of local natural resources are interrelated and both reproductive and environmental externalities exist.

Dasgupta bases his extensive discussion on a systematic accounting of these two pathways. He begins by presenting macroeconomic evidence to support his view on current world poverty, including an inclusive index of wealth that he argues should be used to determine whether long-run well-being is sustainable. By using such an index, he finds that the poorest countries have de-cumulated their wealth relative to their population sizes. In this index he includes manufactured, human, and natural capital, such as oil and minerals, fisheries, forests, grazing land, and aquifers. Each statistic that he presents is worthy of a headline, in that he effectively reveals the destitution in which more than one-fifth of the world's population lives. Dasgupta discusses the measurement of wealth and argues for an inclusive measure that is a comprehensive list of assets, including the three types of capital assets noted earlier.

Dasgupta then argues that nutritional status is positively correlated with capacity to work and that there is also a positive relationship among poverty, population growth, and degradation of the local natural resource base. In the latter case, he concludes that people in the poorest countries have both consumed and invested too little. Dasgupta starts to lay out a theoretical framework concerning economic growth, poverty, and degradation of resources to show how the coping mechanisms prevalently used in developing countries are insufficient to raise them out of poverty. He bases the framework on resource allocation mechanisms and argues that analysts must look at regional and subregional levels, rather than at national levels, as they analyze why poverty can be absent in one region but persist in another, because aggregation leads to erroneous conclusions. He builds upon his arguments as he emphasizes the link between institutions that shape an economy and the distribution of its assets, although he also cautions that "the key issue is access to nutrition and health care, not so much the distribution of assets." Analysts can link the deterioration of institutions, such as property rights, and the efficiency of use of assets.

Dasgupta further develops his theoretical framework by providing considerable information about social norms, threat of sanctions, local commons, and other factors underlying four types of persistent reproductive and environmental externalities: (a) cost sharing, such as the sharing of risks among members of a family; (b) conformity and contagion, in that family size is often dependent on social norms in the community; (c) interactions among institutions, with market and nonmarket institutions often supplying the same good; and (d) property rights in the form of local commons as they relate to households' needs for water, energy, and other basic goods. He notes that none of the analysts in the extensive set of studies he cites has been able to develop the causal connections, although the studies do help support his idea of positive feedback mechanisms.

In summary, Dasgupta outlines the start of a theory of poverty and concludes that "[b]oth the character of human metabolic pathways and weaknesses in human–nature interactions were shown to play significant roles." Safety nets are needed, especially for the poorest of the poor. He maintains that World Bank staff and others who use an environmental Kuznets curve are wrong. He says that implicitly they appear to approve of degradation, because an analyst could conclude from applying the curve that a population can pollute all it wants early in the growth of the economy. As the economy grows, the Kuznets curve would indicate that the amount of pollution will be reduced. Dasgupta believes this is a dangerous path to propose.

Although Dasgupta only outlines a theory he is developing of how and why poverty is occurring throughout the world, he also provides no more than the bare bones of a strategy for getting countries out of the trap, indicating that he cannot do more because (a) the evidence is too sparse, (b) national accounts must be revised to account for environmental degradation, (c) shadow prices must be computed to reflect the scarcity values of natural resources, (d) methods need to be developed to account for the nonlinearity of ecological and economic processes, and (e) more time is needed to gather information prior to promoting strategies. He states that the long run is what matters and that, in the long run, the average well-being of the

population has declined in the world's two poorest regions, South Asia and Sub-Saharan Africa. Thus, rather than indicating a specific strategy to get people out of poverty, he focuses on the two key pathways, noted earlier, of the environment in which rural people in poor countries make their decisions.

Because Dasgupta's analysis is extensive and well done, I hesitate to say that he has neglected some important issues. He himself has made reference to some of these, but I feel he could have relatively easily explored them in greater depth, and so I feel justified in mentioning them. Dasgupta indicates that evidence for the theory he is developing is incomplete, but I suggest that he could fairly easily extend his analysis to include discussions of data, measurements, and tools, and, most important, to elicit the help of analysts from other disciplines. My suggestions follow.

First, at several points in his paper Dasgupta indicates the types of analytical tools and data, such as improved national accounts, that may be needed to help define the poverty problem and its causes. Given the depth of his experience, he could easily outline some of the key missing tools and data. He indicates, for example, that cross-sectional relationships may mislead policymakers and that ecological processes tend to be nonlinear, yet he notes that most economic models reflect linear relationships. What can development economists do to bridge those gaps?

Also Dasgupta is definitely aware of many data issues, but I would like him to indicate more from his own findings concerning where data are inadequate. Does the exclusion of informal sector and underground economy data from most published statistics affect his analysis of ecology and poverty? Which of the missing data would help him the most to discuss ways to alleviate poverty? Are the important data the relative prices of alternative sources of energy and water in rural areas in poor countries versus elsewhere; the proportion of total assets embodied in village ponds, grazing fields, and local forests; or something else? Another question concerns how to measure the total assets in a particular rural area. What role do economic, social, and political power play in controlling access by the poor to those assets (Polenske 2001)? He defers from offering policy prescriptions, because he senses that many of the data and results of analyses may be flawed.

Second, Dasgupta's call for an inclusive measurement of wealth and for the measurement of degradation is not new. Many national accounting analysts and economists have suggested similar measures and, indeed, have constructed such measures for industrial and developing countries; Dasgupta's failure to mention them is curious. These include, to name just a few, Repetto and others (1989) at the World Resources Institute, who are some of the earliest economists to try to account for the degradation of natural resources; Duchin and Lange (1994); Uno (2002); and the United Nations (2003), with staff at the latter proposing the creation of satellite environmental accounts. Repetto is currently working on the theory of nonconvex, nonlinear, dynamic systems for studying environmental policy change and nonlinearity, which is an approach that Dasgupta emphasizes as the way in which many of the ecological factors operate. The extensive accounting literature on ways to incorporate changes in the quality and amount of natural resources into the accounts covers, although by different names, at least some of

the distinction that Dasgupta makes between what he calls inclusive and recorded investment. Specific reference to previous work would have been useful. Even so, I am pleased that Dasgupta emphasizes the need for such resource accounting. Few poverty analysts do.

Third, Dasgupta says that most growth theorists at least imply that "[t]he term globalization is used to signal that location per se does not matter." Here I interpret what other economists are saying differently than Dasgupta. Unless I am misinterpreting his statement, he implies that globalization is the culprit, but I do not believe we can blame globalization for many economists' oversight of the significance of location. Rather, on the one hand, most growth theorists and other economists have never understood the importance of location nor, until recently, have they incorporated location into their economic analyses. On the other hand, regional economists and regional geographers have always analyzed and applied theories of firm location, factor mobility, interregional trade, restructuring, and other spatial theories, and these regional economists and geographers continue to emphasize location as they study globalization precisely because location does matter. Finally, regarding regional transportation and regional supply-chain analyses, globalization is one of the determining factors in the success of a supply-chain management approach by firms; yet it is mainly analysts in fields other than economics who are studying this important phenomenon.

Fourth, Dasgupta mentions the need for efficient risk markets, as Stiglitz (2000) notes. Is it possible to use a regional risk model, such as the one Blanco (2003) proposes for Colombia, as a way to help attract investors to some high-poverty regions, even if the political risk in the country as a whole seems too high for the investor?

Dasgupta is still developing his theory of poverty, so that I hope that the questions I raise may help him continue to expand his theory to answer such questions. He has written an important, thought-provoking paper that reminds us of how critical the issue of poverty is and why more analyses are needed. I maintain that solving the problem and achieving a sustaining world economy will require the help of analysts from many disciplines, including biology, economics, engineering, management, physics, and sociology, to name just a few (Polenske forthcoming). All can provide additional insights into pathways to improve nutrition, reduce environmental degradation, and decrease poverty.

Comments on "Development Strategy: Transition and Challenges of Development in Lagging Regions"

Based on macroeconomic data from 108 countries for 1962–99, Lin and Liu state that, with only a few exceptions, such as China and Vietnam, the per capita GDP gap between industrial and developing countries increased. They argue that the reason for that poor growth performance in developing countries is that the countries followed an inappropriate development strategy, which they call a comparative

advantage defying (CAD) strategy. Lin and Liu claim that after World War II, most developing countries adopted a development strategy that was capital intensive and heavy-industry oriented, which led firms in the developing countries to be nonviable in open, competitive markets, thereby causing the following vicious cycle:

- Nonviable firms led to government-induced trade, finance, and labor distortions.
- These distortions created investment-led growth initially and caused resource misallocations.
- These misallocations caused rent-seeking, macroeconomic instability, and suppressed private initiatives.
- These results led to resource depletion.
- The final consequence was stagnation of the economy.

Lin and Liu require readers to agree with the following four basic statements:

- Most developing countries have followed a CAD strategy of development.
- The growth of labor-intensive sectors is conducive to overall economic growth in the transition process for all countries.
- The neoclassical solution is to upgrade the endowment, rather than the industrial structure, by letting the price structure reflect competitive markets.
- An alternative solution that will allow the countries to grow is for them to adopt what Lin and Liu call a comparative advantage following (CAF) strategy and to consider firm viability in designing reform policies.

To understand Lin and Liu's argument, readers need to understand their definition of firm viability and the terms CAD and CAF. In their current paper, they use Lin's (2003, p. 280) earlier definition of a viable firm: "If, without any external subsidies or protections, a *normally managed* firm is expected to earn socially acceptable profits in a free, open, and competitive market, then the firm is viable. Otherwise the firm is nonviable." Obvious questions are whether there are any developing countries in which markets are free, open, and competitive? If not, does the absence of such a market negate their findings?

Concerning the CAD and CAF terms, the authors never explicitly define what those terms mean. They argue that a country that initiates a CAF strategy will invest in labor-intensive sectors rather than capital-intensive ones. To do this, policymakers must first get rid of the nonviable firms, partly by ending the massive subsidies the authors say that countries are currently giving to such firms.

I have three main concerns with the Lin and Liu proposal as set forth in their paper. One of my major concerns is that the authors do not advocate different versions of the CAF development strategy for different countries. The shoe is pinching and hurting, but one size must fit all, and all that policymakers need to do is to change shoes. Thus, if policymakers substitute the CAF for the CAD strategy for all countries, then the countries will begin to prosper. I do not believe that is so. Housing, transportation, energy, water, land uses, and many other issues vary greatly among developing countries, partly depending on the size of their population, the extent of their natural resource base, the skill of their labor force, the type of governance system and institutions, and so on. Why should all policymakers

follow the same development strategy in our diverse world? Lin and Liu argue that countries should follow their comparative advantage, but a country may well have a comparative advantage in natural resources and not in human skills. How do policymakers determine the comparative advantage of the country? By advocating a single strategy, Lin and Liu tend to simplify the extremely complex set of development factors that occur in different parts of the world. Different countries face different human and natural resource constraints. I believe, therefore, that in discussing development strategies they have omitted important elements.

Second, Lin and Liu's critique could have been substantially bolstered with some of the tables and facts from other papers Lin has written, in particular, his papers on the growth of township and village enterprises in China (for example, Lin 2003; Lin and Yao 2001). The data and some of the analysis presented in Lin and Yao (2001) may help their argument that China followed a CAF strategy. Lin and Yao provide a number of tables to illustrate how the output of rural (township, village, and individual) enterprises has grown rapidly, far outpacing the output growth of China's state-owned enterprises. The amount of labor working in rural industry rose from 9.2 percent in 1978 to 28.4 percent in 1996, and industrial output as a percentage of total national output increased from 9.1 to 57.9 percent (Lin and Yao 2001). Lin and Yao also place China's development in a historical context, which is especially important for understanding the rapid growth of output of these enterprises. Does Lin and Liu's proposed strategy pertain to regions in China that have many township and village enterprises as well as to those in which state-owned enterprises still dominate? In countries as large as China and India, the country may need to use different development strategies in different parts of the country.

Third, and perhaps the most troubling part of Lin and Liu's analysis, is the empirical section of their paper in which they conduct a set of regression analyses for three different options. The unit of analysis they want to examine is developing countries, but in their calculations they include all 23 industrial countries as well as the 85 developing countries. I asked my research assistant, Zhan Guo, to conduct a simple test by removing the 23 industrial countries from the database. When Guo included the industrial countries (table 1), the correlation coefficient between the CAD strategy and GDP is significant at the 0.05 level, but when he excluded the industrial countries, the correlation is not significant (table 2).

TABLE 1.
Correlation between the CAD Strategy and GDP Per Capita Including the Industrial Countries

Pearson correlation significance (2-tailed)	Mean TCI (1963–99)	Growth rate of GDP per capita (%) (1962–99)
Mean TCI (1963–99)	1.00	−0.202 (0.064)
Number of observations	85	85

TCI   Technology choice index of manufacturing (which can be a proxy for the government's adoption of a CAD strategy).

Note: The level of significance is between 5 and 10 percent.

Source: Calculations performed by Zhan Guo based on data for the 108 countries in Lin and Liu's paper.

TABLE 2.
Correlation between the CAD Strategy and GDP Per Capita Excluding the Industrial Countries

Pearson correlation significance (2-tailed)	Mean TCI (1963–99)	Growth rate of GDP per capita (%) (1962–99)
Mean TCI (1963–99)	1.00	−0.238 (0.013)
Number of observations	108	108

TCI   Technology choice index of manufacturing (which can be a proxy for the government's adoption of a CAD strategy).

Note: The level of significance is less than 5 percent.

Source: Calculations performed by Zhan Guo, based on data in the Lin and Liu paper.

I therefore suggest that Lin and Liu redo the sample of countries that they use for their calculations and then rethink their findings based on the new calculations. Overall, Lin and Liu need to present more data to support their arguments for why countries should adopt a CAF strategy. They can do this relatively easily, because Lin has presented considerable data in previous papers to which Lin and Liu could easily refer for their argument.

In addition, the CAF development strategy proposed by Lin and Liu seems limited in that they neglect, or mention only briefly, other important aspects of an overall development strategy, such as (a) the broader asset base of a country or region, including the knowledge base of its laborers; (b) the institutional structure of countries, including the possible role of ambiguous property rights and of enforcement of laws; (c) the role played by the presence of a "gift" economy; (d) the economic, political, and social power relations that affect the effectiveness of any development strategy; and (e) the distributional consequences of the development strategy. Polenske (2001) provides more details about these factors in her presentation of an asset-based strategy of development.

Finally, Lin and Liu refer briefly to some of the factors that created a viable situation in China, but they also mention that Vietnam has managed to follow a CAF strategy and has achieved a growth rate of 6.8 percent in recent years, and Lin and Yao (2001) note that China and Taiwan (China) have both established labor-intensive, low-tech sectors in rural areas. I would like Lin and Liu to present more details about Taiwan (China) and Vietnam in their current paper. In addition, the CAF strategy in China has supposedly led to an increase in income inequality between regions because of the income gap between rural and urban households (Lin and Yao 2001). This disparity, however, varies from region to region, at least in part because the inland regions have a larger rural population than the coastal region. This result would argue for development analysts to pay more attention to the spatial implications of any development strategy. I have argued throughout that a number of different viable development strategies probably exist. I see no reason why a single development strategy would be the most appropriate for all countries or for all regions within a country.

References

Blanco, Alberto E. 2003. "Geographical and Behavioral Economics of Political Risk for Foreign Direct Investment Location." MS thesis, Massachusetts Institute of Technology, Department of Urban Studies and Planning, Cambridge, Mass.

Collier, P., and J. W. Gunning. 1999. "Explaining African Economic Performance." *Journal of Economic Literature*. 37(1): 64–111.

Dasgupta, Partha. 1982. *The Control of Resources*. Cambridge, Mass.: Harvard University Press.

_____. 2000. "Population, Resources, and Poverty: An Exploration of Reproductive and Environmental Externalities." *Population and Development Review* 26(4): 748–54.

_____. 2001. *Human Well-Being and the Natural Environment*. Oxford, U.K.: Oxford University Press.

Duchin, Faye, and Glenn-Marie Lange. 1994. "Strategies for Environmentally Sound Economic Development." In A. M. Jansson, C. Folke, R. Costanza, and M. Hammer, eds., *Investing in Natural Capital: The Ecological Economics Approach to Sustainability*. Covelo, Calif.: Island Press.

Easterly, W. 2002. *The Elusive Quest for Growth*. Cambridge, Mass.: MIT Press.

International Food Policy Research Institute. 1995. *A 2020 Vision for Food, Agriculture, and the Environment*. Washington, D.C.

Lin, Justin Yifu. 2003. "Development Strategy, Viability, and Economic Convergence." *Economic Development and Cultural Change* 53(2): 277–308.

Lin, Justin Yifu, and Yang Yao. 2001. "Chinese Rural Industrialization in the Context of the East Asian Miracle." In Joseph E. Stiglitz and Shahid Yusuf, eds., *Rethinking the East Asian Miracle*. Oxford, U.K.: Oxford University Press.

Polenske, Karen R. 2001. "Taking Advantage of a Region's Competitive Assets: An Asset-Based Regional Economic Development Strategy." In *Entrepreneurship, Firm Growth, and Regional Development in the New Economic Geography*. Trollhättan, Sweden: Uddevalla Symposium.

_____, ed. Forthcoming. *The Technology, Energy, Environmental, Health (TEEH) Chain in China: A Case Study of the Cokemaking Sector in Shanxi Province, China*. Boston: Kluwer.

Repetto, R., W. Magrath, M. Wells, C. Beer, and F. Rossini. 1989. *Wasting Assets: Natural Resources in the National Income Accounts*. Washington, D.C.: World Resources Institute.

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

Stiglitz, Joseph. 2000. "Whither Reform? Ten Years of the Transition." In Boris Pleskovic and Joseph E. Stiglitz, eds., *Annual Bank Conference on Economic Development*, Washington, D.C.: World Bank.

United Nations. 2003. *Handbook of National Accounting: Integrated Environmental and Economic Accounting 2003*. New York.

Uno, Kimio, ed. 2002. *Economy-Energy-Environment Simulation*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Participation, Inclusion, and Results

Community-Based Development in the Context of Within-Group Heterogeneity

JEAN-PHILIPPE PLATTEAU

With a great deal of determination, the international donor community has recently adopted a new approach to fighting poverty known as the community-based development (CBD) approach. Such an abrupt shift in donors' aid strategies is questionable, not because the approach is wrong (the opposite is actually the case), but because massive injections of aid funds into CBD projects have given rise to the entry into the field of numerous agencies with little or no experience in participatory development and a pressing need for quick and visible results, both of which threaten to undermine CBD's effectiveness in reducing poverty. The worry arises from the elite capture problem, which risks deflecting a large portion of the resources devoted to CBD into the hands of powerful groups dominating target communities. This paper both documents the elite capture problem and discusses the use of sequential and conditional disbursement procedures as a way of surmounting it. It also examines how various elements of the aid environment, including the pressure of competition among donor agencies and the availability of aid funds, influence the share of CBD aid that reaches the poor. Finally, multilateral reputation mechanisms and intra-community competition for leadership are assessed as possible alternatives to sequential disbursement procedures.

Introduction

That community participation leads to improved project performance and better targeting compared with top-down service delivery and poverty reduction approaches has become a sort of received wisdom today. The idea is not new. On the one hand, it has

been continuously advocated by development scholars during the last two decades (see, for example, Baland and Platteau 1996; Cernea 1985; Chambers 1983; Hirschman 1984; Ostrom 1990). On the other hand, not only was it attempted during the 1950s by the Ford Foundation and U.S. foreign assistance programs (by 1960, as many as 60 countries were involved in this community development thrust) before being abandoned (Holdcroft 1984), but nongovernmental organizations (NGOs) working in developing areas claim that it is their key approach. What is striking today is that largely as a response to critiques of top-down development, most bilateral donors and large international organizations have started to include participatory elements in the design of their large-scale development assistance programs (think, for example, of the World Bank's social investment funds or of participatory development programs sponsored the International Fund for Agricultural Development), or to channel substantial amounts of aid money through international or local NGOs (Stiles 2002). The move to put participation and empowerment of the poor squarely on the agenda is especially noticeable in the case of the World Bank, which has made so-called community-driven development one of the cornerstones of its Comprehensive Development Framework. The *World Development Report 2000/2001* (World Bank 2001) duly reflected this shift in approach (Mansuri and Rao 2003).

Given the high hopes placed on community-based development (CBD) and the determined attempts to scale up projects based on this new approach, assessing the strength of the case in support of it is important. Such an assessment is all the more necessary, as the available empirical evidence—so far as we can judge from recent surveys (Bardhan 2002; Conning and Kevane 2002; Mansuri and Rao 2003)—does not unambiguously confirm the view that CBD projects are more effective than more conventional approaches in terms of efficiency, equity (reaching the poor), and sustainability. NGOs themselves, contrary to a widespread belief, have not produced impressive results, even with respect to alleviating poverty and promoting participation (Carroll 1992; White and Eicher 1999). The same agnostic conclusion emerges from a recent review of empirical studies of decentralized delivery of public services (Bardhan 2002). According to the author, even though the studies suggest generally positive effects of decentralization, "[I]t is hard to draw conclusive lessons" (Bardhan 2002, p. 200; see also, for example, Asthana 2003). Caution is required, because most studies are essentially descriptive and point to correlations rather than to carefully tested causal relationships.

Few reliable evaluations of participatory development projects have been undertaken (that is, evaluations grounded in representative samples with treatment and control groups as well as baseline and follow-up data), even by NGOs that have used the approach for several decades (White and Eicher 1999, Brett 2003). The skepticism of many NGOs about the need for and purpose of evaluation and their emphasis of action over analysis are largely responsible for this situation (Ebrahim 2003). Thus the evidence produced in favor of CBD tends to be anecdotal or based on unqualified generalizations. Indeed, when evaluations do take place, they are often biased in a direction favorable to CBD projects. A "praise culture" is pervasive among all the actors involved, who have a tendency to "resist the presence of evaluators and

make efforts to influence their work and present results that will provide a more favorable impression" (Mansuri and Rao 2003, pp. 30–31; see also Riddel 1999, pp. 223–34). Finally, to the extent that serious evaluations of NGO projects exist, they are almost never released for public scrutiny.[1]

Unfortunately, the criticism can be extended to almost all organizations that have adopted the CBD approach. Thus, for example, interventions funded by the World Bank have also been inadequately assessed, a fact deemed inexcusable by Mansuri and Rao (2003). The consequence of this dearth of reliable evaluations is ineffective learning by doing where it is badly needed. Clearly, the available evidence does not justify the speed and the enthusiasm with which many agencies, especially large bureaucracies, have started to implement CBD (Platteau and Abraham 2001, 2002, forthcoming). In particular, approaches that use rapid disbursement procedures to channel substantial external resources toward local development are likely to seriously undermine the welfare and long-term objectives of CBD. Such a point has been brought to the fore in a recent report assessing the effectiveness of the World Bank's social funds (World Bank 2002). The relevant conclusion is worth quoting at some length (see also Edwards and Hulme 1996; White and Eicher 1999):

> While social fund projects have been successful in channeling substantial external resources toward local development, disbursing rapidly and achieving their physical output targets, their impacts on outcome and welfare variables, and on community capacity building and social capital, have been mixed . . . Many social funds that were initially set up as emergency response mechanisms are now being called upon to shift their focus to longer-term development impact and institutional development objectives, but this transition is proving difficult to implement in some cases . . . [S]ocial funds are liable to meet the same constraints as other kinds of interventions and may lose the strengths on which their reputation has been built. For example, building capacity and social capital at the community level are time- and human resource-intensive processes, making disbursements potentially slower and less predictable. Experience suggests that the constraints to accomplishing this transition should not be underestimated and the tradeoffs should be explicitly addressed (World Bank 2002, p. 48).

The issue of whether CBD is more effective than more centralized approaches to tackling poverty can be framed in terms of a trade-off between information advantages and the risk of elite capture. If the former advantages outweigh the latter risk, the case in favor of CBD seems strong, but if the risk of elite capture is high and the information advantages are not significant, the grounds for embarking upon CBD become shaky. The next section of this paper describes the trade-off in more detail, while in the following section its relevance is further appraised in the specific context of foreign aid relations. The paper then turns to the question of how the entrenched elites that are likely to misappropriate aid resources could be disciplined through an appropriate mechanism. This issue is addressed within the framework of aid projects financed by donor agencies from industrial countries. As the discussion in this section shows, the issue is complex: there is a serious risk that in the competition prevailing among donor agencies to get access to local communities quickly, local leaders eventually misappropriate many resources transferred from rich to poor countries. The last section concludes.

Note that throughout the paper the focus is on the problem of elite capture. This follows from the fact that CBD is essentially aimed at relieving poverty and promoting widely spread rural development more effectively than what has been achieved in the past through more centralized approaches.

Trade-Off between Information Advantages and Elite Capture: The Problem

The main advantage associated with CBD lies in communities' or user groups' better knowledge of local conditions and constraints (environmental, social, and economic), as well as in the dense network of continuous interactions among individuals that constitute community life (often referred to as social capital in the recent literature). As a result of these two features, authors writing on the subject typically assume that communities are better able than a central government or an external donor not only to set up priorities, identify deserving beneficiaries, design projects, and select techniques and inputs, but also to enforce rules, monitor behavior, and verify actions. In addition, people's motivation to apply effort and to contribute resources is expected to be stronger when they are allowed to choose their objectives and how to achieve them rather than being told from above what to do and how to do it.

If, in theory, a central agent might procure for itself the same information advantage of proximity by posting local agents in the field, there is apparently no way in which it could avail itself of the social capital available in a community. Even the information advantage is likely to be thwarted by considerations of political accountability. Indeed, according to Bardhan (2002, p. 191): "[L]ocal politicians may have more incentive to use local information than national or provincial politicians, since the former are answerable to the local electorate while the latter have wider constituencies, where the local issues may get diluted." The political accountability argument is, however, less pertinent when applied to a context in which external donors rather than central governments are the purveyors of funds.

The other side of the coin is that local governments or communities may be more prone to capture and thus less accountable than central governments (or external donors), and, if that is the case, decentralization can also be subject to misappropriation and targeting failures. In the words of Bardhan (2002, p. 192):

> Political accountability in poor countries is particularly affected by the likelihood of corruption or capture by interest groups. While local governments may have better local information and accountability pressure, they may be more vulnerable to capture by local elites, who will then receive a disproportionate share of spending on public goods.

A plausible argument is that, at least in situations of high inequality, the poor and minorities are more easily oppressed by local power groups that can easily collude beyond the control of higher-level institutions and the attention of the media. Moreover, social capital may be harnessed against rather than in favor of vulnerable segments of the population. This is because "the multiplex interlocking social

and economic relationships among local influential people may act as formidable barriers to entry into these cozy rental havens" (Bardhan 2002, p. 194). Facing these strong collusive networks, the poor are often helpless because their own networks, typically geared to cope with immediate subsistence problems, are not in a position to dispute the power the rich wield. In these circumstances, the poor may naturally look to the central state for protection and relief (Bardhan 2002), and if the central government is not responsive to their needs, their predicament persists.

Thus the conventional presumption is as follows: the lower the level of government, the greater the extent of capture by vested interests. If this is correct, the information advantage of CBD programs would be compromised by their greater diversion to the benefit of local elites (Bardhan and Mookherjee 2000a). The case for CBD would then hinge on the relative strengths of the two opposing effects.

Bardhan and Mookherjee (2002) develop a tight framework to analyze the trade-off between the two conflicting aspects of centralized versus decentralized systems of service provision and delivery in the context of infrastructure services, such as roads, water, electricity, and telecommunications. In the centralized system, authority is assumed to be assigned entirely to bureaucrats whose objective is to maximize their net incomes (that is, bribes less the costs of delivery). These bureaucrats behave like unregulated monopolists. The effect of decentralization is to shift control rights to a local government that, under the pressure of electoral forces, seeks to maximize a weighted sum of the welfare of two types of local user: elites and non-elites. Two other assumptions are crucial to their theoretical exercise: elites value the service provided more than non-elites, and the phenomenon of capture of local governments is reflected in the fact that elites receive a higher welfare weight.[2]

What Bardhan and Mookherjee (2002) show is that decentralization tends to expand service delivery as authority is devolved to those most responsive to users' needs. Yet, given local elite capture in the sense defined earlier, the local government tends to overprovide the service to local elites at the expense of the non-elites. The amount of such overprovision actually increases with the degree of fiscal autonomy granted to the local government. This is because with local tax financing there is the risk that the captured local government may resort to a regressive financing pattern (whereby the non-elite bear the tax burden of providing services to the elite). Therefore restrictions on local governments' ability to raise local taxes can be justified on efficiency and equity grounds. User fee mechanisms, by contrast, ensure that decentralization welfare dominates centralization, irrespective of the degree of local capture. This is because no one is being compelled to use the service, and thus user charges impose a limit on the extent of cross-subsidization of the rich by the poor.

Here is an obviously distressing conclusion if the problem is to relieve poverty by catering to the poor's basic needs (food, health, and education). Indeed, the poor do not have the ability to pay for the services intended for them (or for bribes to the bureaucrats). In such cases, as Bardhan and Mookherjee (2000b) show, the extent of elite capture at the local level relative to that occurring at the central level is a critical determinant of the welfare impact of decentralization.

In still another paper, Bardhan and Mookherjee (1999) investigate theoretically the determinants of relative capture of local and national governments in the context of a model of a two-party electoral competition with probabilistic voting behavior and lobbying by special interest groups (the non-poor are organized in a lobby and can make campaign contributions). One interesting result is that relative capture depends on heterogeneity with respect to levels of local inequality and poverty: decentralization will tend to increase elite capture in high-inequality localities (because higher inequality reduces the level of awareness of the poor, decreasing the level of their political participation) and lower it in low-inequality ones. Nevertheless, while several factors tend to increase the relative proneness to capture of local governments, other factors have the opposite effect. The contrasting roles of these diverse factors—cohesiveness of interest groups, degree of voter ignorance at the local level, relative extent of electoral competition, and so on—suggest that local governments are unlikely to be universally vulnerable to greater elite capture. The extent of elite capture at the local level "may well turn out to be context- and system-specific" (Bardhan and Mookherjee 2000a, p. 139), which creates a need for empirical research to appraise the potential pitfalls of decentralization in various settings.

This theory does not enable us to obtain clear-cut answers to the question of the relative desirability of decentralized versus centralized development. However, it has the merit of drawing our attention to crucial factors, such as within-community heterogeneity, that impinge upon the comparative effectiveness, in both efficiency and equity terms, of the two approaches. Except for the study by Galasso and Ravallion (forthcoming), to which I shall return later, one of the few serious attempts to test the sort of models discussed is Foster and Rosenzweig (2002). These authors use a model of two-party (the poor and the non-poor) representative democracy with probabilistic voting in which local governments must choose to allocate public resources among different public goods for which the preferences of the poor presumably differ from those of the rich. A key prediction of the model is that in villages with democratic governance, an increase in the population share of the landless should result in outcomes that are, other things being equal, more favorable to the poor; that is, more road construction or improvements, which are relatively labor intensive, and less public irrigation infrastructure, which benefits landed households in particular. Applying the econometrics to a 21-panel data set from 250 villages in rural India bears out this prediction.

Caution is nevertheless needed in the interpretation of such results insofar as they are based on a comparison of predicted and realized outcomes in the absence of strong direct testing of the underlying assumptions. In particular, there is doubt about whether improved roads benefit the poor more than irrigation infrastructure. Thus:

> [I]t is often the case that non-poor households corner most of the wage work opportunities within their home village, especially when this work is provided by government agencies at an official wage rate that is two to three times the traditional village rate (Kumar 2002, p. 776).

Moreover, we would obviously like to know more about how village democracy works in practice. Indeed, to show that democratic governance enables the poor to express their preferences and make them prevail, there is no escape from analyzing the concrete process through which they raise their "voice" in the relevant institutions.

By relying on formal voting processes and formal rules of electoral competition, political economy models also ignore other, potentially effective local accountability institutions. It is thus revealing that in nondemocratic countries such as China and the Republic of Korea, ingenious mechanisms exist at the local level to develop trust and cooperation within the ambit of incentive-based organizations and bureaucratic procedures, whereas in democratic countries such as India, local-level accountability mechanisms are often quite deficient (see, for example, Wade 1985, 1990). Indeed, because of the multiplicity of intervening factors (see Agrawal 1999, chapter 3, for other considerations), the abstract stylization of political economy models does not easily lead to reliable, testable propositions.

When we contemplate decentralized or participatory development as practiced by external donors rather than by central governments, the picture appears to be somewhat neater. To the extent that external donor agencies can be deemed to be genuinely committed to relieving poverty, the risk of elite capture at the central level is unambiguously lower than the same risk at the local level. A trade-off between information advantages and the risk of elite capture is then certain to exist, and if the latter is high compared with the former, the desirability of CBD should be called into question.

The Trade-Off between Information Advantages and Elite Capture in the Context of Foreign Aid Relations

Let us now review a number of considerations that should help us assess the relevance of the trade-off described in the previous section, with special reference to relations between external donors and target communities.

The Information Advantage of Communities: Some Qualifications

While communities or user groups no doubt possess information advantages over an external donor agency, several problems may arise that the CBD literature generally overlooks. People may not have a clear perception about the critical dimensions of poverty reduction strategies; their views may diverge from those of donor agencies, especially if the poor have internalized the values of the local elites; or people's preferences may be heterogeneous, giving rise to conflicts of interest.

Members of a community may not have reached consensus on some critical dimensions of an aid program. In particular, they may not agree on who is poor and who is not, or on the nature of the most important problems to be addressed and how best to do so. For example, Bergeron, Morris, and Banegas (1998) show that in

Honduras, when different randomly selected subgroups of community members were asked to establish wealth and food security ratings, the correspondence between the rankings obtained was quite weak. My own experiment with wealth and power rankings in fishing villages in south India (Kerala state) led me to a similar conclusion. Moreover, my experience with NGO work in participatory development in West Africa has shown that villagers are not always clear or correct about the causes of their problems, what their priorities should be, and what strategies ought to be followed to meet those priorities. Confusion or ignorance is especially likely when the matter concerned is somewhat technical or complex.[3] These are the kinds of circumstances that make people especially prone to being influenced by external agencies, in the sense that they tend to demand the sorts of things that they know will appeal to such agencies, especially if they are simply asked to answer an invitation to submit subproject proposals. Moreover, "if the name of the game was poverty, then poverty could surely be demonstrated" (Kumar and Corbridge 2002:80).

If participation is to mean anything in such a context, the intervention of outside facilitators is required. Their role should consist of initiating and supervising a process whereby a community can form an opinion about (a) a list of valid objectives, (b) a suitable sequencing of their realization over time, and (c) a coherent and feasible action program to achieve them through appropriate methods. This process will necessarily be slow, because the facilitators must not impose their own ideas on people. Instead, they must carefully listen and then make suggestions intended for stimulating discussions within the community that will drive its members to think of critical issues and eventually agree on some way to address them.

This is obviously highly subtle work that requires facilitators with the right kinds of motivations and combinations of qualities, as well as patient donor agencies ready to wait before disbursing funds. These two conditions are rarely met in reality. For one thing, facilitators are too often young, poorly paid, and inexperienced individuals who are driven by incentives that are not well aligned with the needs of CBD projects. For another thing, project implementers—especially, but not exclusively, when they belong to large aid bureaucracies—are typically concerned with showing rapid results, while increased participation does not necessarily improve project performance, at least in the short and medium term (Isham, Narayan, and Pritchett 1995; Khwaja 2002; Mansuri and Rao 2003; White and Eicher 1999). Too often, participatory planning is an ideal that exists in speeches rather than in reality. Aid agents initiate a process of analysis within the target community that ends as soon as posters reporting the agreed upon objectives and methods have been taken to the agency to form the basis of its project interventions (Birch and Shuria 2001; Ebrahim 2003; Eversole 2003; Vivian and Maseko 1994).

In contrast with the foregoing case—even though in practice the two situations may be rather hard to disentangle—community members may have a clear and consensual perception about who needs to be helped, what the cause of their predicament is, what needs to be done, and how it should be done, yet their views and preferences may diverge substantially from those of the donor agency. Thus observers often note that the intended beneficiaries pay much less attention to long-term, strategic considerations, including the building of autonomous organizational capacities, and attach

much greater weight to immediate improvements of life conditions than external aid agencies. They also tend to place too much hope in externally-provided resources and to demand that the scale of development activities is increased beyond the limit of their own absorptive capacity. More fundamentally, target groups may just not understand the very concept of development that lies at the heart of the donors' approach (Laurent 1998).

Community members may also have an idea of eligibility that is inconsistent with that of the donor agency. Thus poor members deemed undeserving because they are known to be lazy, frequently drunk, or undisciplined, or because they have broken some local social norm (for example, a son who has not shown respect to his father, or a daughter who has separated from her husband and returned to her native village against the wishes of her parents) will be considered ineligible for aid relief, whereas the donor agency might think differently on the basis of other criteria or principles of justice. Insofar as the undeserving members have internalized the values and norms prevailing locally, a community versus donor preference dichotomy is observed. When such is the situation, discussions are required in the hope that the stances of the two parties will converge without the donor imposing its will. But this is a time-consuming process (Birch and Shuria 2001), and the danger always looms that the intended beneficiaries will again strategically adapt to the demands of the donors and pursue their own agenda while using the aid resources. In the words of an anthropologist with long field experience in the Mossi villages of Burkina Faso:

> Confronted with the hegemonic "project" of the donor, the local population, for fear of losing the aid offer, prefer to remain silent about their practices and aspirations. This is because these practices and aspirations are perceived to be so far away from those of the donor that they are better not disclosed. Such is the vicious circle of development cooperation: the fear of avowing the discrepancy between the two views because it could lead to the discontinuation of the aid relationship has the effect of strengthening the donor's confidence in the validity of its approach (Laurent 1998, p. 212, my translation).

A further complication arises when preference heterogeneity exists within the target community. Thus rural communities are often concerned with preserving a sense of social inclusiveness, which leads them to exclude certain segments of the poor while insisting on the eligibility of the rich (Conning and Kevane 2002). Immigrants of more or less recent origin, nomadic people, erstwhile slaves in caste societies, and widows may thus be precluded from benefiting from an external intervention. For example, a recent study of southern Sudan finds that local views about who should benefit from famine relief efforts were very much at variance with those of the aid workers, which caused a great many problems in relation to project implementation (Harragin forthcoming). A similar difficulty emerges from another study dealing with a CBD project designed to promote community-organized and community-funded schools in Kenya (Gugerty and Kremer 1999). However, a study in rural Bangladesh reaches a more optimistic conclusion. Adams and others (1997) find a good match between wealth-ranking judgments arrived at through a rapid rural appraisal technique on the one hand, and ratings obtained by using standard socioeconomic indicators from a household survey on the other hand.

Tagging by the external agency—that is, categorical targeting that offers eligibility to all members of a group defined by an easily identifiable characteristic or trait (Conning and Kevane 2002)—seems to be the obvious way to surmount such a divergence. Unfortunately, things may not be so simple. For one thing, community members can subvert a program in many ways if they think that it runs against some local social norm. An external observer may not be readily able to detect these ways, especially if the benefits received by, say, nomads or migrants, are not openly taken away from them but are canceled out through the withdrawal of some other benefit that they were previously enjoying. In addition, by imposing eligibility or other criteria that are incompatible with the local culture, the external agency may cause tensions within the community that could hamper its ability to act collectively in other circumstances. Again, time is needed to overcome such differences.

Preferences can differ not only between community insiders and marginalized groups whose membership is questioned but also among the community members themselves. In such cases, the question as to how heterogeneous preferences are aggregated comes to the fore. Rather than through majority voting, decisions tend to be made by the elite alone or through unanimity voting (see Platteau and Abraham 2002, 2004). In highly differentiated societies, mechanisms whereby a consensus is forged among contending parties are almost always a tool the elite use to impose their own views behind a screen of democratic discussions. Clearly when this happens, the dominating community preferences will differ from those of the donor agencies. From a case study on the Jamaica Social Investment Fund, Rao and Ibanez (2001) conclude that the overall quality of the match between local preferences and project achievements was poor. Only in two of the five communities studied was the project obtained consistent with the preferences of a majority in that community. Furthermore, better educated and better networked people were more likely to obtain projects that matched their preferences.

Because powerful elites can easily manipulate disadvantaged people, reserving seats on a village council for the disadvantaged along the line of a positive discrimination strategy is likely to prove insufficient. In the presence of asymmetrical social structures, empowering underprivileged groups is the only solution, that is, mobilizing and organizing them in such a way that they can assert their rights to participate in decisionmaking even if that implies challenging existing social structures and antagonizing the elite. This is an arduous task that goes well beyond the usual understanding of CBD. As Brett (2003: 14–15) aptly notes:

> [P]articipatory systems are rarely a response to demands from local people who may well be locked into hierarchical and deferential structures, but rather promoted in response to western values imported by donors. This obliges local communities to develop different kinds of organization from those they have used in the past, thus demanding new skills and the ability to overcome local opposition if they are to succeed.

Therefore participatory development "cannot be treated as a process in which facilitators merely 'enable' local people to do what they would have wanted to do anyway" (Brett 2003: 15; in the same vein, see Platteau and Abraham 2002, 2004).

Given the kinds of incentive systems present, are large bureaucracies really equipped to perform such a delicate job? In this respect, the evidence is worrying. Project facilitators tend to easily fall prey to local elites either because they are in a rush to show results and therefore gloss over local power relations (Mansuri and Rao 2003), or because they are too weak to resist their pressure and the donor agency is not supporting them enough.

Elite Capture and Development Brokers

The problem of elite capture is especially serious, because donor agencies are enthusiastically rushing to adopt the participatory approach, either because they are eager to relieve poverty in the most disadvantaged countries or because they need rapid and visible results to persuade their constituencies or sponsors that the new strategy works well. Clearly such urgency runs against the requirements of effective CBD, because CBD cannot succeed unless it is based on genuine empowerment of the rural poor (see, for example, Edwards and Hulme 1996; Rahman 1993). If donor agencies do not spend the time required to ensure that the poor acquire real bargaining strength and organizational skills, "ownership" of projects by beneficiary groups is most likely to remain an elusive objective, such as observed in the case of the World Bank's social investment funds (Narayan and Ebbe 1997; Tendler 2000).

The perverse mechanism that risks undermining CBD is triggered by donor agencies' being tempted to skip the empowerment phase by asking intended beneficiaries to form groups or partner associations and to "elect" leaders to direct them (Platteau 2004; Platteau and Abraham 2002). As Esman and Uphoff point out (1984, p. 249):

> The most prominent members are invariably selected and then given training and control over resources for the community, without any detailed and extended communication with the other members about objectives, rights, or duties. Creating the groups through these leaders, in effect, establishes a power relationship that is open to abuse. The agency has little or no communication with the community except through these leaders. The more training and resources they are given, the more distance is created between leaders and members. The shortcut of trying to mobilize rural people from outside through leaders, rather than taking the time to gain direct understanding and support from members, is likely to be unproductive or even counterproductive, entrenching a privileged minority and discrediting the idea of group action for self-improvement.[4]

Confirming Esman and Uphoff's prediction, several studies conclude that the formation and training of village groups in community-based projects have the effect of encouraging the entry of wealthier and more educated people into leadership positions because of the attractiveness of outside funding (Gugerty and Kremer 1999, 2000; Rao and Ibanez 2001). In point of fact, a major problem confronted by the community development movement of the 1950s lay in its inability to effectively counter the vested interests of local elites (Holdcroft 1984). Being adept at representing their own interests as community concerns expressed in the light of project deliverables, local leaders often succeed in deluding donors into thinking that their motivations are guided by the collective good (Harrison 2002; Mosse 2001; Ribot

1996, 2002). Their demands are replete with the sorts of pleas and vocabulary that strongly appeal to donors, and to create the appearance of participation, they may go as far as spending resources to build community centers; hold rallies; and initiate showcase, labor-intensive activities (Conning and Kevane 2002).

In lineage-based societies, local chiefs and elders from dominant lineages are ideally positioned to capture the benefits of CBD projects. Instead of father figures clinging to their traditional duties of guaranteeing people's livelihoods, redistributing wealth, and settling conflicts in such a way as to maintain the existing social order, the erstwhile elite often become transformed into greedy individuals who show all the less restraint in enriching themselves at the expense of their communities, because they are actually legitimated by outside actors (Platteau 2004; Platteau and Abraham 2002).[5] As many NGOs working in Sub-Saharan Africa have experienced, local chiefs who are "elected" as representatives of their village community tend to require that any equipment or facilities made available through external assistance should benefit them as a matter of priority. When the aid agency concerned resists such a demand, the chiefs often succeed in concealing their misbehavior from the agency's scrutiny.

The mismanagement of aid transfers can obviously occur in class- or caste-based village societies in which landed elites use their dominant economic, social, and political position to appropriate for themselves whatever portion of the resources they want and let the poor have just the leftovers (Bardhan 2002; Conning and Kevane 2002; Sara and Katz 1997). In their study of a decentralized food-for-education program in Bangladesh, Galasso and Ravallion (forthcoming) find that the program was mildly pro-poor, in the sense that a somewhat larger fraction of poor than non-poor received benefits. They also find evidence of local capture, particularly in highly unequal or in remote villages.

The traditional elite are not the only category of people to benefit from the newly channeled resources because of their frequent involvement in tactical alliances with educated people and politicians operating outside the village domain. As I have emphasized elsewhere (Platteau 2004, p. 112; Platteau and Abraham 2002, p. 122), in Sub-Saharan Africa, "chiefs often co-opt new elites in their village 'associations,' for example, by creating neotraditional titles that they then sell to the new rich eager to acquire a political base in the countryside" (see Bayart 1989 and Geschiere 1994 for an illustration of these possibilities). Moreover, "the urban rather than the rural elite may be responsible for initiating the process that deflects CBD from its intended purpose." Witness the rapid multiplication of national NGOs that are created at the initiative of educated unemployed individuals, politicians, or state employees who may have been laid off or deprived of access to key logistical resources as a result of structural adjustment measures.[6] Acting as development brokers, political entrepreneurs have been quick to understand that "the creation of an NGO has become one of the best means of procuring funds from the international community" (Platteau 2004, pp. 112–13; see also Bierschenk, Chauveau, and de Sardan 2000). In the words of Chabal and Daloz (1999, pp. 22–24, 105):

> [A] large number of key political actors have now shifted their operations to the local level, which currently enjoys wide international favour and receives substantial assistance. . . . [A] massive proliferation of NGOs . . . is less the outcome of the increasing political weight of civil society than the consequence of the very pragmatic realization that resources are now largely channelled through NGOs. . . . Indeed, NGOs are often nothing other than the new "structures" with which Africans can seek to establish an instrumentally profitable position within the existing system of neo-patrimonialism. . . . Above and beyond the new discourse of NGO ideology . . . , the political economy of foreign aid has not changed significantly. The use of NGO resources can today serve the strategic interests of the classical entrepreneurial Big Man just as well as access to state coffers did in the past. . . . Furthermore, NGO-linked networks are inevitably intertwined with those emanating from the state.

Thus in the case of Benin, a West African country especially spoiled by donors, we learn that local NGOs and associations—which are often "empty shells established with the sole purpose of capturing aid" (*Le Monde* February 26, 2001)—have multiplied within a short time to number several thousands and many others are waiting to receive approval from the Ministry of the Interior. In non-African countries also, NGOs often constitute "an opportunistic response of downsized bureaucrats, with no real participation or local empowerment" (Conning and Kevane 2002, pp. 383–84), and inevitably program officers themselves become involved in the creation of community institutions (Bebbington 1997; Gray 1999; Meyer 1995). Such a risk is obviously high when self-conscious, organized local communities do not actually exist prior to the opening up of new development opportunities by state agencies or international donors (see Li 2001 for a well-documented illustration of this possibility), while the latter presume their existence on a priori grounds (McDermott 2001).

Of course, not all local leaders are opportunists ready to divert foreign aid from the intended beneficiaries. Several studies actually point to substantial variations in targeting effectiveness across villages (Jalan and Ravallion forthcoming; Ravallion 2000). Interestingly, investigators often find that inequality within villages is inversely related to this effectiveness (Galasso and Ravallion forthcoming), confirming the prediction derived from Bardhan and Mookherjee's (1999) political economy model and suggesting that the local elite tend to appropriate a larger share of the transfers in communities that are highly unequal to begin with.

Future evidence will settle the issue of the extent to which cases of embezzling behavior outweigh cases of leaders who are either unable (because of sufficient empowerment of the grassroots) or unwilling (because they somehow share the ethical code of aid agencies) to embezzle aid funds. Results are likely to vary from one region to another, depending on the strength of social movements, levels of rural literacy, and so on. Yet I believe that anecdotal evidence about the misdeeds of the local elite is plentiful enough, at least in poor countries such as those of Sub-Saharan Africa, to justify a cautious attitude about the possible impact of the community-driven development approach. While thinking about mechanisms aimed at keeping fraudulent behavior in check, I will therefore assume that leaders do not share the values of foreign aid agencies and that the grassroots are not empowered to dispute their decisions.

Before pursuing that line of thought, the next subsection provides some evidence of elite capture in the context of decentralized development experiences.

Misappropriation of Funds by Local Governments

In Indonesia, the new devolution system has resulted in a situation where the provincial regents

> exercise their new administrative and financial clout so imperiously that locals refer to them as "little kings." Stories abound of reckless extravagance or outright corruption. . . . [R]egents have simply seized companies belonging to the central government, or imposed arbitrary new rules on businesses. Fears of decentralization run amok are beginning to replace fears of Indonesia's disintegration (*The Economist* February 15–21, 2003, pp. 54–55).

In Nepal, decentralization created opportunities for elites to dominate decisionmaking at the local levels (Bienen and others 1990). In India, as testified to by one of the best-documented studies available regarding one of the most comprehensive attempts at decentralization (Kumar 2002), under current joint forest management arrangements the poor are net losers and are likely to remain so over a 40-year time horizon. If they participate in joint forest management, it is just to "state their loyalty to the village leadership" (Kumar 2002, p. 776). In behaving thus, "the poor ensure that they can partake of at least some village institutions, and they build up their stocks of social capital" (Kumar 2002, p. 776). In Bangladesh, under President Ershad's decentralization reforms, even though people were mobilized at the local level, elites of politically-based factions that controlled the local governments continued to make decisions about the allocation of resources (Westergaard and Alam 1995; see also Das 2000; Inbanathan and Gopalappa 2002; and Véron 2001 about the interference of patronage politics in local governance programs in Kerala and Karnataka states in India).

Turning to Sub-Saharan Africa, municipal bodies or rural councils in Senegal used the new prerogatives accorded them under a decentralization scheme to become involved in dubious dealings, such as sales of rural lands to touristic and other business interests, without consulting the communities concerned as they should have done (Mosse 2001). In Uganda, participatory planning appears to be "more a matter of form than substance." Indeed, local participation is reduced to a minimum, being "limited to counterfeit mechanisms of enfranchisement such as the 'Participatory Poverty Assessments' so alluring to Uganda's donors, which provide the desired facade of consultation" (Francis and James 2003, pp. 334–36). While important resources are channeled to local governments through conditional grants that leave little room for genuine participation by people in decisionmaking (such grants are essentially decided in a technocratic, top-down manner), other resources are made available through unconditional grants and locally generated revenue that create an ideal ground for the exercise of unfettered local patronage. Revealingly, Reinikka and Svensson (2001) find that during 1991–95, only 13 percent of the total flow of funds granted by the central government for nonwage expenditures in schools (for textbooks, instruction costs, and so on) actually reached the schools after having transited through local intermediary bodies. Most schools did not receive any of the funds.[7] In Cameroon, as soon

as they understood that new money opportunities were opened by the decentralized program of forestry management, urban elites manifested "a new interest" in their geographical origin. Once self-appointed or co-opted in the village councils, they "instantaneously establish alliances with term-based companies to whom they have promised their villages' forests" (Oyono 2004: 102).

What must be stressed is that the attitudes involved partake of the logic of clientelistic politics characteristic of the African continent (and of poor countries, such as Bangladesh, Haiti, and Nepal). For those at the very bottom of the social order, "the material prosperity of their betters is not itself reprehensible so long as they too can benefit materially from their association with a patron linking them to the elites" (Chabal and Daloz 1999, p. 42). As a result, abuses of power are tolerated so long as patrons are able to meet the demands made by their clients, who are concerned above all with ensuring their daily livelihoods. Ultimately, it is because they overlook the genuine nature of the links between elites and commoners, rulers and ruled in poor areas that international donor agencies overestimate the capacity of CBD to deliver development gains more effectively and more equitably than previously applied strategies.

As Platteau (2004, p. 114) points out, a rush to embrace CBD not only entails the risk of creating and reinforcing an opportunistic, rent-seeking elite, but

> it also involves a serious bias in the selection of communities. Indeed, communities within easy reach tend to be privileged, but they are not the most needy precisely because of their easy accessibility. They are better off because they have good access to markets, education facilities, and all sorts of information.

Note that their advantage in attracting donor funds under participatory programs does not lie only in comparatively low transportation and other transaction costs but also in their greater ability to set up an appropriate collective structure and "elect" a leader who speaks a foreign language (Platteau 2004).

The Difficult Challenge of Disciplining Local Leaders

We are now in a position to present a tighter framework to analyze possible mechanisms for disciplining local leaders and to identify key factors bearing on donors' capacity to achieve that end.

Aid Agencies as Local Monopolies

What I will now argue is that a massive injection of aid funds for participatory development will most likely corrupt the very process of community development that these funds are intended to promote. Let us first consider the situation in which the supply of aid money for participatory development is scarce. The aid agencies concerned can then be considered to be in a monopolistic position in relation to the community with which they establish a relationship. Platteau and Gaspart (2003a) develop a model analyzing the problem of a donor or aid agency (henceforth designated by A) that contemplates providing funds to a particular community or grassroots group (designated by G) that does not have any alternative funding

possibility (see also Platteau 2004 for a first sketch of the main argument). No direct contact takes place between A and G, however, because A deals with a local leader or intermediary (designated by L) acting on behalf of the intended beneficiaries. Typically, L has organized G into an development association and has been "elected" president.

What A can do, however, is to check whether G genuinely supports L by, say, asking G to confirm that L is its authentic leader, whether through a formal voting procedure or otherwise. As a matter of principle, A will not disburse funds through L unless it has received such a confirmation. A does not observe how the money is being shared within the target community, but A acts strategically, taking the behavior of L into account when making its decision to support or not to support the community.

As for the strategic interactions between L and G, they are assumed to be deeply asymmetrical and are represented by an ultimatum game. That is, L has the first move and makes an offer to G regarding the apportionment of the aid fund. Then G must say whether it accepts the offer or not, knowing that its rejection would mean the collapse of the group consensus required to receive aid from A. In such a game, as is well known, it is in the interests of the second mover (G) to accept the proposal made by the first mover (L), and it is therefore in L's interests to set the share accruing to G at as low a level as possible. This is because G does not wield sufficient leverage to dispute the self-asserted right of L to appropriate a large share of the aid proceeds. Indeed, as illustrated in African case study material described elsewhere (Platteau and Gaspart 2003b), G may not resent L's disproportionate share insofar as its own situation is simultaneously improved, however small that improvement may be.

This is obviously a depressing result in view of the objectives of CBD. Therefore some mechanism that could discipline local leaders in the absence of democratic governance within target groups or communities would be useful. Sequential disbursing of aid money is obviously such a mechanism. Instead of releasing money in a single shot, aid funds would be disbursed in successive tranches, with the disbursement of each of them being conditioned on L's proper behavior regarding the use of previous tranches. Inherent in such a strategy is the recourse to a fraud detection technology without which local leaders would not be incited to behave. Detection is necessarily costly, yet it is in the interests of an aid agency to incur the related expenses, because it can thereby hope to better achieve its own objective of poverty relief (fraud detection is incentive compatible).

As we know from repeated game theory, however, as long as the duration of the game is finite, and no matter how high the number of periods in the game, the equilibrium outcome will be the same as that obtained in the one-period game (Kreps 1990; Kreps and Wilson 1982). The effort, including the monitoring resources, the donor agency expends during the successive stages of the project will be of no avail. Assuming that the local leaders are selfishly rational, they will embezzle funds from the very beginning; knowing that, aid agencies should refrain from disbursing even the first tranche of money. At least, this will happen if donors cannot credibly commit themselves to releasing money even when fraud remains undetected. Of course, if aid agencies interact with communities over an

infinite (or indeterminate) period of time, this awkward result can be avoided. But this is hardly a consoling thought inasmuch as CBD aid, in particular, is precisely aimed at making communities self-supporting after a certain period of time and the limited duration of the external intervention is better made clear from the beginning. (It is only in the context of decentralized development, when aid transfers to local governments are anchored in a framework of fiscal decentralization, that there would be an endless round of disbursement periods creating the conditions of an infinitely repeated game.)

That said, the assumption of strategic rationality underlying the foregoing reasoning is questionable. This is not only because actors may not perfectly anticipate the future consequences of their actions and the reactions of others, or because they may entertain doubts about the rationality of the people with whom they interact (in which case we know that even in a finitely repeated game, cooperation may be established as an equilibrium), but also because some social norms may exist that have the effect of constraining rational calculations.

The existence of a norm of intertemporal fairness among the members of G may thus make gradual, conditional disbursing of aid money effective even in the context of finitely repeated interactions between A and L. The reason becomes evident if such a norm dictates that a division rule adopted during one period may not be changed at will by L during a later period, especially if the change is made at the expense of G. In other words, L is not allowed to reduce the share of aid transfers accruing to G over the successive stages of a project. In a two-period CBD game, L will thus be unable to strategically lower the share allotted to G between the first and the second rounds. As a result, because the granting of the second tranche is conditional upon L's proper behavior in the previous round, and because the probability of fraud detection can be assumed to increase with the extent of the embezzlement, the portion that L grants to G will be the minimum share compatible with an acceptably low risk of detection at the end of the first round, and this share will be applied again during the second round. Clearly the norm of intertemporal fairness serves the purpose of conferring genuine bargaining power upon G during the second round.

This is not sufficient, however. For the mechanism to be effective, G must also be able to perfectly enforce L's promise to pay the agreed share of the aid transfer once A has released the money. The story told elsewhere (Platteau and Gaspart 2003b) about a CBD project in a Sahelian country seems to attest that enforcement is not the real problem: even though their leader embezzled substantial amounts of aid money, villagers did not feel cheated and actually voted for the predatory leader again even after his malpractices had been fully revealed and he had confessed to them. It therefore appears that G must be empowered enough to enforce L's promise, but not enough to actively debate the sharing rule with him or her. If G were not empowered enough even in the first sense, it would be doomed to be seriously exploited by L and little could be done to relieve G's poverty until, through a time-consuming process of becoming more conscious of its rights and through learning processes, it becomes better able to defend its rights and effectively participate in decisionmaking. However, if G were empowered enough in both senses, the

sharing rule would be determined as the outcome of a bargaining process between *L* and *G* and not by *L* only.[8]

In the two-period, game-theoretical model proposed by Platteau and Gaspart (2003a), *A*, which is concerned with relieving poverty, decides how to allocate the available aid budget between two successive periods, as well as the amount of monitoring expenses on which the effectiveness of fraud detection partly depends. Given the amounts of the first and second aid tranches as well as the size of the monitoring effort made by *A*, *L* chooses the share of the aid transfers that he or she will hand over to *G*, among whom a norm of intertemporal fairness is known to prevail. While making its decisions, *A* faces the following trade-off. On the one hand, *A* would like to disburse as much money as possible during the first period because it is impatient to see the poverty of *G* alleviated. On the other hand, *A* wants to defer its disbursement of aid until the second period, because late payments serve to discipline *L*. Indeed, the higher the amount of the second tranche relative to that of the first, the more *L* is encouraged to use the aid transfers according to *A*'s prescriptions, that is, for the benefit of *G*.[9]

One important result derived from the comparative static of the model is the following: the more impatient the donor agency, that is, the more *A* discounts the benefits enjoyed by the target population during the second period, the smaller the amount of the second aid tranche relative to that of the first tranche, and the lower the share accruing to *G*. In other words, because the subjective cost of waiting is higher, *A* is less ready to use the leader-disciplining mechanism and to postpone the disbursement of aid funds. As a consequence, *L* is less effectively induced to behave during the initial period, holding monitoring expenditures constant. At the new equilibrium, however, the amount of these expenditures is being increased. The net effect of these two opposite forces is shown to be detrimental to *G*: the share appropriated by *L* increases and the absolute amount of aid money that will accrue to *G* if *A* does not detect any fraud is smaller. Bear in mind that monitoring expenses, which have been increased to substitute for the smaller use of the conditional mechanism of aid disbursement, are to be subtracted from the gross aid budget before transfers to *G* are made.

Wishing for rapid results in the struggle against poverty is therefore counterproductive. Indeed, the main effect would be to enrich and consolidate local elites, much in the same way as windfall incomes from natural resources can be a curse because they give rise to greater rent-seeking activity (see, for example, Tornell and Lane 1998). At the limit, if *A* is extremely impatient, the share accruing to *G* will tend to a value as low as that obtained under a one-shot disbursement procedure.

A second interesting comparative static result is that the higher the cost of recycling aid funds—or the smaller the proportion of aid money earmarked for the second tranche that can be costlessly redirected to another group or association in the event of fraud being detected in the initial project—the lower the relative amount of the second aid tranche, the smaller the share accruing to *G*, and the lower the amount of aid money accruing to *G* in the absence of fraud detection. In other words, if reallocating funds intended for a particular project is difficult, say, because of larger set-up costs, a donor agency will find that deferring its disbursement is less attractive

in equilibrium. As a result, *L* will be in a better position to appropriate the aid money. Donors may therefore be tempted to avoid working in low-density and remote areas where high set-up costs arising from long distances to be traveled, low education levels, and so on tend to reduce the effectiveness of their efforts to reach the poor.

Clearly, the logic underlying the effects of a rise in the cost of recycling aid funds is the same as that obtained for an increase in the discount rate of future benefits accruing to the poor. This is not surprising, inasmuch as in both cases the cost of using the leader-disciplining mechanism is higher, and the aid agency is therefore induced to disburse its available funds more quickly.

Competition among Donor Agencies Using Bilateral Reputation Mechanisms

If, in complete contrast to the foregoing situation, competition among donor agencies is perfect in a context characterized by an abundant supply of aid funds, the grassroots will only get crumbs. This is the worst scenario: aid agencies compete for access to communities and, in their attempts to lure the local leaders "representing" the grassroots, they are ready to drop their safeguards against the appetite of these leaders. Money is disbursed quickly without paying much attention to the manner in which it is shared between *L* and *G*.

A less pessimistic scenario arises if we consider, perhaps more realistically, that aid agencies do not produce a homogeneous service but differentiated, multi-attribute services that comprise the total aid budget on offer, the timing of its disbursement over successive tranches, and the monitoring effort. Monopolistic competition would then prevail among donor agencies and Platteau and Gaspart's analytical framework could be adjusted accordingly. This would imply that an exit option now exists for *L,* and that two critical parameters of the model, namely, *A*'s intertemporal preference and the cost of recycling aid funds, can be reinterpreted as possibly reflecting the intensity of prevailing competition among donor agencies. Regarding these latter two factors, the comparative static of the model indicates that acute competition is an unambiguously regrettable feature of the aid environment. Indeed, by driving aid agencies to disburse funds quickly to prevent rival agencies from destabilizing a particular aid supply relationship, and by increasing the cost of recycling funds in the event of fraud detection, acute competition causes the share of aid funds that local intermediaries appropriate to increase at the expense of the intended beneficiaries. In other words, intermediaries can skillfully play on inter-agency competition, because they know both that aid agencies are keen to find partners through whom to channel their aid budget and that the budget is more or less tied to the initially chosen project or community.

It is also evident that the emergence of exit options following the multiplication of aid agencies has the effect of raising the share that local leaders are allowed to appropriate at equilibrium, that is, the share that will deter them from pursuing a shifting strategy. A shifting strategy is a strategy whereby *L* does not care about staying with the same agency over the whole course of the aid project, because he or she is ready,

if caught cheating, to shift to another agency and start cheating again. What is at work here is a so-called bilateral reputation mechanism: if caught embezzling funds, a local leader is punished only by the aid agency that has actually provided the funds embezzled.

The elite capture problem can apparently be mitigated in two ways. Reducing competition by concentrating the supply of aid in the hands of fewer agencies is the first way. Indeed, by diminishing the exit options available to local intermediaries, especially if aid agencies are geographically specialized, such concentration in the market for aid would have the same effect as a reduction in the aggregate supply of aid. When projects are geographically concentrated, the presence of scale economies in the technology of fraud detection would constitute an additional advantage of this first solution. The second solution consists of a coordination mechanism whereby aid agencies would inform each other about fraudulent acts committed by intermediaries.[10] While such a device, known as a multilateral reputation mechanism (MRM), is apparently more feasible than reducing competition, it is not devoid of serious practical difficulties, as the following discussion indicates.

Competition among Donor Agencies Using Multilateral Reputation Mechanisms

Greif (1989, 1994) documents the MRM with respect to relationships between traders (see also Aoki 2001, chapter 4; Platteau 2000, chapter 6). Applied to our problem, the mechanism would work as follows. Operating within a repeated game framework, a donor agency would adopt a strategy whereby it grants money to a local leader, but only provided that he or she is not known to have cheated another agency at some time in the past. If the money is disbursed and the benefiting leader is later found to have cheated the agency, the latter dutifully reports the fraud and communicates the name of the dishonest leader to other members of the donor community. Before embezzling funds, a leader would thus be incited to think twice, because by cheating today he or she would spoil his or her reputation for future interactions with the entire donor community. The multilateral reputation strategy can be shown to be an equilibrium strategy. That is, if a leader expects every donor agency to adopt such a strategy, his or her interest is to share aid funds equitably among the intended project beneficiaries. Knowing that reaction, the interest of all donor agencies is to cling to the multilateral reputation strategy. Honest behavior therefore becomes established as a (Nash) equilibrium.

However, several problems are associated with the MRM.[11] The first arises because the information conditions that must be fulfilled for it to work are extremely stringent: information must circulate perfectly between donor agencies. This is unlikely to be the case in reality, because there are many donor agencies, they are scattered around the industrial world, and they are extremely heterogeneous in terms of several key characteristics (size, ideology, methods, time horizon, and so on). These are hardly ideal conditions for the existence of a dense information network.

Is the establishment of a private third party charged with centralizing information (as suggested, for example, in the Law Merchant system analyzed by Milgrom, North, and Weingast 1990) the solution to the problem caused by the costliness of generating and communicating information? Such a system can work effectively only if donors have an incentive to detect fraud and report fraudulent experiences to the third party, so that the blacklist of dubious intermediaries is exhaustive and regularly updated (otherwise donors would not be induced to consult it). Yet, insofar as the detection and reporting of a fraud once it has occurred entails costs but brings no benefits to the individual agency that has been cheated, such an incentive does not exist unless, of course, donor agencies are so genuinely committed to the cause of poverty relief that they are not concerned about whether they or another aid agency are the ones to reduce poverty. (The critical argument of a donor's utility function is then the extent of general poverty relief rather than the relief accomplished by its own efforts.)

To create the adequate incentive, the third party should be able to exercise pressure on the detected fraudulent leader so as to make him or her return the stolen money. A provision that donors would not be entitled to use the system to obtain compensation unless they appropriately queried the third party about the reliability of their current partners would also make it in donors' interests to query about past dealings of partner-leaders before disbursing money. As a result, so the theory goes, the threat against potential leaders would be effective, and, if caught, fraudulent leaders would be prompted to comply with the third party by returning the stolen money so that their names would be removed from the blacklist. This said, Milgrom, North, and Weingast (1990) nevertheless show that honesty will be established as a (symmetric sequential) equilibrium under the foregoing mechanism only if a number of conditions are met, in particular, the cost of information querying, the cost of appeal to the third party, and the cost for the latter to recover stolen money from fraudulent local leaders ought not to be too high. Unfortunately, these assumptions are likely to be violated, especially because the headquarters of aid agencies are located at great distances from one another and because the cost of acquiring information is high, including evidence of fraud in the opaque context of alien cultural environments. The mechanism is therefore not self-enforcing.

A second problem lies in the fact that local leaders may not be concerned about preserving their reputations, because their time horizons are short and they could be quite happy with running away with the money stolen from a single project. In other words, the payoff from dishonest behavior is so large compared with the payoff from honest behavior that honesty cannot be induced at equilibrium.

Finally, one key actor has been missing from the foregoing discussion, namely, the ultimate purveyors of funds (designated as *P*) from whom donor agencies obtain their financial resources. These are taxpayers for national and international organizations or taxpayers and the general public mobilized during fund-raising campaigns for NGOs. These ultimate purveyors of funds create a further link in the game, giving rise to a new space of strategic relationships between donor agencies and themselves. The MRM becomes seriously dysfunctional if donors expect their ultimate sponsors

to react negatively to news of embezzlement in their projects, for instance, through the revocation of funds (Ebrahim 2003). In these circumstances, a donor organization has no incentive to report the acts of malfeasance detected in its projects. This is because it may entertain the hope that other agencies will candidly reveal their own bad experiences or because it fears that if it did convey such information, others might not have done so and would then exploit the situation to their own advantage. That this risk is real is evident from the atmosphere of secrecy that surrounds the activities of many donor organizations, including NGOs. The only way to reduce such risk is to improve the general public's understanding of CBD processes and the possibility of failures, so that honest donors that openly admit to cases of cheating are not unfairly sanctioned to the benefit of more opportunistic ones.

Central funding bureaucracies, such as the European Community or the cooperation administrations of national governments, rather than the scattered contributors to fund-raising campaigns organized by NGOs, could help tame the opportunism of local leaders through indirect measures aimed at donor agencies. One way to achieve coordination would be to introduce a rating of donor or aid agencies that these bureaucracies would systematically use to decide which agencies deserve to be financially supported. But once again, things do not look quite that simple once the question of a yardstick on which to base the rating is raised. Resorting to measures of output, such as improvements in the standard of living of the poor in the communities chosen, is an ideal procedure but is likely to be too costly to be feasible, especially in the case of NGOs with their typically diverse and long-term objectives (see Edwards and Hulme 1996; Ebrahim 2003). Moreover, such measures could introduce biases in rated agencies' selection of communities. Indeed, the agencies would be induced to choose communities in which poverty could be more easily reduced for reasons other than the prevailing power structure (for example, easy accessibility).

The foregoing discussion suggests another, more feasible criterion, namely, the disbursement and monitoring procedures used by donor agencies, as well as the duration of their CBD projects. In this perspective, self-reported cases of fraud detection could be considered as indirect evidence of the effectiveness of monitoring activities rather than as signs of failure. Not only are such characteristics relatively easy to observe, but they also offer the advantage of not creating perverse incentives for the rated agencies.[12]

The need for a proper evaluation of aid agencies is all the more pressing, as careless organizations that do not implement sequential disbursement mechanisms with a view to disciplining local leaders exist alongside serious agencies. The former tend to disburse funds quickly, either because they do not have a proper understanding of the one-period game thus being played,[13] or because they are not single-mindedly pursuing the objective of poverty alleviation. For example, despite all their pro-poor rhetoric, they are also concerned with reproducing themselves as job- and income-providing organizations in the West. The presence of careless organizations further complicates the problem of elite capture, not only because it increases the exit options available to local intermediaries, but also because it makes the establishment of an MRM among all donor agencies impossible. In the same way that bad money

chases good money, the operation of these opportunistic aid agencies risks driving good agencies out of business or forcing them to relax or altogether give up their gradual and conditional disbursement procedures. Indeed, if offered the choice, local leaders will normally prefer to work with bad agencies, and if the latter are numerous enough, good agencies will be unable to attract partner communities unless they soften their approach to aid disbursement.

Competition among Local Leaders

Platteau and Gaspart (2003a) also consider the possibility of aid agencies relying on competition among local leaders in order to discipline them. Assuming the presence of two village leaders with unequal leadership skills, they show that leader competition makes the two-stage reputation mechanism analyzed earlier unnecessary. Note, however, that as long as the competing leaders are not equally proficient, some elite capture will subsist in equilibrium, regardless of *A*'s willingness to effectively reach *G*. The wider the gap between the competences of the two leaders, the greater the misappropriation observed under the competitive equilibrium.

Moreover, and more important, whenever several competing leaders are present there is a serious risk of collusion between them. If the candidates do effectively collude, the leader-disciplining mechanism becomes necessary again to present significant exploitation of *G*. If collusion is not feasible because of intense rivalry between leaders, the negative externalities of a mechanism that fosters competition rather than cooperation among elites are possible shortcomings of that mechanism. The existence of such a dilemma—not particularly good relationships between local leaders are necessary for the competitive mechanism to be effective, yet they are a liability that threaten collective action at the village or community level—may undermine the case for relying on competition among elites as a way to protect the poor's entitlement to external assistance.

When the foregoing dilemma does not exist, channeling aid through several local organizations or groupings that are potentially competing with each other may prove a more reliable or less costly solution to the elite capture problem than the leader-disciplining mechanism. Empowering the poor and disadvantaged by relying on competition within a community is also less likely to arouse hostile reactions from the elites. Indeed, a monolithic elite will likely try its best to counter empowering attempts by outsiders in a strategic effort to preserve its own power, for instance, by criticizing actions that empower the local poor as constituting undue interference with the recipient country's politics.

Conclusion

CBD must be accompanied by an empowerment strategy that is necessarily time-consuming and requires slow disbursement procedures, as well as by widespread training efforts, especially in literacy and organizational skills. Sequential and conditional release of aid funds is a useful approach to participatory development.

It obviously implies that fraud can be detected ex post, which requires donor agencies to devote substantial resources to project monitoring, thereby substituting external control for missing internal democratic governance. If things go well, the hope is that during the process the poor will gradually learn to better defend their rights; monitor the actions of their leaders; compel their leaders to follow through on their promises; and spawn new, alternative leaders who can compete with the existing elite.

Yet the aid environment must be conducive to effective application of conditional disbursement procedures. If competition among donor agencies is too stiff and they are not linked by any coordinating, information-sharing mechanism, the interests of the poor are unlikely to be well protected. This is because of at least three reasons. First, when numerous aid agencies offer an ample supply of funds the agencies tend to rush to help communities for fear that they may be overtaken by rival agencies in their search for good partners. Gradual disbursement procedures will be shunned if the expectation prevails that rival agencies are disloyal competitors ready to disburse money immediately and unconditionally. Second, competition among aid agencies raises the cost of recycling funds for those that have detected fraud in one of their projects. Third, in the presence of stiff competition, preventing local leaders from following an opportunistic strategy of shifting donor partners is more difficult, and this is especially true if some aid agencies are careless. Thus more competition implies that a smaller share of aid proceeds will accrue to the poor.

Unfortunately, pressures to spend money on poverty alleviation through CBD are mounting, especially because many aid agencies are finding it difficult to use the entirety of their available budgets given the lack of effective aid absorption on the part of poor countries.[14] If many aid agencies rush in with a lot of money in hand in a frantic search for communities and community-based projects, their efforts will have the effect of reinforcing the privileges of local elites, who will thereby gain increasing legitimacy from the outside world rather than from their own people. Moreover, the agencies will help to create an unhealthy situation in which an excessively high value is placed on the skills needed to relate to the donor community, skills that tend to be heavily concentrated among a narrow educated class. Outside money can therefore corrupt the process of local institutional development by allowing leaders to eschew negotiation with members in matters pertaining to support and material contributions, thereby preventing autonomous organization building based on the leaders' accountability in relation to community or group members.

Notes

1. As Ebrahim (2003, p. 825) points out: "For a sector that views itself as largely mission-driven, there is an urgent need to take performance assessment seriously in order to justify activities with substantiated evidence rather than with anecdote or rhetoric."

2. Galasso and Ravallion (forthcoming) have likewise assumed that a community is maximizing a positively weighted sum of utilities featuring the situation of two population groups, poor and non-poor. Communities are therefore assumed to be able to achieve an

efficient allocation of the resources put at their disposal by a central agent (the so-called project office), which does not observe how much is going to the poor in each area but takes the behavior of communities into account while setting the budget allocation between them. The weights on the utilities of the poor and the non-poor are interpreted as capture coefficients arising endogenously in a probabilistic voting model with differences in voter information between the poor and the non-poor. The authors also postulate that the weights depend on characteristics of the poor and non-poor, as well as on the local political and economic environment and the program itself.

3. On the basis of data collected on 132 community-maintained infrastructure projects in northern Pakistan, Khwaja (2002) has shown that increased community participation positively effects performance for non-technical project decisions yet has the opposite effect for technical decisions. Infrastructure maintenance is also better in non-complex projects, or in those made as extensions of old ones.

4. In light of this diagnosis, Cernea's (1988, p. 10) contention that "NGOs insert themselves not as a third and different/independent actor, but as an emanation and representation of the community" appears almost surrealistic.

5. In some areas, they have been accustomed to doing just that since colonial or precolonial (slavery) times (Bayart 1989).

6. In the Philippines, Ebrahim (2003) estimates the number of NGOs at more than 60,000!

7. Note that problems of misaligned incentives, conflicts of objectives between higher-level principals and middle-level agents, manipulation of the rules of the games, and so on are not confined to developing countries. A recent example is the Job Training Partnership Act in the United States (Courty and Marschke 1997).

8. Provided that the bargaining strength of G is strong enough, disciplining L with the help of a phased process of aid disbursement will not have the effect of raising the share of aid money accruing to the intended beneficiaries. To achieve its objective, the aid agency could therefore rely on the bargaining strength of the latter. To be sure, some embezzlement would still occur, but the agency would not be able to do better by using such a phased mechanism.

9. Note that the amount granted under the first tranche must be large enough to ensure that L's behavior can be effectively tested before making a decision about whether or not to disburse the second tranche.

10. For a discussion of alternative coordination mechanisms, such as codes of conduct, ombudsmen, social auditing, and accreditations, see Ebrahim (2003).

11. I ignore the awkward problem that, in order to counter leaders' temptation to embezzle funds, donors should, in theory, give them a flow payment or rent each period, and this flow should be at least equal to the interest on the one-off embezzlement of stock they could carry out.

12. This is the opposite of evaluations that reward success while punishing failure, because those encourage NGOs to exaggerate successes and to refrain from revealing failures (Edwards and Hulme 1996).

13. Imperfect knowledge of the game typically arises because aid agencies tend to underestimate the leverage of the local leader within the group, overestimate his or her degree of altruism as a result of the leader's cunning ability to deceive them, or are simply naive.

14. During the 1990s, the European Union's commitments for overseas development assistance exceeded its gross disbursements by more than US$1.6 billion each year, peaking at US$2.2 billion in 1994 (Heller and Gupta 2002). In particular, in 1996–97, £4.5 million of the budget of the United Kingdom's Department for International Development for

Africa was unallocated. In 2000–01, this figure rose to £18 million (*The Economist* November 2–8, 2002, p. 39). As all agencies seriously concerned with genuine development know, the scarcity of good projects and reliable groups and associations is probably the most important constraint to the effectiveness of aid programs.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Adams, Alayne M., Timothy G. Evans, Rafi Mohammed, and Jennifer Farnsworth. 1997. "Socioeconomic Stratification by Wealth Rankings: Is It Valid?", *World Development* 25(7): 1165–72.

Agrawal, Arun. 1999. *Decentralization in Nepal: A Comparative Analysis.* Oakland, Calif.: Institute for Contemporary Studies Press.

Aoki, Masao. 2001. *Toward a Comparative Institutional Analysis.* Cambridge, Mass.: MIT Press.

Asthana, Anand. 2003. "Decentralisation and Supply Efficiency: The Case of Rural Water Supply in Central India." *Journal of Development Studies* 39(4): 148–59.

Baland, Jean-Marie, and Jean-Philippe Platteau. 1996. *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* Oxford, U.K.: Clarendon Press.

Bardhan, Pranab. 2002. "Decentralization of Governance and Development." *Journal of Economic Perspectives* 16(4): 185–205.

Bardhan, Pranab, and Dilip Mookherjee. 1999. "Relative Capture of Local and Central Governments: An Essay in the Political Economy of Decentralization." Working Paper. Boston University, Department of Economics.

_____. 2000a. "Capture and Governance at Local and National Levels." *American Economic Review* 90(2): 135–39.

_____. 2000b. "Decentralizing Anti-Poverty Program Delivery in Developing Countries." Working Paper. Boston University, Department of Economics, Boston.

_____. 2002. "Corruption and Decentralization of Infrastructure Delivery in Developing Countries." Working Paper. Boston University, Department of Economics, Boston.

Bayart, Jean-François. 1989. *L'Etat en Afrique.* Paris: Fayard.

Bebbington, Tony. 1997. "New States, New NGOs? Crises and Transitions among Rural Development NGOs in the Andean Region." *World Development* 25(11): 1755–65.

Bergeron, Gilles, Saul Sutkover Morris, and Juan Manuel Banegas. 1998. "How Reliable Are Group Informant Ratings? A Test of Food Security Ratings in Honduras." *World Development* 26(10): 1893–1902.

Bienen, Henry, Devesh Kapur, James Parks, and Jeffrey Riedlinger. 1990. "Decentralization in Nepal." *World Development* 18(1): 61–75.

Bierschenk, Thomas, Jean-Pierre Chauveau, and Jean-Pierre Olivier de Sardan, eds. 2000. *Courtiers en développement: Les villages africains en quête de projets.* Paris: Karthala.

Birch, Isobel, and Halima A. O. Shuria. 2001. *Perspectives on Pastoral Development.* Oxford, U.K.: Oxfam Development Casebooks.

Brett, E. A. 2003. "Participation and Accountability in Development Management." Journal of Development Studies 40(2):1–29.

Carroll, Thomas. 1992. *Intermediary NGOs: Supporting the Link in Grassroots Development*. West Hartford, Conn.: Kumarian Press.

Cernea, Michael. 1985. *Putting People First: Sociological Variables in Rural Development*. Oxford, U.K.: Oxford University Press.

_____. 1988. *Nongovernmental Organizations and Local Development*. Discussion Paper no. 40. Washington, D.C.: World Bank.

Chabal, Patrick, and Jean-Pascal Daloz. 1999. *Africa Works. Disorder as Political Instrument.*" Oxford, U.K.: James Currey.

Chambers, Robert. 1983. *Rural Development: Putting the Last First.* London: Longman.

Conning, Jonathan, and Michael Kevane. 2002. "Community-Based Targeting Mechanisms for Social Safety Nets: A Critical Review." *World Development* 30(3): 375–94.

Courty, Pascal, and Gerald Marschke. 1997. "Measuring Government Performance: Lessons from a Federal Job-Training Program." *American Economic Review Papers and Proceedings* 87(2): 383–88.

Das, M. K. 2000. "Kerala's Decentralised Planning: Floundering Experiment." *Economic and Political Weekly* 35(49): 4300–03.

Ebrahim, Alnoor. 2003. "Accountability in Practice: Mechanisms for NGOs." *World Development* 31(5): 813–29.

Edwards, Michael, and David Hulme, eds. 1996. *Nongovernmental Organisations: Performance and Accountability: Beyond the Magic Bullet*. London: Earthscan.

Esman, Milton J., and Norman T. Uphoff. 1984. *Local Organizations: Intermediaries in Rural Development*. Ithaca, N.Y.: Cornell University Press.

Eversole, R. 2003. "Managing the Pitfalls of Participatory Development: Some Insight from Australia." *World Development* 31(5): 781–95.

Foster, Andrew D., and Mark R. Rosenzweig. 2002. "Democratization, Decentralization, and the Distribution of Local Public Goods in a Poor Rural Economy." Harvard University, Kennedy School of Government, Cambridge, Mass.: Processed.

Francis, Paul, and Robert James. 2003. "Balancing Rural Poverty Reduction and Citizen Participation: The Contradictions of Uganda's Decentralization Program." *World Development* 31(2): 325–37.

Galasso, Emanuella, and Martin Ravallion. Forthcoming. "Decentralized Targeting of an Anti-Poverty Program." *Journal of Public Economics*.

Geschiere, Peter. 1994. "Parenté et argent dans une société lignagère." In Jean-François Bayart, ed., *La réinvention du capitalisme*. Paris: Karthala.

Gray, M. L. 1999. "Creating Civil Society? The Emergence of NGOs in Vietnam." *Development and Change* 30(4): 693–713.

Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 49(4): 857–82.

_____. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102(5): 912–50.

Gugerty, Mary Kay, and Michael Kremer. 1999. *The Impacts of Development Funding on Social Capital: The Kenya Local Community Action Project*. Washington, D.C.: World Bank.

_____. 2000. "Outside Funding of Community Organizations: Benefiting or Displacing the Poor?" Working Paper no. 7896. National Bureau of Economic Research, Cambridge, Mass.

Harragin, Simon. Forthcoming. "Relief and an Understanding of Local Knowledge: The Case of Southern Sudan." In Vijayendra Rao and Michael Walton, eds., *Culture and Public Action: A Cross-Disciplinary Dialogue on Development Policy.* Stanford, Calif.: Stanford University Press.

Harrison, Elizabeth. 2002. "The Problem with the Locals: Partnership and Participation in Ethiopia." *Development and Change* 33(4): 587–610.

Heller, Peter S., and Sanjeev Gupta. 2002. "More Aid: Making It Work for the Poor." *World Economics* 3(4): 131–46.

Hirschman, Albert O. 1984. *Getting Ahead Collectively: Grassroots Experiences in Latin America.* New York: Pergamon Press.

Holdcroft, Lane. 1984. "The Rise and Fall of Community Development, 1950–65: A Critical Assessment." In Carl Eicher and John M. Staatz, eds., *Agricultural Development in the Third World.* Baltimore, Md.: The Johns Hopkins University Press.

Inbanathan, Anand, and D. V. Gopalappa. 2002. "Fixers, Patronage, 'Fixing,' and Local Governance in Karnataka." Working Paper no. 12. Institute for Social and Economic Change, Bangalore, India.

Isham, Jonathan, Deepa Narayan, and Lant Pritchett. 1995. "Does Participation Improve Performance? Establishing Causality with Subjective Data." *World Bank Economic Review* 9(2): 175–200.

Jalan, Jyotsna, and Martin Ravallion. Forthcoming. "Estimating the Benefit Incidence of an Anti-Poverty Program by Propensity-Score Matching." *Journal of Business and Economic Statistics.*

Khwaja, Asim Ijaz. 2002. "Can Good Projects Succeed in Bad Communities? Collective Action in Public Good Provision." Working Paper. Harvard University, Cambridge, Mass.

Kreps, David. 1990. *A Course in Microeconomic Theory.* New York: Harvester Wheatsheaf.

Kreps, David, and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27(2): 253–79.

Kumar, Sanjay. 2002. "Does 'Participation' in Common Pool Resource Management Help the Poor? A Social Cost-Benefit Analysis of Joint Forest Management in Jharkand, India." *World Development* 30(5): 763–82.

Kumar, Sanjay, and Stuart Corbridge. 2002. "Programmed to Fail? Development Projects and the Politics of Participation." *Journal of Development Studies* 39(2): 73–103.

Laurent, Pierre-Joseph. 1998. *Une association de développement en pays Mossi: Le don comme ruse.* Paris: Karthala.

Li, Tania Murray. 2001. "Boundary Work: Community, Market, and State Reconsidered." In Arun Agrawal and Clark Gibson, eds., *Communities and the Environment: Ethnicity, Gender, and the State in Community-Based Conservation.* New Brunswick, N.J.: Rutgers University Press.

Mansuri, Ghazala, and Vijayendra Rao. 2003. *Evaluating Community Driven Development: A Review of the Evidence.* Washington, D.C.: World Bank, Development Research Group.

McDermott, Melanie. 2001. "Invoking Community: Indigenous People and Ancestral Domain in Palawan, the Philippines." In Arun Agrawal and Clark Gibson, eds., *Communities and the Environment: Ethnicity, Gender, and the State in Community-Based Conservation.* New Brunswick, N.J.: Rutgers University Press.

Meyer, C. A. 1995. "Opportunism and NGOs: Entrepreneurship and Green North-South Transfers." *World Development* 23(8): 1277–89.

Milgrom, Paul, Douglass North, and Barry Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics and Politics* 2(1): 1–23.

Mosse, David. 2001. "People Knowledge, Participation, and Patronage: Operations and Representations in Rural Development." In Bill Cooke and Uma Kothari, eds., *Participation: The New Tyranny?* London and New York: Zed Books.

Narayan, Deepa, and Katrinka Ebbe. 1997. *Design of Social Funds: Participation, Demand Orientation, and Local Organizational Capacity.* Discussion Paper no. 375. Washington, D.C.: World Bank.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, U.K.: Cambridge University Press.

Oyono, Phil René. 2004. "One Step Forward, Two Steps Back? Paradoxes of Natural Resources Management Decentralisation in Cameroon." *Journal of Modern African Studies* 42(1): 91–111.

Platteau, Jean-Philippe. 2000. *Institutions, Social Norms, and Economic Development.* London: Harwood Academic Publishers.

_____. 2004. "Rushing to Help the Poor through Participation May Be Self-Defeating." In G. Kochendörfer-Lucius and B. Pleskovic, eds., *Service Provision for the Poor—Public and Private Sector Cooperation.* Berlin Workshop Series 2004. Washington, D.C.: World Bank.

Platteau, Jean-Philippe, and Anita Abraham. 2001. "Imperfections des communautés rurales traditionnelles et développement participatif." *Revue d'Economie du Développement* 1–2(June): 197–231.

_____. 2002. "Participatory Development in the Presence of Endogenous Community Imperfections." *Journal of Development Studies* 39(2): 104–36.

_____. 2004. "Participatory Development: Where Culture Creeps in." In Vijayendra Rao and Michael Walton, eds., *Culture and Public Action: A Cross-Disciplinary Dialogue on Development Policy,* Stanford, Calif.: Stanford University Press.

Platteau, Jean-Philippe, and Frederic Gaspart. 2003a. "Disciplining Local Leaders in Community-Based Development." Working Paper. University of Namur, Department of Economics, Centre for Research on the Economics of Development, Namur, Belgium.

_____. 2003b. "The Risk of Resource Misappropriation in Community-Driven Development." *World Development* 31(10): 1687–1703.

Rahman, Anisur. 1993. *People's Self-Development.* London and Dhaka: Zed Books and University Press.

Rao, Vijayendra, and Ana Maria Ibanez. 2001. *The Social Impact of Social Funds in Jamaica: A Mixed-Methods Analysis of Participation, Targeting, and Collective Action in Community Driven Development.* Washington, D.C.: World Bank, Development Research Group.

Ravallion, Martin. 2000. "Monitoring Targeting Performance When Decentralized Allocations to the Poor Are Unobserved." *World Bank Economic Review* 14(2): 331–45.

Reinikka, Ritva, and Jakob Svensson. 2001. "Explaining Leakage of Public Funds." Working Paper. World Bank, Development Research Group, Washington, D.C.

Ribot, Jesse. 1996. "Participation Without Representation: Chiefs, Councils, and Forestry Law in the West African Sahel." *Cultural Survival Quarterly* (Fall): 40–44.

_____. 2002. "Democratic Decentralization of Natural Resources: Institutionalizing Popular Participation." World Resources Institute, Washington, D.C.

Riddel, R. C. 1999. "Evaluating NGO Development Interventions." In D. Lewis, ed., *International Perspectives on Voluntary Action: Reshaping the Third Sector.* London: Earthscan.

Sara, Jennifer, and Travis Katz. 1997. *Making Rural Water Supply Sustainable: Report on the Impact of Project Rules.* Washington, D.C.: United Nations Development Programme and World Bank Water and Sanitation Program.

Stiles, Kendall. 2002. "International Support for NGOs in Bangladesh: Some Unintended Consequences." *World Development* 30(5): 835–46.

Tendler, Judith. 2000. "Why Are Social Funds So Popular?" In S. Yusuf, W. Wu, and S. Evenett, eds., *Local Dynamics in an Era of Globalization: 21st Century Catalysts for Development.* Oxford, U.K.: Oxford University Press.

Tornell, Aaron, and Philip Lane. 1998. "Are Windfalls a Curse? A Non-Representative Agent Model of the Current Account." *Journal of International Economics* 44(1): 83–112.

Véron, René. 2001. "The 'New' Kerala Model: Lessons for Sustainable Development." *World Development* 29(4): 601–17.

Vivian, Jessica, and Gladys Maseko. 1994. "NGOs, Participation, and Rural Development: Testing the Assumptions with Evidence from Zimbabwe." Discussion Paper no. 49. United Nations Research Institute for Social Development, Geneva.

Wade, Robert. 1985. "The Market for Public Office: Why the Indian State Is Not Better at Development." *World Development* 13(4): 467–97.

_____. 1990. "Employment, Water Control, and Water Supply Institutions: India and South Korea." In Wilbert Gooneratne and S. Hirashima, eds., *Irrigation and Water Management in Asia.* New Delhi: Sterling Publishers.

Westergaard, Kirsten, and Muhammad Alam. 1995. "Local Government in Bangladesh: Past Experiences and Yet Another Try." *World Development* 23(4): 679–90.

White, Robert, and Carl Eicher. 1999. "NGOs and the African Farmer: A Skeptical Perspective." Staff Paper no. 99–01. Michigan State University, Department of Agricultural Economics, East Lansing.

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

_____. 2002. *Social Funds: Assessing Effectiveness.* Washington, D.C.: Operations Evaluation Department (OED).

Inclusiveness, Accountability, and Effectiveness of Development Assistance in Sub-Saharan Africa

BENNO J. NDULU

In the continuing search for answers to the elusiveness of development in Africa, this paper stakes out the increasingly important territory of inquiry exploring the linkages between aid, macroeconomic policy, politics, and the micro concerns of governance and project design in the African context. One of the main reasons for elusive growth and development in Africa is the absence or weakness of citizen voice, which is necessary for engendering restraint against predatory behavior and for giving governments an incentive to be accountable. While markets create managerial discipline and induce efficacy through the exercise of private choice, governments are principally disciplined through the exercise of voice to enforce representative public choice. I argue that in the absence of effective domestic mechanisms for restraint, weak autocratic regimes in Africa sacrifice overall growth for the sake of transfers to themselves and the narrow elite they serve. When unconditional aid is given in this context, it tends to increase transfers where they are already active, and policy distortions will persist despite the aid inflows and the external conditionality aimed at enforcing restraint against predatory behavior. A major concern of the paper is that external conditionality also tends to crowd out the domestic institutional and political development that is ultimately needed to discipline policy choices in the recipient countries in a sustained way.

With these themes in mind, I first lay out a conceptual framework linking inclusiveness and accountability to the quality of policy choice and results. Economy-wide and project-level empirical evidence reviewed in the paper broadly supports the association between inclusiveness, accountability, and results. The paper concludes by emphasizing that the next big push for the African development agenda is improved

governance and poses the key challenges for reforming aid principles and practice to support this push.

Introduction

Sub-Saharan African economies have not yet achieved sustained growth and poverty reduction despite protracted and high levels of foreign assistance. Overall, the long-term growth of Sub-Saharan African countries has been slow relative to that of other developing countries: less than half the growth and about half the investment efficiency levels.[1] A general lack of persistent growth is, to a large extent, a major factor behind Sub-Saharan Africa's inability to achieve poverty reduction and relatively higher living standards in the long run. This concern has recently heightened from the perspective of poverty reduction, especially in light of the emerging view about poverty being a global public "bad" and poor countries being the weakest link in the global initiative to promote peace and security as a global public good (Kaul, Le Goulven, and Schnupf 2002).

A related concern continues to be the ineffectiveness of aid in engendering results and sustained growth in Sub-Saharan Africa. A vast amount of empirical research shows that aid in general has had no observable impact on economic growth in poor countries, but it has promoted growth in countries with good policies and a good institutional environment. Recent work by Burnside and Dollar (2000) and Collier and Dollar (2000) concludes that success has been achieved where countries have set up and sustained a policy and institutional environment conducive to growth and poverty reduction. Given such an environment, countries could gainfully absorb much larger amounts of aid than what is currently available to them.

The search for answers to the reasons behind the elusiveness of development in Sub-Saharan Africa has focused largely on weaknesses in policy and the institutional environment in relation to addressing resource shortages and inefficiency. In the context of the search for answers and solutions, the pendulum has swung from an earlier emphasis on market failures (which indicated that the state needed to play a large role in engendering development) to concerns about state failure, and hence the need to improve governance structures and make political processes more inclusive. Inclusiveness is a crucial feature of a developmental state and can be broadly characterized by representational political processes, accountability and openness, broad participation in national policy choice, and feedback on the outcomes and impact of public service delivery.

As Pritchett and Woolcock (2002) argue, while markets create managerial discipline and induce efficacy through the exercise of private choice, governments are principally disciplined through the exercise of voice to enforce representative public choice. I attempt to show that in the case of Sub-Saharan Africa's development experience, space for exercising voice in the choice of national development policies and management is necessary to sustain reforms, achieve durable growth, and realize

results from public programs on the ground. This is a theme that the *World Development Report 2004* (World Bank 2003) pursues in remarkable detail, following on the equally remarkable work on voice in an earlier *World Development Report* (World Bank 2000b). I agree with Mansuri and Rao's (2003) and Cleaver's (2000) conclusions that the development of the conceptual framework explaining the link between participation and inclusiveness on the one hand, and project efficiency on the other, is at an early stage of development, as is the empirical research to establish a systematic link. Nevertheless, this paper uses the substantial work in this area to draw out this important relationship in the Sub-Saharan African context.

The basic argument of this paper is that under authoritarian rule, a narrow group of elites operating under relatively weak institutional constraints has tended to capture development policy and management in many Sub-Saharan African countries. Characterized by weak legitimacy and tenuous bureaucracy, until the 1990s weak, autocratic regimes in Sub-Saharan Africa did not function as agencies of public interest, but rather sacrificed growth in favor of distributive politics that benefited the elite and through predation (Adam and O'Connell 1999; Bates 1983; Collier and Gunning 1999; Humphreys and Bates 2001; Ndulu 1986; Ndulu and O'Connell 1999). With the exception of Botswana and Mauritius, which espoused democratic developmental states and pursued and sustained high growth rates, the region typically experienced huge transfers to narrow elites, little reform, and minimal growth. As Adam and O'Connell (1999) and McGuire and Olson (1996) argue, in systems of autocratic rule, the predation problem intensifies with shorter ex ante planning horizons. Even though some autocratic regimes in Sub-Saharan Africa have had long tenure ex post—and Sub-Saharan Africa has had some of the longest lasting ruling regimes in the world (Humphreys and Bates 2001)—the high ex ante probability of violent regime change, which has sometimes involved leaders' death or exile, has influenced their planning horizons. As political contestation and inclusive processes for national development have been emerging since the 1990s, Sub-Saharan Africa has an alternative to its doomsday regime transition approach, one that would avoid the large economic and social costs of civil strife and political instability.

Aid given to regimes dominated by narrow elite interests is likely to be ineffective because, as Adam and O'Connell (1999) argue, a rise in unconditional aid is likely to simply increase transfers to elites. Moreover, unless the government represents all strata of society, there is a level of aid above which transfers will be initiated in preference to growth-enhancing tax cuts or increases in public investment spending. Aid does not directly reduce growth in these cases, but policy distortions will persist despite inflows. In the absence of a transfer of political power, slow growth is a locally stable political equilibrium (Collier and Gunning 1999). The primary effects of aid in this situation may operate through domestic competition for politically-motivated transfers and the power to dispense them.

Furthermore, Braetigum (2000) argues that aid intensity and the approaches to aid delivery have tended to blunt the development of domestic accountability systems in recipient Sub-Saharan African countries by generating perverse incentives for accountable behavior by governments, and by providing additional room for financing

patron-client networks and for elite capture. Aid has insulated governments from pressures to adopt inclusiveness in relation to development management. As a result, accountability systems have been fragmented and revolve around dialogue between the government and the multitude of donors typically working in a specific Sub-Saharan African country. Much national energy and political capital is applied to conducting dialogue on aid and managing a diverse project pipeline, resulting in large transaction costs (Killick 1995; see Wuyts 1996 for the case of Mozambique). More worrisome is that often these engagements compete with paying attention to the domestic development debate and to consensus building.

To contain predation, the international development community has emphasized the application of external conditionality to discipline policy choices or reduce policy distortions. Note that traditional aid conditionality emerged during the 1970s to counter deteriorating policy choices associated with increasingly nonrepresentative political structures and weak autocracies. Case studies from Sub-Saharan Africa strongly suggest that even though conditionality may be helpful in establishing a country's credibility during a period of rapid reform, its usefulness is short-lived (Devarajan, Dollar, and Holmgren 2001). This is because donors' continued use of conditionality goes against the grain of efforts to reform ownership, particularly in environments where such reform is taking root (Tsikata 2002). The domestic constituency for reform could be strengthened if citizens had a better appreciation of their own interests through improved access to knowledge and greater capacity to absorb and process information. Furthermore, while some progress has taken place in turning the focus toward strengthening the institutional framework from the supply side, fewer efforts have been made to raise the demand for good governance, a necessary complement to strengthening accountability. The absence of feedback from the ultimate beneficiaries of aid in recipient countries has compromised the efficacy of aid.

More recently, donors have added so-called process conditionality to the tool kit, for example, making aid conditional on countries adopting inclusive processes in the preparation of poverty reduction strategies or setting up democratic institutions. The purpose of process conditionality is to increase the influence of the poor on policy choices in recipient countries by increasing their participation (Hefeker and Michaelowa 2003) with the hope that this will boost the likelihood of achieving pro-poor growth. However, external partners have also typically sought to influence whom the poor choose as their representatives in the domestic dialogue rather than allowing representational processes to emerge. As Hefeker and Michaelowa (2003) argue, the success of this approach will depend on the adequacy of information available to external partners in relation to assessing the political situation and selecting the right people to represent the poor. Apart from the tension between process conditionality and policy conditionality that Adam and O'Connell (1999) discuss, worries are emerging that donor-driven processes could crowd out the critical domestic institutional and political development that is ultimately needed to discipline policy choices in the recipient countries in a sustained way.

Until recently, much of the analysis on the effectiveness of conditionality in engendering reforms has been carried out in the context of external agencies dealing with

unitary governments. The problem was perceived to pertain to time inconsistency or conflicts of interest among bargaining parties in aid relationships complicated by asymmetric information. The policies of recipient governments were considered time inconsistent, with governments accepting ex ante the need for policy changes as a condition for receiving aid, but having a strong incentive for breaking the deal once they had received the aid (Diwan and Rodrik 1992; Sachs 1989). More recently, analysis has moved toward discussions of the tension between conditionality and ownership by incorporating multiple (conflicting) interests in recipient countries in relation to reform programs (Drazen 2002) and in both recipient and donor agencies (Azam and Laffont 2003). Drazen (2002) argues that because policymaking is a process of collective choice in the face of conflicting interests, ownership by some important policymakers in a country is not necessarily ownership by the policymaking apparatus, which includes the entire domestic constituency (see also Khan and Sharma 2001).

Against this background, the paper will attempt to stake out the increasingly important exploration of the linkages between aid, macroeconomic policy, politics, and the micro concerns of governance and project design in the Sub-Saharan African context. It seeks to address the following three questions:

- What difference has inclusiveness and participation in national policy formulation and development management made in achieving and sustaining results at the economy-wide and subnational levels in Sub-Saharan Africa?

- What are the main constraints to fostering inclusiveness and accountability in development processes at these levels and what role has aid played in these constraints?

- In a path-dependent framework and given strategic incrementalism, what transitional actions are needed for building a strong domestic constituency for achieving and sustaining results?

The next section provides a brief background of the roots and logic of the Sub-Saharan African development autocracy as the main driver of policy distortions. Against this background, the following section lays out the conceptual framework linking inclusiveness and accountability in the management of public service delivery on the one hand, and the quality of policy choice and results on the other. The paper then turns to an overview of economy-wide and project-level empirical evidence on the association between inclusiveness, accountability, and results. The final section focuses on the way forward by suggesting that the next big push for the Sub-Saharan African development agenda is improving governance and reforming aid principles and practice to be consistent with governance reform.

The Roots and Evolution of Development Autocracy in Sub-Saharan Africa

This section offers an overview of the roots and evolution of Sub-Saharan African development autocracy as the main driver of policy distortions, looking first at

inclusiveness in national policy choice and then at inclusiveness through the devolution of responsibility to subnational governments.

National Level

At Sub-Saharan African countries' independence, observers considered their predominantly peasant economies not only to be technologically backward, but also as lacking the requisite dynamism for autonomous development. They argued that the state needed to play a central role as the main agent for modernizing the economy (Ndulu 1986). A government was therefore to use its fiscal powers, that is, the external resources channeled through it, along with indirect controls on private sector resource allocation, to this end. The government promised development in exchange for the state's right to maintain a centralized, authoritarian system of governance. Thus in most Sub-Saharan African countries, a so-called social contract was struck early in the postindependence period that traded the right to open governance structures for patronage and the promise of rapid growth in what has been referred to as developmental autocracy (Gordon 1990). The political legitimation of this paradigm drew on the liberal assumption that the state is a neutral, and even a benevolent, arbiter among different interest groups that would further national interests in relation to economic growth, efficiency, and social welfare (Sandbrook and Barker 1985). Donor support of all types during the first two decades of independence was channeled through the state with the same understanding.

In these economies the private sector was, in essence, part of the covenant with the state, which provided protection and granted access to subsidized resources and to significant rents derived from protection. In many cases, a strong symmetry of interests existed between the public sector and the politically well-connected private sector as far as price distortions, protective measures, and access to subsidized resources were concerned. Economic nationalism was promoted through state enterprises and by encouraging the indigenous private sector by means of preferential arrangements of the type described (Ndulu 2002).

The logic of the weak autocracy framework of governance encompasses two main underlying trends that make the demise of the regime imminent. The first is the time inconsistency of the system of rent extraction in the absence of growth caused by the distortionary effects of taxation and the fact that soon these types of distortions move the system to the wrong side of the Laffer curve (Adam and O'Connell 1999). The second is the assumption that the multitude of unorganized interest groups, including the peasantry, is politically passive and acquiescent to the benevolence of the state, even if the state's actions are geared toward catering to the narrow interests of favored and vocal groups.

The regimes in power totally underestimated the potential of grassroots responses to frustrate predation. Because the revenue base is the most important political resource for servicing patron-client networks, its erosion through faltering revenue collection undermines the capacity of political patrons to service their networks of clients. Economic stagnation caused by overtaxing productive sectors and the increased proportion

of the economy that went underground because of the preponderance of controls had a negative impact on the tax base. The steady increase in controls to create rents for the favored groups was the main reason behind the mushrooming of parallel markets and illegal cross-border trade to evade the controls. This development not only undermined the fiscal base of the state but also significantly reduced the rental incomes that had hitherto been available to service patron-client networks.

Another undercurrent was the latent tension between grassroots interests and the purported national interests advanced through state policies. Sub-Saharan African social structures are characterized by the dominance of relatively autonomous networks bound by kinship, tribe, religion, race, or community ties. These networks span rural-urban boundaries in what Hyden (1983, 1986) calls the economy of affection. The state is structurally superfluous and is only acknowledged to the extent that these networks consider its actions beneficial. Where they deem government policy to be unbeneficial or to threaten their interests, these networks, through their own communication systems, have frustrated state policies through the use of a variety of exit options. These include participation in parallel markets, illegal cross-border trade, and even informal banking systems.

Thus grassroots networks have proved to be a potent force against predation when their own interests are ignored. The networks, by contrast, are a form of social capital, which if harnessed present an opportunity for successful project implementation.

The roots of recent changes toward a more encompassing governance system in Sub-Saharan African countries lie in the foregoing developments. Conditionality associated with adjustment lending has, to a large extent, amplified the latent pressures for change and catalyzed the process. By the time the externally enforced adjustment had taken effect, the state's traditional role had become unsustainable and the social contract struck at independence had been broken. Governments had overextended themselves relative to their financial and managerial resources and their populations had become cynical about the effectiveness of development institutions. Rampant evasion of controls through vigorous parallel markets exerted pressure for dismantling them, and in most cases parallel markets were eventually formalized and made more efficient under reform programs. The downsizing of government operations was essentially a forgone conclusion given the emerging unsustainable resource gaps and the rising debt service burden. Meanwhile, global views about minimalist government gathered momentum, and aid relationships and international governance institutions reinforced pressures for change. Finally, as democracy is setting in, albeit haltingly in some cases, the removal of unwanted regimes from power through elections is becoming a potent alternative to using either force or latent exit options (Ndulu 2002).

Subnational level

At the subnational level, the former colonial powers had used local, indigenous structures to maintain a two-track legal and administrative system: "city" law for

urban areas and tribal law for rural local authorities based on traditional power structures. Soon after independence, many of these structures were dismantled and replaced with a deconcentrated system of central authority. In many respects, this undermined local governments' capacity in relation to the choice and delivery of public services at the local level. Indeed, this change also led to the disappearance of voluntary cooperatives built around traditional structures, which were typically replaced with imposed cooperative societies built around administration boundaries. An important exception is Botswana, which has not only maintained traditional authority but has integrated it into the modern state system.

Strong state regulation of local development pursued in the postindependence period was partly influenced by the work of Demsetz (1967); Hardin (1968); North and Thomas (1977); and Olson (1973), who raised concerns about the tragedy of commons in the absence of coercion and regulation for achieving common interest. This work generated a great deal of pessimism about local authorities' capacity to provide public goods.

By the mid-1980s, it became clear that centralized service delivery programs were unsuccessful and unsustainable, reawakening interest in local management and decisionmaking building on local knowledge and accountability systems. Inclusiveness in this context involves a willingness to devolve power to subnational entities as a way to enhance participation in development management and strengthen accountability systems. Notions of participatory development were built on collective action by the poor to improve their own well-being (Cernea 1985; Chambers 1983; Hirschman 1984). The World Bank's 2000/2001 *World Development Report* (2000b) contains perhaps the strongest explanation of the rise of decentralization and local authorities and the shift of focus to empowerment as one of the key priorities of development policy.

National Policy Choices under Weak Autocracy: A Conceptual Framework

The first part of this section draws on three recent papers to formalize the process of national policy choices under weak autocracy in the Sub-Saharan African context. Following from this conceptual framework, the second part of the section analyzes the role and effectiveness of the government as an aid intermediary under inclusive and noninclusive political regimes.

Inclusion and National Policy Choices

Adam and O'Connell (1999); Humphreys and Bates (2001); and Ndulu and O'Connell (1999) develop models that explain Sub-Saharan Africa's slow growth by emphasizing policy choices that favor transfers to narrow political elite at the expense of productive investment. For simplicity and clarity of exposition, I focus on the thread of argument in the Ndulu and O'Connell (1999) variant of the models.

Applying a variant of Olson's (1982) encompassing interests paradigm to fiscal policy choices under authoritarian rule, the main explanation for poor growth performance in authoritarian regimes Ndulu and O'Connell (1999) offer is that ruling elites in these regimes sacrificed the general interest in order to extract rents and retain power. Outright predation can emerge when rulers believe that wealth accumulation outside the control of the elite will increase political contestability. A narrow elite overtaxes and underprovides public goods, sacrificing overall growth for the sake of transfers to itself, while a broad and more representative elite internalizes the general interest in growth.

The scope of the government's inclusiveness is the exogenous fraction, f, of national income directly controlled by the elite. Current consumption by the elite is given by $C = fY - I$, where Y is the (exogenous) current national income and I is infrastructure spending. In the absence of aid, domestic revenues constrain the choices open to the government. The opportunity cost of a unit of infrastructure spending is therefore a unit of current consumption. Future consumption of the elite, $C_F = fY_F(I)$, increases with infrastructure spending, because today's road building increases the economy's future productive capacity.[2] Leaders of autocratic regimes with high discount rates because of the perception that their continued rule faces high risk will maximize transfers to the narrow elite at the expense of investment. Risks are perceived to be particularly high in Sub-Saharan Africa, where regime changes have tended to be violent. A more encompassing regime would choose to expand future productive capacity.

The basic model as applied in Ndulu and O'Connell (1999) poses a policy choice in the allocation of budgetary resources between transfers to a narrow elite in a patronage system and developmental spending (represented here by expenditure on infrastructure). Tax revenues are thus split between public infrastructure spending and per capita transfers of size T to a favored group of size f (with the population normalized at 1).

$$fT + \text{infrastructure spending} = tB(t), \quad B'(t) < 0 \qquad (1)$$

Domestic taxation is distortionary, so that a rise in t shrinks the tax base $B(t)$ and creates a deadweight loss. For fixed t, growth effects emerge over time by undermining investment in the taxed activity.

To motivate the choice of t and T, we think of the government as maximizing the welfare of the favored group. As long as infrastructure spending is positive, some amount of distortionary taxation cannot be avoided. In this situation, a tiny transfer creates a finite additional distortion at the margin and is therefore socially inefficient. Thus a government with fully encompassing preferences—indeed, even one with "sufficiently" but not fully encompassing preferences (see Adam and O'Connell 1999; Boone 1996; McGuire and Olson 1996)—will forego transfers. There is some level of f, however, below which the concentration of economy-wide revenues in the hands of the favored group will justify transfers. Narrower power structures will generate larger transfers and slower growth.

The model is consistent with the broad features of Sub-Saharan African economic policy, in particular, the heavy discrimination against international trade and the widespread use of quantitative controls in preference to price interventions during the 1970s and 1980s. Given the exclusion of peasant interests from the favored group and the low administrative costs of taxing external trade, transfers from export agriculture are often large. The tendency to overtax export sectors and to rely heavily on revenues to service patron-client networks created vulnerable balance of payments situations and tenuous fiscal positions. Transfers were effected largely by establishing and implementing control regimes that typically involved (a) restricting import licenses to the favored group, (b) giving a scarcity premium of underpriced capital to those who are politically connected, and (c) sustaining a large public sector for serving political patronage through jobs.

In a similar vein, Humphreys and Bates (2001) consider a class of models in which politicians choose from a set of policies that have implications for aggregate income and its distribution. The main hypothesis these authors test revolves around a tradeoff ruling regimes face between choosing pro-growth policies that maximize aggregate economic activity and freedom to engage in distributive politics to deliver benefits to the "right" constituencies. The model emphasizes that elections influence policy choice by putting the political fortunes of politicians who seek to retain office at risk. Policy choice tends to be pro-growth if the discounted returns from retaining office in a competitive political framework exceed the discounted rewards of predation for the private benefit of the regime's leaders.

Equation (2) extends the basic model, equation (1), to include the line items for budgetary financing. Thus the extended budget constraint includes net inflows of foreign aid and borrowing from the banking system or inflation tax revenue. Given the fairly undeveloped market for government securities in the region, the equation excludes nonbank borrowing from significant sources of budget financing.

$$fT + \text{infrastructure spending} = tB(t) + FA + pm(p) \quad pm'(p) > 0$$
$$\text{for } p < p^* \text{ and } < 0 \text{ for } p > p^*, \tag{2}$$

where p is the inflation rate, m is the real money demand, and p^* is the optimal inflation rate beyond which revenue from inflation tax declines in absolute terms (Ndulu 2002).

In this extended version of the basic model, aid and revenue from the inflation tax provide additional resources to the government for financing infrastructure and transfers to the favored groups. Until the early 1990s, foreign assistance was readily available and was channeled through the state apparatus in exchange for what leadership elites could supply. Sometimes this did not include rapid development, which occurred in some cases, but appears to have been incidental to aid flows (Burnside and Dollar 2000; Collier and Dollar 2000; Killick 1997; Lensink and White 1999; World Bank 1999).

In equation (2), untied or fungible aid inflows can enhance growth by allowing a reduction in distortionary taxation or public infrastructure financing. For a sufficiently encompassing government, this is what the model predicts (Boone 1996). It is easy to see that if conditionality were effective, a similar effect would emerge if, for example, aid financed infrastructure spending. However, empirical evidence suggests

that when aid was predicated on fulfilling conditions linked to improvements in the policy and institutional environment, it did not make any significant difference to the adoption of desired policies, mainly because of a failure to enforce conditionality. Empirical research has also found little association, let alone causation, between the application of ex ante donor conditionality and improvements in the policy environment (Burnside and Dollar 2000; Collier and Gunning 1999; Dollar and Svensson 2000; Kanbur 1999; Killick 1997; Lensink and White 1999).

From the perspective of governance, seignorage revenue plays a similar role to unconditional or fungible aid. As long as it stays on the correct side of the inflation tax Laffer curve and the government retains credibility about its intentions to keep inflation low (Azam 1997), an all-encompassing government can use seignorage revenue to reduce distortionary taxation or raise spending on infrastructure. However, this form of financing is limited by the fact that beyond the optimal inflation tax rate the government moves to the wrong side of the Laffer curve and revenue from this source falls in absolute terms. Furthermore, tension between the government and favored groups may ensue from the effects of high inflation as the latter experience real erosion of their wages or profits, and distortions in the incentive structure lead to a reduction of the tax base, particularly as the local currency appreciates and exports decline.

During the first half of the 1980s, many Sub-Saharan African countries outside the franc zone resorted to this kind of financing, and most of them overshot the optimal inflation tax rate (see Adam, Ndulu, and Sowa 1996 for examples of such overshooting in Kenya, Ghana, and Tanzania). They took this course of action partly to cushion themselves against the consequences of fiscal crises, and partly because of a temporary decline in external aid inflows associated with the ongoing global recession and debt crisis. In a number of cases, however, excessive use of the inflation tax persisted well into the 1990s. The absence of a strong and independent central bank was typically the reason for the lack of restraint in relation to staying on the correct side of the curve.

As concerns the foregoing model, two important caveats need to be highlighted:

• One weakness of the model and much of the related literature is that they gloss over the question of whether participatory political institutions necessarily produce policy choices that best serve the public interest.[3] Quite often, inclusive or participatory political systems have an encompassing point of view, but it does not follow that, for example, a majoritarian political system, which may be viewed as highly participatory, succeeds in aggregating many individual interests into a truly encompassing welfare function. One could envisage such regimes resulting in populist results; for instance, a high tax rate on capital that seeks to finance progressive redistribution but fails by undermining growth. In nascent democracies, policy performance could be fitful because of the emergence of populist pressures during and after the transition to democracy or during election cycles. Developing a system with a high encompassment of interests is a constitutional problem and confronts us with the question of whether participation and ownership as domestic political processes can be prone to populism and other "mistakes." What

are the sufficient conditions for emerging, encompassing regimes to generate sustained pro-growth policies?

- Almost all the most spectacular episodes of growth and poverty reduction in East Asia occurred under strong autocratic regimes and not democracies. It appears that demands for democracy followed success in both stellar growth performance and deep poverty reduction. The difference lay in these governments' pursuing an encompassing interest (such as shared growth) with low discount rates and facing few external threats. By contrast, authoritarian governments in Sub-Saharan Africa had high discount rates and tended to privatize the rewards of office. Most of these regimes faced a threat of violent termination of their tenures as corroborated by the frequency of coups, and hence faced high risks of finite office terms. If finite terms prevent policymakers from investing in policies that may benefit them while they are in office, then finite terms may have a negative influence on policy choice (Humphreys and Bates 2001).

Sub-Saharan Africa does have examples of regimes that were highly "personalized," faced a low risk of a sudden and violent termination, and produced better policies; for example, Jomo Kenyatta in Kenya and Houphouet Boigny in Côte d'Ivoire oversaw fairly long periods of rapid growth. However, the two countries in Sub-Saharan Africa that have sustained high growth rates even by world standards are Botswana and Mauritius, where national-level policy has been contested through regular elections and meaningful legislative roles, both of which have enforced a discipline of accountability for results that has broadly benefited the electorate. Furthermore, by adopting inclusive politics Mauritius has demonstrated that ethnic diversity need not lead to divisive politics and conflict. Similarly, political tolerance and inclusive politics have characterized the peaceful transition in South Africa since 1994 (Chegge 1999). Sub-Saharan Africa clearly seems to have embarked on a path toward democracy, and the option of benevolent autocracy appears to be petering out.

Role of Aid under Weak and More Inclusive Regimes

Within the conceptual framework developed earlier, this section analyzes the effectiveness of government as an aid intermediary under inclusive and noninclusive political regimes. It employs the narrower (and sometimes considered distasteful when applied to the aid relationship) framework of the principal-agent theory, distinguishing between a weak autocratic regime versus an encompassing political regime, in relation to the different behavior of the recipient government as an aid intermediary. The discussion does not consider the incentive problems in donor agencies (see Azam and Laffont 2003; Easterly 2002; and Hefeker and Michaelowa 2003 for discussions of donor behavior in the context of the political economy of aid).

Kanbur (1999) succinctly summarizes the aid relationship as a principal-agent problem and models the relationship in terms of a Stackleberg leader-follower

interaction. The donor is the leader and decides on the level and composition of aid. The recipient is the follower who, taking as given the level of aid, decides on actions, (for example, public expenditure patterns or trade policy) that affect outcomes for the recipient, such as economic growth and access by the poor to education. However, the donor also values these outcomes and chooses a level of aid that will influence the choice of actions by the recipient, and hence the outcomes for the recipient. The level of aid is thus chosen to maximize the donor's preferences, subject to the reaction function of the recipient, which in turn derives from the recipient's preferences and shows the actions the latter would choose for each level of aid.

The application of the principal-agent theory in a weak autocratic regime highlights the problem of differences in relation to development preferences and priorities or imperfect information about the real intentions of the aid intermediary (the agent's hidden information or hidden action) and the objectives of the donor (the principal). This problem appears in terms of time inconsistency in an aid relationship, particularly where mechanisms for enforcing the principal's (donor's) preferences are ineffective. Because in the typical case aid is additional to domestic resources, the fungibility of aid adds to the likelihood that the government's composition of expenditure will diverge from that preferred by the donor.

The simplest versions of the theory assume the donor and the intermediary to be unitary entities, represented only by a set of preferences. They rarely incorporate processes for determining preferences and accountability arrangements, even though these influence the enforceability of contracts entered into between the principal and the agent.[4] Most models assume that the donor is more concerned about the poor (the ultimate beneficiaries) than the intermediary is. In general, all that is needed to make contracts between the principal and the agent tenuous is for their preferences to differ. The donor typically relies on aid conditionality to enforce compliance by pricing aid in terms of the actions taken. The recipient may prefer unconditional aid, but has no choice (Kanbur 1999).

The application of the principal-agent theory in a more inclusive political regime treats the government as an aid intermediary (agent) with preferences that have to reflect those derived from an inclusive domestic political process for determining development priorities. Thus a slight variant of the traditional application of the principal-agent problem to aid relationships is appropriate. In this variant, we consider the autocratic government to be an aid intermediary (agent) with preferences that do not necessarily reflect those deriving from an inclusive domestic political process for determining development priorities. The government intermediates between the principal (the financier) and the electorate, which is the ultimate beneficiary of aid, and indeed, the main contributor to the total public resource envelope. In contrast to a government in an authoritarian regime, in a democratic system the recipient government is, in effect, accountable to two principals for outcomes resulting from the application of aid resources and the country's own resources. The electorate can vote the government out of office in case of failure, while the donor retains the power to withdraw aid if the intermediary does not deliver on promises. Where the preferences of the two principals coincide, the leverage for enforcing compliance

is greatly enhanced. Where they do not coincide, the government becomes a roving agent with its choices weighted toward the principal closer to its preferences.

This variant is similar to the theoretical literature that models the recipient not as a unitary government, but as a combination of interest groups interacting among themselves, and treats the government as representing a democratically processed set of preferences or as an arbiter among competing interests (see, for example, Adam and O'Connell 1999; Coate and Morris 1996; Drazen 2002). Through aid flows, the donor then tries to influence this process of domestic political economy, strengthening the hand of one group against another, thereby ensuring that some actions are more likely to be taken, and hence that some outcomes are more likely. Problems include both the limited availability of channels for exerting such influence and the sustainability of the strategy as the fortunes of interest groups change (Ndulu 2002).

Conditionality is a commitment technology that is expected to overcome the time inconsistency inherent in these problems, but only if it is strictly adhered to. If the donor has no way to ensure the enforcement of conditionality—that is, it cannot make a credible threat in the event of violation—the outcomes will essentially depend on the agent's behavior. Donors are typically unable to enforce conditionality on governments except, as noted earlier, where the conditions coincide with beneficiaries' concerns. To ensure the effectiveness of conditionality, donors need to minimize any divergence in preferences and institute a credible system of accountability and threats.

In an open governance system, the tension between two channels of accountability, one to donors and the other to electorates, becomes more pronounced and needs resolution. The tension manifests itself essentially as a conflict between conditionality and local ownership in an aid relationship. Donor agencies have to be accountable to their own political systems and, more critically, within those to a constituency for aid. Tied aid, technical cooperation, and resources devoted to "consultants," for example, are subject to internal pressures in donor countries. Tension also arises between pressure for short-term, well-identified results linked to specific support measured in physical, objective terms and project designs conducive to helping recipient countries create effective governance institutions in the longer term. The pressure for results—schools built, roads constructed, babies immunized—in turn creates pressures for cocooning donor projects from the general institutional weaknesses by creating alternative channels of accountability to donors. Furthermore, when donors and recipients have different preferences about budgetary allocations, donors have also used conditionality to help implement their preferences. An example here is priority spending on social sectors that sometimes comes at the expense of other sectors that could more directly address income poverty.

The international development community recently adopted the Millennium Development Goals as the main framework for monitoring the implementation of poverty reduction programs. In this context, the grand bargain seems to be more money for aid in return for more accountability for results. However, the results are phrased in terms of output targets, for instance, enrollment levels. The pressure for short-term numerical targets—which are perhaps necessary to mobilize support for aid—may conflict with long-run effectiveness in terms of seeking to create local

processes for absorbing and using resources effectively. The focus should therefore be on building a more credible monitoring and feedback system, which will also encourage the development of an accountability system that will integrate domestic and external concerns.

This conceptual framework can also be extended to a devolved system of public service delivery and development management. In a multilevel government structure, each level is typically simultaneously a principal and an agent (Hamdok 2002). The relationship in this case is between the central government agencies (as principals) and subnational governments (as agents) in receipt of grants. The local government is modeled not as a unitary local government but as a combination of interest groups interacting among themselves, and the local government represents a democratically processed set of preferences. Akin to the case of national policy choice, through aid and grant flows the central government and donors may try to influence this process of localized political economy, strengthening the hand of one group against another to ensure that some actions are more likely to be taken, and hence making some outcomes more likely. The main problem is again the limited availability of channels for exerting such influence in a situation where the prominence of interest groups changes. The problem of contract enforcement applies here as it does in the national case, calling for a process to ensure the coincidence of local electorates' preferences with those of the central government and donors through an inclusive approach to development management and information flow to support contract enforcement.

Community-driven development as an organizational framework for conceiving and implementing projects was designed partly to circumvent the foregoing problems and to enhance the positive impact of projects. This approach relies on communities using their social capital (trust, norms, networks) to ensure that projects are better designed, benefits are better targeted, project inputs are delivered in a more cost-effective and timely manner, and project benefits are distributed more equitably and with smaller leakages resulting from corruption and other rent-seeking activities (Mansuri and Rao 2003).

One problem is how donor activity can strengthen rather than weaken a country's social fabric. After all, one of the major problems with the social capital agenda is whether purposive actions can create social capital.[5] Project design can undermine previously successful strategies for collective action, but can they be built as a result of project interventions? Possibly existing effective modes of collective action (for example, religious groups, informal sharing) are effective and can work informally with little conflict precisely because the stakes are so low, but if an outside group decides to funnel a largish amount of resources through that group the very nature of the social interactions might change because so much is at stake (see Cleaver 2000). This is one of the hard problems project design faces, namely, the use of existing groups, perhaps with strengthening, versus the establishment of new groups, such as user associations, that could undermine existing groups.

Another issue is the typically small share of local revenue in financing development programs and the large flows of grants to local authorities, which dampen the demand for accountability. Taxation has historically been a powerful stimulus to the

development of popular restraints over the public sector, that is, people who are being heavily taxed will develop institutions capable of reducing the excess burden and will require good performance in relation to the delivery of public services. This is particularly true if the intended beneficiaries of public service programs are well informed about expenditures and results. By analogy, user fees would accomplish the same thing. There is an understandable strong stance against the nondiscriminatory application of user fees on poverty grounds, but projects have used other forms of contributions by local communities for reasons of sustainability and enhancement of accountability.

Does Inclusiveness Affect Results and Effectiveness: Empirical Evidence

A growing body of multidisciplinary empirical research provides significant evidence of a strong positive association between participation and inclusiveness on the one hand, and between successful policy reforms and project results on the other. China's phenomenal success in engendering rapid pro-poor growth, reducing the number of poor by nearly 90 percent within a decade, is perhaps the most striking case where inclusiveness and participation have had the most significant results (Stern 2002). The success has been credited partly to the implementation of the household responsibility system and the township and village enterprise program. Narayan (2002) pulls together a wide range of emerging evidence of this relationship at the economy-wide and project levels. The following draws from this source and others to illustrate selectively the types of empirical evidence available, particularly for Sub-Saharan Africa.

Economy-Wide Evidence

Evidence of differential growth performance across different political regimes in Sub-Saharan Africa exercising varying degrees of participation in decisionmaking at the macroeconomic level and of restraints against predatory behavior corroborates the view that performance is better where inclusiveness in national policy choice is applied. Following the categorization of Bratton and van de Walle (1997), Ndulu and O'Connell (1999) disaggregate growth performance by regime, distinguishing between multiparty systems, single-party systems (whether competitive or plebiscitary), and military oligarchies (figure 1). The classification applies to 1988, and with relatively few exceptions it was established by the mid-1970s. At this level of aggregation, countries with multiparty systems had higher incomes on average in 1960 and thereafter. Equally striking is their contrarian growth experience: on average, the multiparty systems diverged from the rest, avoiding the extended contraction that was such a prominent feature of Sub-Saharan Africa's overall experience. Within the authoritarian group, the military oligarchies were poorer on average at the outset of the period under review and remained so throughout. Neither of the authoritarian subgroups avoided the dramatic post-1970s growth slowdown.

FIGURE 1.
Real Gross Domestic Product Per Capita by Political Regime, 1960–97
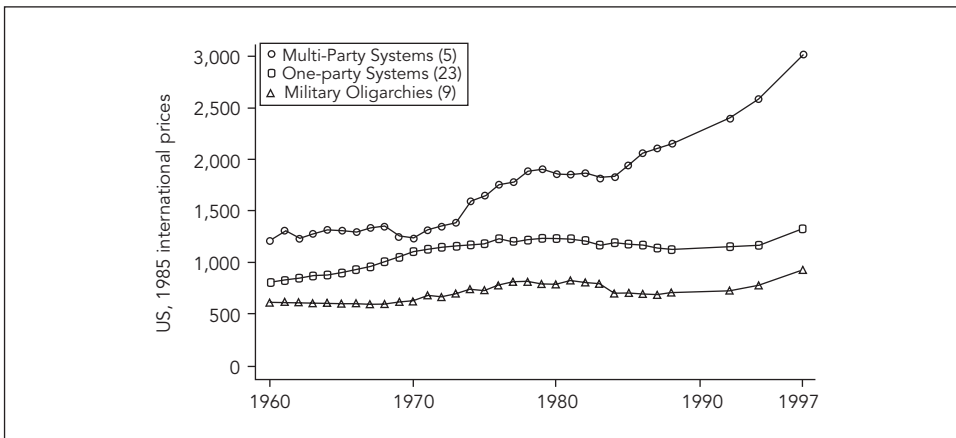

Source: Ndulu and O'Connell 1999, figure 3.

Using a measure of the quality of economic policies in the 1970s, 1980s, and 1990s in a panel of Sub-Saharan African and other countries, Humphreys and Bates (2001) investigate the extent to which this measure is influenced by the characteristics of a polity and the incentives for the government to provide public goods rather than engage in distributive politics. These factors include the characteristics and size of the constituency required to keep a government in power, the ease with which a government can redirect public goods into private benefits, the presence of term limits, the existence of checks and balances, and the competitiveness of electoral institutions. They find that policymakers who need to satisfy larger constituencies, who face checks and balances in relation to the decisionmaking process, who are subject to electoral review, and who function in stable institutional environments are more likely to produce good policies. What is striking is that the authors also find that the magnitudes of effects are typically larger for their Sub-Saharan African sample than for a global sample.

In a study of 81 World Bank adjustment programs in 38 countries between 1980 and 1988, Johnson and Wasty (1993) find a positive correlation between ownership and how satisfactorily the programs met their objectives. They also find that ownership was strongly predictive of program success in 73 percent of the cases. As Tsikata (2002) emphasizes, commitment by a country's political leadership is the most important determinant of success.

Kaufmann (2002) finds a significant relationship between the extent of transparent information provided by the government and accountability on the one hand, and the level of per capita income on the other (figure 2). Income per capita rises sharply for countries with above average transparency and accountability and with a strong institutional framework for upholding social contracts reached through the influence of voice (also see Addison 2003). Civil liberties are a proxy for empowerment, particularly inclusiveness and participation. Based on a sample of 150 countries, Kaufmann (2000) finds that a high level of civil liberties is strongly related to low corruption

FIGURE 2.
Transparency and Gross Domestic Product Growth, 150 Countries
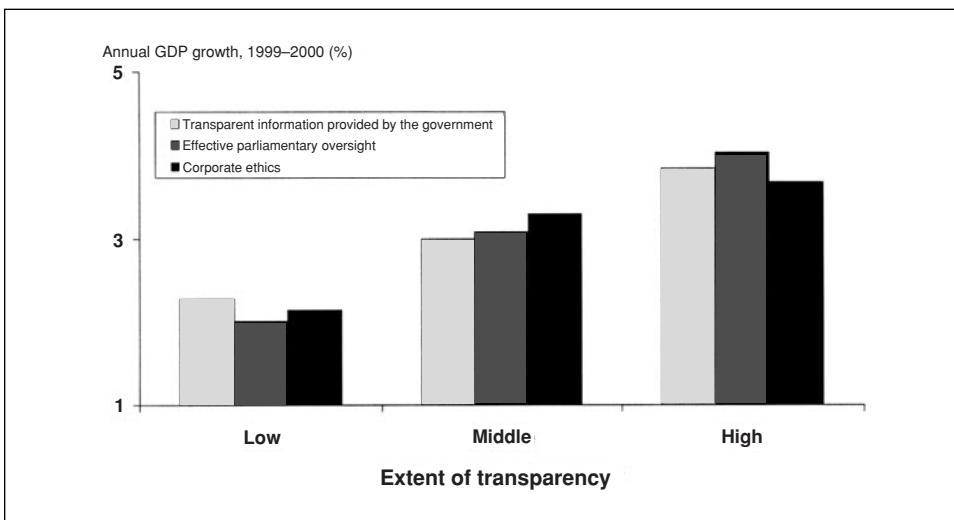
Source: Kaufmann 2002, figure 5.

levels. The link is associated with the strength of public watchdog systems armed with information and the freedom to provide feedback.

Economy-wide evidence exists at a more disaggregated level and relates empowerment, elements of good governance, and public expenditure management on the one hand, and growth and poverty reduction on the other. Narayan (2002) collates evidence of these relationships and cites the main channels of the influence of the above factors on growth to be an improved investment climate, reduced state capture, improved efficacy of public service delivery, enhanced judicial efficiency, and unfettered private initiative at the national and local levels.

Reinikka and Svensson (2001) show that disclosing budgetary allocations to intended beneficiaries of a school project in Uganda substantially improved the proportion of those resources reaching the primary education delivery units. By providing information to strengthen public watchdog systems, leakage was reduced by nearly 75 percent. In Tanzania, switching the flow of funds system directly to schools rather than through the local government bureaucracy and making budgetary allocation information public is having similar impacts. Allowing the public to review public expenditure in Tanzania and Uganda has strengthened public scrutiny of the integrity of spending and value for money.

Kaufmann, Mehrez, and Gurgur (2002) conduct an in-depth diagnostic survey of public officials in Bolivia to analyze the importance of voice, external accountability, and feedback mechanisms given to citizens in relation to the oversight of public institutions. Figure 3 shows that in Bolivia's case, an improvement in external accountability or voice improves access by the poor to public services. Furthermore, transparency and the depoliticization of agencies are significant determinants of good governance (Kaufmann 2002).

FIGURE 3.
Relationship between Voice and External Accountability and Access
by the Poor to Public Services, Bolivia
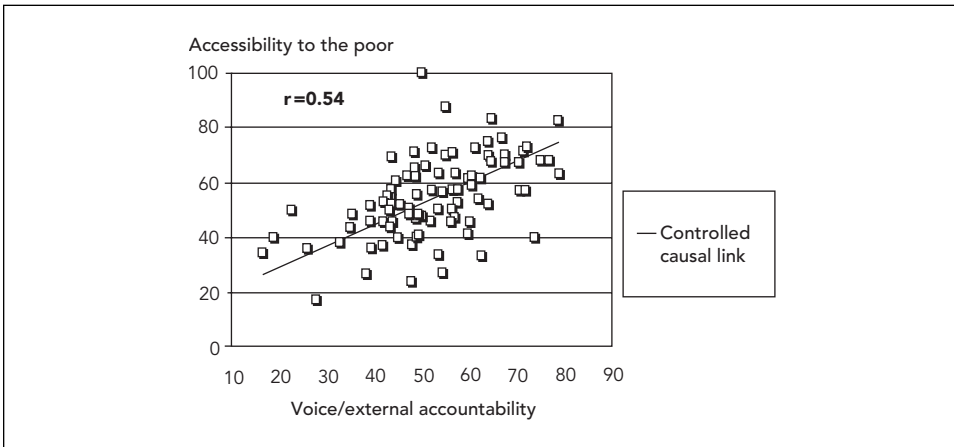
Note: The sample of institutions includes 44 national, departmental municipal agencies that
provide services to the poor. Each point depicts an institution. The units of both axes are per-
ception indexes. The higher the index the better the service and the higher the extent of
voice/accountability.

Source: Kaufmann (2002, box 2, figure A).

Ritzen, Easterly, and Woolcock (2000) demonstrate the importance of social cohe-
sion and inclusiveness for generating the trust needed to implement reforms, while
Knack and Keefer (1997) establish the relationship between growth and measures of
trust and civic cooperation.

Knack and Rahaman (2003), using the Development Gateway database, show
that donor fragmentation (based on the number of projects or on aid disbursements)
is associated with lower bureaucratic quality. This is a result of the proliferation of
enclave project management units, with the best government employees being
attracted to much better paying jobs in these units. In turn, Kaufmann (2000) shows
a strong, significant, positive relationship between bureaucratic quality (capability)
and economic growth in a study of 94 countries for 1964–94.[6]

Micro Evidence

At the micro level, a growing volume of evidence shows that when projects are con-
ceived, designed, and executed with the participation of local populations, they tend
to be more successful than those without such participation (Isham, Narayan, and
Pritchett 1995; Narayan 2002). This conclusion receives further support from stud-
ies by Isham and Kaufmann (1999) and Isham, Kaufmann, and Pritchett (1997),
which confirm the critical importance of a strong citizen voice in ensuring better proj-
ect performance through participation and better governance. Specifically, observers
have found that participation is effective in better targeting of transfers to the poor,
in improved selection of subprojects, and in enhanced locating of small community

projects through better informed decisionmaking. They have also found that participation by beneficiaries in running projects and programs also improves public service delivery. The rates of return of projects in countries with a higher degree of civil liberties were higher.

Based on available evidence we can make the following important broad associations:

- Centralized systems are better at identifying poor communities via poverty maps than at identifying individual households or beneficiaries within communities (Demombynes and others 2002).

- Community-based targeting is more effective than targeting based on survey proxy indicators (Alderman 2002; Galasso and Ravallion 2001).

- When communities have a greater say, the pro-poor distribution of gains improves (Ravallion and Jalan 2000; Ravallion 2000).

- Community-managed projects are better maintained than projects managed by local governments (see Khwaja 2001 for an example from northern Pakistan).

- Communities that are allowed greater autonomy over the management of schools seem to increase the effectiveness of the schools and reduce absenteeism (see Jimenez and Sawada 1999 for an example from El Salvador's community-managed schools program and King and Ozler 2000 for an example from Nicaragua).

- Water systems perform markedly better in communities where households are able to make informed choices about the type of system and the level of service they require and where decisionmaking is genuinely democratic and inclusive (Katz and Sara 1997; see also Isham and Kahkonen 2002).

In Sub-Saharan Africa, several recent case studies point to similar conclusions. Successful cases span a wide area of activity as follows:

- Decentralization strategies include investment in rural government in Guinea using a bottom-up approach for the choice of services, management information systems, budget tracking systems, and "citizens score card" feedback from citizens; in Mbabane, Swaziland, a two-way communication system for better access to information and feedback from beneficiaries; and a public expenditure tracking system in Uganda.

- Community-driven development is being successfully implemented for a water supply project in Côte d'Ivoire, the social action funds in Malawi and Tanzania, the Zambia Social Investment Fund, and the Tunisia Northwest Mountainous Areas Development Project.

- Empowerment in education is under way in Madagascar, Tanzania, and Uganda.

- Citizen report cards, pioneered in Bangalore, India, are now being successfully implemented in Ghana.

- Participatory poverty diagnostics include participatory poverty assessments in Tanzania and Uganda.

Notwithstanding the early evidence, Mansuri and Rao (2003) argue that research does not yet provide clear evidence on how long-lasting or deep the results of participation in projects are. Does participation improve project sustainability? Does it develop social capital (the capacity for collective action is often considered to be an important by-product, and often an important objective, of participation)? Does it empower the participants, that is, enable them to influence institutions that affect their lives outside the boundaries of a project? These are relevant questions that point toward the need for understanding when participation in projects may not, by itself, produce better results, or for that matter, the limitations of inclusiveness as a strategy for achieving societal goals. Key problems relate to elite capture, agency capture, information asymmetry, and donor capture.

Key Challenges for Reforming Governance and Aid Principles and Practice

The hope for Sub-Saharan Africa's development is based partly on the continent's own history and partly on the experience of successful newly industrializing countries. In particular, the experience of the East Asian countries over the last 35 years shows that countries can break out of the poverty cycle and embark on sustained growth and poverty reduction, and more recently shows that they can recover from a crisis and resume robust growth. Based on Sub-Saharan Africa's own growth experience, roughly half of 21 Sub-Saharan African economies contained in a sample from a study by Pritchett (1997), accounting for nearly 80 percent of Sub-Saharan Africa's gross domestic product and population, experienced reasonably robust growth between 1960 and 1973, and the region as a whole achieved nearly 5 percent growth in per capita income during this time. During the 1960s, countries such as Côte d'Ivoire, Ghana, and Zambia had per capita incomes that exceeded those in several East Asian countries. Botswana, Mauritius, and the Seychelles have maintained steady high growth throughout the past four decades, with the first two being among the fastest growers worldwide. Botswana and Mauritius have maintained per capita annual income growth rates above 3 percent for nearly four decades, making major strides in living standards and placing them in the group of middle-income countries.

There is further cause for cautious optimism from recent developments taking place in the region, even as it continues to struggle with the legacy that characterizes much of the continent's negative reputation. During the last decade, many reforming Sub-Saharan African countries have made significant progress in reestablishing sustained macroeconomic stability, more open and liberal markets, and greater involvement of the private sector. Together with peace and security, African governments are increasingly accepting these changes as prerequisites for growth, development, and ultimately for the region's prosperity. About half of the countries in Sub-Saharan Africa are now

growing at 5 percent per year or more.[7] Many more countries in the region are successfully carrying out macroeconomic, structural, and trade reforms. Macroeconomic stability is returning to the continent with significantly lower inflation and narrower fiscal and external trade deficits, and more countries now have convertible domestic currencies for current account transactions.[8] In addition, since the early 1990s Sub-Saharan Africa has been undergoing a relatively rapid (by historical standards) process of political liberalization, including rising pressure for the devolution of authority to subnational government entities. This development, albeit sometimes halting, sets the stage for facing the development challenge in a more inclusive manner.

Yet despite the progress, sustained growth and development remain elusive and poverty remains preponderant. A World Bank (2000a) study identifies improvements in governance and conflict resolution as among the most basic requirements for faster development in the region. It asserts that widespread civil conflicts impose enormous costs, including on neighboring countries. The study argues that contrary to popular belief, Sub-Saharan Africa's conflicts do not stem primarily from ethnic diversity. Rather, in a pattern found around the world, conflicts are driven by poverty, under-development, and lack of economic diversification, as well as by political systems that marginalize large parts of a population. Undeniably, conflicts perpetuate poverty, creating a vicious cycle that can be reversed only through special development efforts, including long-run peace building and political reforms. Furthermore, strong evidence at a more aggregate level shows that economic inequality increases political instability and discourages investment (Alesina and Perotti 1993). With success in addressing these concerns, countries can grow rapidly, foreign capital inflows can increase, and flight capital can return.

Conclusions from studies of growth in 30 Sub-Saharan African countries under the African Economic Research Consortium emphasize the strong role of politics in policymaking. They also highlight the impact of politically motivated conflicts in halting growth and development. Note that most Sub-Saharan African countries studied under the consortium project were unable to break out of growth deceleration for nearly 15 years after most of them experienced major shocks during the 1970s and early 1980s. According to the studies, and in line with Rodrik (1998), the ability to manage the domestic social conflicts triggered by the turbulence of the world economy during the 1970s is what made the difference between continued growth and economic collapse. Where the onset of shocks coincided with regime changes toward weak autocracies, the downturn was more severe and protracted.

The New Partnership for African Development (NEPAD), an initiative by African heads of state for promoting recovery, has endorsed improved governance as a top priority, with an emphasis on inclusiveness, conflict prevention, and anticorruption activities. NEPAD sees this focus on governance as a way to improve Africa's collective reputation and credibility and to reduce impediments to attracting investment to the region. Participating countries have emphasized the need to ensure support for the initiative by a broad range of stakeholders in each country. They also have given priority to what Africans need to do for themselves, with external assistance and foreign private investment supplementing these efforts. The African peer review system has

been endorsed by the African Union as a key mechanism for region-wide pressure for positive change.

A major challenge to implementing NEPAD's vision is for governments to commit themselves to effective development contracts. Without such a commitment, ruling regimes will continue to maximize their own private gains or those of interest groups to which they belong (Dethier 1999; Hamdok 2002) at the expense of overall growth and development. As Hamdok (2002) cites, Bardhan (1997) and Singh (1999) suggest four approaches toward obtaining such commitment and to help ameliorate the effects of information asymmetries that typically make contracts incomplete: (a) governments should commit with credibility, (b) administrative systems should be accountable, (c) political and economic environments should minimize risk to enable longer time horizons for actors, and (d) domestic politics should encompass most interest groups and the political system should be open to contestation. In the same vein, this paper sees the devolution of decisionmaking and development management to subnational levels not merely as a tactic for making better use of funds, important as this is, but also as part of a strategy for better governance by rebuilding a social fabric and a social contract that were broken by the colonial imposition of alien forms of power and the subsequent abuse of those institutions by authoritarian governments.

A significant part of the challenge has to be faced by reengineering the aid relationship so that it is supportive of the development of durable accountability systems to make them consistent with more open governance structures. Domestic politics will ultimately dictate the speed with which Sub-Saharan African countries can move toward greater transparency and increased accountability: good governance cannot be sustained in an environment characterized by exclusive politics (Hamdok 2002). While donors have pursued conditionality as a commitment technology or external agency of restraint, failure to enforce it points to a need to find other ways to tie the hands of governments and to enable citizens to threaten the very existence of avenues for continued predation through a credible, but peaceful, threat of regime turnover. Political contestation would encourage greater transparency and increased accountability.

A rapid process of liberalizing political systems is under way in Sub-Saharan Africa, with greater acceptance of democracy and increased pressure for devolution of authority to subnational government entities. This increases the pressure for greater inclusiveness of stakeholders in designing development programs and for fundamental changes to be made in accountability systems to make them fit in with more open governance structures. This development poses perhaps the greatest challenge to reformulating the aid relationship. The current dual accountability system whereby recipient governments are accountable both to donors and to local constituencies may need to be reconfigured into a single integrated system. The growing emphasis on a partnership approach and local ownership is one potentially useful response, but to make partnership and local ownership a reality, the immense challenges involved in surmounting attitudes and changing procedures for aid management first have to be faced.

The main challenge at the national level is to redefine aid relationships in a manner that permits inclusive processes in designing, implementing, and monitoring development programs. Local ownership is at the heart of the new relationship. The strength of this type of relationship lies in its ability to elicit commitment and create conditions for local pressure to ensure accountability and prevent the wasteful use of resources. For this to occur, however, important preconditions must be fulfilled. The first is adequate capacity and resolve on the part of the recipient country to take up the lead role. The second is the development of an open political system and the upgrading of civil liberties to bring the policy process and information into the public domain and subject it to citizens' voice. The third is donors' willingness to wait for and adhere to coherent, home-grown strategies on which to base their specific assistance. In this last respect, promoting local ownership is consistent with better aid coordination and will help avoid the problems currently associated with the proliferation of enclave projects and unmanageable recurrent cost profiles resulting from this proliferation.

In this connection, important changes are underway that, if supported, hold good promise for a credible transition to a partnership approach. The advent of the Poverty Reduction Strategy (PRS) as a focal point for development programs in more than 25 Sub-Saharan African countries perhaps best exemplifies the convergence of these various developments into the practical arena. The ideal PRS is developed through a consultative process with civil society and donors as development partners. The PRS process emphasizes local ownership; a participatory approach to determining priority areas for poverty reduction programs; and a participatory monitoring system, which requires open accountability and feedback in relation to efforts made and the ensuing results and in order to bring the scrutiny of public spending programs into the public domain. The PRS also incorporates the reforms needed to enhance overall efficiency in resource application to achieve higher growth along with the necessary institutional reforms for more effective and transparent execution of programs. Given the inclusive process for developing the PRS, it opens up the discussion of policy reforms to a wider range of domestic stakeholders.

More important, the PRS is being adopted as a key strategic instrument outside the procedural requirements for debt relief. NEPAD has adopted the PRS as an inclusive approach to attack abject poverty by means of Africa's own undertakings and efforts, with assistance sought from development partners to supplement African efforts. Durable processes are also being developed for reviewing the annual progress of PRS implementation, including open scrutiny by local stakeholders. The African peer review system adopted under the NEPAD initiative provides for cross-country peer pressure to encourage progress toward good development management in the region. It supplements the domestic pressure for commitment, which helps tie governments' hands and signals the credibility of ongoing reforms to citizens and development partners. NEPAD is supplemented by several other regional and subregional associations and committees that buttress the cross-country peer review system, such as the Association of African Central Bank Governors, consultative forums, policy

convergence councils, a variety of private sector associations, and the Cross Border Initiative. Such associations and networks are typically endorsed by the formal institutions for regional cooperation. They operate through peer pressure for good practice and mutual support for improved performance; promote collaborative programs; and, in the case of private associations, exert pressure on public agencies to improve their operating environments.

Notes

1. Since attaining independence, economies in Sub-Saharan Africa have undergone a pattern of relatively high growth until 1973, a long period of stagnation between 1973 and 1993, and a modest recovery since 1994. The growth in Sub-Saharan African countries in comparison with that in other developing countries exhibits a sample difference in average growth rates for 34 years of 1.7 percent (Ndulu and O'Connell 1999). In per capita terms, the difference is nearly 2.2 percent because of Sub-Saharan Africa's higher population growth. Investment productivity in Sub-Saharan Africa for the entire postindependence period was only half of that in other developing regions (World Bank 2000a).

2. The government chooses I to maximize the discounted consumption of the elite:

   Max $W = C + \beta C_F = (fY - I) + \beta fY_F(I)$,

   where $\beta < 1$ is an exogenous discount factor. The government's optimal choice equates the marginal cost of providing I to its marginal benefit. Recasting the problem as equivalent to maximizing W/f rather than W, the marginal cost of infrastructure (in terms of W/f) is $1/f$, reflecting the concentration of its financing in the hands of the elite. The marginal benefit, in turn, is just the discounted marginal product of public infrastructure, $\beta \cdot dY_F/dI$. With diminishing marginal returns to infrastructure, this is declining in I. The socially efficient level of infrastructure is $1/f$, where the full social marginal cost ($= 1$) equals the social marginal benefit ($= \beta \cdot dY_F/dI$). The government chooses instead to trade off infrastructure for higher current consumption ($I^E < I^*$). A more encompassing elite chooses better policy.

3. I am grateful to Steve O'Connell for raising this point.

4. Important recent exceptions to this include Drazen (2002) and, in the case of multiple agents, governments, and donor agencies, Azam and Laffont (2003) and Hefeker and Michaelowa (2003).

5. I am greatly indebted to an anonymous referee of the outline for this point.

6. Bureaucratic quality includes the following characteristics: rule of law, predictability, extent of corruption, regulatory burden, voice, and accountability.

7. Sub-Saharan African countries with growth rates above 5 percent during 2001–02 include Botswana, Burkina Faso, Côte d'Ivoire, Equatorial Guinea, Ethiopia, Ghana, Malawi, Mali, Mauritius, Mozambique, Namibia, Rwanda, Senegal, the Seychelles, Swaziland, Tanzania, and Uganda.

8. The group of Sub-Saharan African countries with single-digit inflation is now larger than the group with double-digit inflation. Except in a few countries, black market premiums average just 4 percent. Through unilateral trade reforms, Sub-Saharan African countries have compressed both the tariff rates and categories with average rates of 15 percent. In a survey of 22 Sub-Saharan African countries in 1997, fewer business people saw the state as an opponent than they had 10 years earlier (World Bank 2000a).

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Adam, C., and Stephen A. O'Connell. 1999. "Aid, Taxation, and Development in Sub-Saharan Africa." *Journal of Economics and Politics* 2(3): 225–53.

Adam, C., B. Ndulu, and N. Sowa. 1996. "Liberalization and Seignorage Revenue in Kenya, Ghana, and Tanzania." *Journal of Development Studies* 32(4): 523–37.

Addison, T. 2003. "Do Donors Matter for Institutional Reform in Africa?" In S. Kayizzi-Mugerwa, ed., *Reforming Africa's Institutions: Ownership, Incentives, and Capabilities.* Tokyo: United Nations University Press.

Alderman, H. 2002. "Do Local Officials Know Something We Don't? Decentralization of Targeted Transfers in Albania." *Journal of Public Economics* 83.

Alesina, A., and R. Perotti. 1993. "Income Distribution, Political Instability, and Investment." Working Paper no. 4173. National Bureau of Economic Research, Cambridge, Mass.

Azam, J.-P. 1997. "Public Debt and Exchange Rate in the CFA Zone." *Journal of African Economies* 6: 54–84.

Azam, J.-P., and J.-J. Laffont. 2003. "Contracting for Aid." *Journal of Development Economics* 70: 25–58.

Bardhan, P. 1997. *The Role of Governance in Economic Development.* Paris: Organisation for Economic Co-operation and Development, Development Centre.

Bates, R. 1983. "Governments and Agricultural Markets in Africa." In D. Gale. Johnson and G. Edward Schuh, eds., *The Role of Markets in the World Food Economy.* Boulder, Colo.: West View Press.

Boone, P. 1996. "Politics and the Effectiveness of Foreign Aid." *European Economic Review* 40(2): 289–329.

Braetigum, D. 2000. *Aid Dependence and Governance.* Stockholm: Ministry of Foreign Affairs, Expert Group on Development Issues.

Bratton, M., and N. van de Walle. 1997. *Democratic Experiments in Africa.* Cambridge, U.K.: Cambridge University Press.

Burnside, C. and D. Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90(4).

Cernea, M. 1985. *Putting People First: Sociological Variables in Rural Development.* New York: Oxford University Press.

Chambers, R. 1983. *Rural Development: Putting the First Last.* London: Longman.

Chegge, M. 1999. "Politics of Development: Institutions and Governance." Paper presented at the workshop on Can Africa Claim the 21st Century? Abidjan, Côte d' Ivoire.

Cleaver, F. 2000, "Participatory Development at the World Bank: The Primacy of Process." In B. Cooke and U. Kothari, eds., *Participation the New Tyranny*? London: Zed Books.

Coate, S., and S. Morris. 1996. "Policy Conditionality." University of Pennsylvania, Department of Economics, Philadelphia. Processed.

Collier, P., and D. Dollar. 2000. "Does Africa Need a Marshal Plan?" *Cambridge Review of International Affairs* 14(2).

Collier, P., and J. Gunning. 1999. "Explaining African Economic Performance." *Journal of Economic Literature* 37(1): 64–111.

_____. 1998. "Trade Policy and Regional Integration: Implications for the Relations between Europe and Africa." In C. Miller, ed., *Developing and Newly Industrializing Countries*. Cheltenham, U.K.: Edward Elgar.

Demombynes, G., C. Elbers, Jean O. Lanjouw, and Peter Lanjouw. 2002. *Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries*. World Institute for Development Economics Research Discussion Paper no. 2002/39. Tokyo: United Nations University.

Demsetz, H. 1967. "Towards a Theory of Property Rights." *American Economic Review Papers and Proceedings* 57(2): 347–59.

Dethier, J. J. 1999. *Governance and Economic Performance: A Survey*. Discussion Paper on Development Policy no. 5. Bonn: University of Bonn, Centre for Development Research.

Devarajan S, D. Dollar, and T. Holmgren. 2001. *Aid and Reform in Africa*. Washington, D.C.: World Bank.

Diwan, I., and D. Rodrik. 1992. *External Debt, Adjustment, and Burden Sharing: A Unified Framework*. Princeton Studies in International Finance no. 73. Princeton, N.J.: Princeton University.

Dollar, D., and J. Svensson. 2000. "What Explains the Success or Failure of Structural Adjustment Programs?" *Economic Journal* 110(October): 894–917.

Drazen, A. 2002. *Conditionality and Ownership in IMF Lending: A Political Economy Approach*. IMF Staff Papers vol. 49. Washington, D.C.: International Monetary Fund.

Easterly, W. 2002. "The Cartel of Good Intentions: Bureaucracy Versus Markets in Foreign Aid." Working Paper no. 4. Washington, D.C., Institute for International Economics, Center for Global Development.

Galasso, E., and M. Ravallion. 2001. "Decentralized Targeting of an Anti-Poverty Program." World Bank, Development Research Group, Washington, D.C. Processed.

Gordon, A. A. 1990. *Transforming Capitalization and Patriarchy*. Boulder, Colo., and London: Lynne Peenner.

Hamdok, A. 2002. "Governance and Policy in Africa: Recent Experiences." In Steve Kayizi-Mugerwa, ed., *Reforming Africa's Institutions: Ownership, Incentives, and Capabilities*. Tokyo: United Nations University Press.

Hardin, G. 1968. "The Tragedy of the Commons." *Science* 162: 1243–48.

Hefeker, C., and K. Michaelowa. 2003. *Can Process Conditionality Enhance Aid Effectiveness? The Role of Bureaucratic Interest and Public Pressure*. Discussion Paper no. 239. Hamburg, Germany: Hamburg Institute of International Economics.

Hirschman, A. 1984. *Getting Ahead Collectively: Grassroots Experiences in Latin America*. New York: Pergamon Press.

Humphreys, M., and R. Bates. 2001. "Economic Policies in Africa: The Good, the Bad, and the Ugly." Paper presented at the 97th Annual Meeting of the American Political Science Association, August 30–September 2, 2001, San Francisco.

Hyden, G. 1983. *No Short Cut to Progress: African Development Management in Perspective*. London: Heinemann.

_____. 1986. "The Anomaly of the African Peasantry." *Development and Change* 17(4): 677–704.

Isham, J., and S. Kahkonen. 2002. "Institutional Determinants of the Impact of Community-Based Water Services: Evidence from Sri Lanka and India." *Economic Development and Cultural Change* 50(2): 667–91.

Isham, J., and D. Kaufmann. 1999. "The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects." *Quarterly Journal of Economics*. 11(2): 149–84.

Isham, J., D. Kaufmann, and L. H. Pritchett. 1997. "Civil Liberties, Democracy, and the Performance of Government Projects." *World Bank Economic Review* II(2): 219–42.

Isham, J., D. Narayan, and L. H. Pritchett. 1995. "Does Participation Improve Performance? Enabling Causality with Subjective Data." *World Bank Economic Review* 9(2): 175–200.

Jiminez, E., and Y. Sawada. 1999. "Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program." *World Bank Economic Review* 13(3): 415–42.

Johnson, J. H., and S. Wasty. 1993. *Borrower Ownership of Adjustment Programs and the Political Economy of Reform*. Discussion Paper no. 199. Washington, D.C.: World Bank.

Kanbur, R. 1999. "Prospective and Retrospective Conditionality: Practicalities and Fundamentals." In P. Collier and C. Pattilo, eds., *Investment and Risk in Africa*. Basingstoke, U.K.: Macmillan.

Katz, Travis, and J. Sara. 1997. "Making Rural Water Supply Sustainable: Recommendations from 'Global Studying'." United Nations Development Programme and World Bank, Water and Sanitation Program, Washington, D.C. Processed.

Kaufmann, D. 2000. " 'Unbundling' Governance and Corruption: Some New Empirical Findings and Collective Action Approaches." Paper presented at the Anticorruption Summit, September, World Bank Institute, Washington, D.C.

_____. 2002. "Rethinking Governance: Questioning from Empirical Perspective." Paper prepared for Governance Flagship Course, World Bank Institute, December 10, 2002, Washington, D.C.

Kaufmann D., G. Mehrez, and T. Gurgur. 2002. "Voice or Public Sector Management? An Empirical Investigation of the Determinants of Public Sector Performance Based on a Survey of Public Officials." World Bank, Washington, D.C. Processed.

Kaul, I., K. Le Goulven, and M. Schnupf. 2002. "Financing Global Public Goods: Policy Experience and Future Challenges." In I. Kaul, K. Le Goulven, and M. Schnupf, eds., *Global Public Goods Financing: New Tools for New Challenges—A Policy Dialogue*. New York: United Nations Development Programme.

Khan, M., and S. Sharma. 2001. "IMF Conditionality and Country Ownership of Programs." Working Paper no. 01/142. International Monetary Fund, Washington, D.C.

Khwaja, A. I. 2001. "Can Good Projects Succeed in Bad Communities? Collective Action in the Himalayas." Harvard University, Cambridge, Mass.: Processed.

Killick, T. 1995. "Conditionality and the Adjustment-Development Connection." *Pakistan Journal of Applied Economics* 11(Summer/Winter): 17–36.

_____. 1997. "Principals and Agents and the Failings of Conditionality." *Journal of International Development* 9(4): 435–95.

King, E., and B. Ozler. 2000. "What's Decentralization Got to Do with Learning? Endogenous School Quality and Student Performance in Nicaragua." World Bank, Development Economics Research Group, Washington, D.C. Processed.

Knack, Stephen, and Phillip Keefer. 1997. "Does Social Capital Have an Economic Pay Off? A Cross Country Investigation." *Quarterly Journal of Economics* 112(November): 1251–88.

Knack, S., and A. Rahaman. 2003. "Donor Fragmentation and Bureaucratic Quality in Aid." Public Sector Management Working Paper no. 3186. World Bank, Washington, D.C.

Lensink, R., and H. White. 1999. *Aid Dependence: Issues and Indicators*. Stockholm: Ministry of Foreign Affairs, Expert Group on Development Issues.

Mansuri, G., and V. Rao. 2003. "Evaluating Community Driven Development: A Review of the Evidence." Policy Research Working Paper no. 3209. World Bank, Washington, D.C.

McGuire, M., and M. Olson, Jr. 1996. "The Economics of Autocracy and Majority Rule." *Journal of Economics Literature* 34: 72–96.

Miguel, E. 2000. "Ethnic Diversity and School Funding in Kenya." University of California at Berkeley, Department of Economics, Berkeley, Calif.: Processed.

Narayan, D. 2002. *Empowerment and Poverty Reduction: A Source Book*. Washington, D.C.: World Bank.

Ndulu, B. J. 1986. "Governance and Economic Management." In R. Berg and J. Whitaker, eds., *Strategies for African Development*. Berkeley, Calif.: University of California Press.

_____. 2002. "Partnership, Inclusiveness, and Aid Effectiveness in Africa." In V. FitzGerald, ed., *Social Institutions and Economic Development: A Tribute to Kurt Martin*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Ndulu, B. J., and S. A. O'Connell. 1999. "Governance and Growth in Sub-Saharan Africa." *Journal of Economics Perspectives* 13(3): 41–66.

North, D., and R. P. Thomas. 1977. "The First Economic Revolution." *Economic History Review* 30(2): 228–241.

Olson, M. 1973. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, Mass.: Harvard University Press.

_____. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*. New Haven, Conn.: Yale University Press.

Pritchett, L. 1997. "Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains." Working Paper no. 1947. World Bank, Washington, D.C.

Pritchett, L., and M. Woolcock. 2002. "Solutions When the Solution Is in the Problem: Arraying the Disarray in Development." Working Paper no. 10. Harvard University, Center for Global Development, Cambridge, Mass.

Ravallion, M. 2000. "Monitoring Targeting Performance When Decentralized Allocations to the Poor Are Unobserved." *World Bank Economic Review* 14(2): 331–46.

Ravallion, M., and J. Jalan. 2000. "Is Transient Poverty Different? Evidence for Rural China." *Journal of Development Studies* 36(6).

Reinikka, R., and J. Svensson. 2001. *Explaining Leakage of Public Funds*. Discussion Paper no. 3227. London: Centre for Economic Policy Research.

Ritzen, J., W. Easterly, and M. Woolcock. 2000. "On 'Good' Politicians and Bad Policies: Social Cohesion, Institutions, and Growth." World Bank, Research Development Group, Washington, D.C.

Rodrik, Dani. 1998. "Where Did All The Growth Go? External Shocks, Social Conflict, and Growth Collapses." Working Paper no. 6350. National Bureau of Economic Research. Cambridge, Mass.

Sachs J. 1989. "Conditionality, Debt Relief, and Developing Country Debt Crisis." In *Developing Country Debt and Economic Performance,* vol. 1, *International Financial System.* Chicago: Chicago University Press.

Sandbrook, R., and J. Barker. 1985. *The Politics of Africa's Economic Stagnation*. Cambridge, U.K.: Cambridge University Press.

Singh, S. 1999. "The Political Underpinnings of Economic Governance." Paper presented at the workshop on Indicators for Monitoring Progress towards Good Governance in Africa, United Nations Economic Commission for Africa, Addis Ababa, Ethiopia.

Stern, N. 2002. *A Strategy for Development*. Washington, D.C.: World Bank.

Tsikata, Y. 2002. "Owning Economic Reforms: A Comparative Study of Ghana and Tanzania." In Steve Kayizi-Mugerwa, ed., *Reforming Africa's Institutions: Ownership, Incentives, and Capabilities*. Tokyo: United Nations University Press.

World Bank. 1999. *Assessing Aid: What Works, What Doesn't, and Why*. New York: Oxford University Press.

_____. 2000a. *Can Africa Claim the 21st Century?* Washington, D.C.

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

_____. 2003. *World Development Report 2004: Making Services Work for Poor People*. New York: Oxford University Press.

Wuyts, M. 1996. "Foreign Aid, Structural Adjustment, and Public Management: The Mozambican Experience." *Development and Change* 27(4): 717–49.

Comment on "Community-Based Development in the Context of Within-Group Heterogeneity" by Jean-Philippe Platteau and "Inclusiveness, Accountability, and Effectiveness of Development Assistance in Sub-Saharan Africa" by Benno J. Ndulu

ROHINI NAYYAR

The papers by Jean-Philippe Platteau and Benno Ndulu both address an important issue, namely, the role of participation in the development process. The analyses focus on Africa, with special reference to development assistance. The papers recognize that people's voice in the design and execution of development projects and programs is imperative if they are to have a positive effect on people's lives. A system run by a few for the benefit of the many poor and marginalized does not have the desired outcomes. This is true in the African context where, on average, growth has been sluggish and poverty persists.

The two papers diverge in their views on the role of community-based development (CBD) in Africa. Platteau argues that elite capture is the likely outcome at the local level given populations' underdevelopment. This happens in all development programs, especially in donor-assisted programs. The paper summarizes the literature on the dilemma of decentralization, where the advantages that could accrue to project design and implementation because of involvement by local people could be offset by the threat of more vocal and organized groups cornering the benefits to the detriment of the truly poor for whom the program was intended. The paper accepts the advantages of CBD and goes on to suggest ways to surmount the problem of elite capture. One approach is for a donor agency to alert all other donors if particular local governments or community-based organizations misappropriate funds, or what Platteau refers to as the multilateral reputation mechanism. Another is to opt for sequential disbursement of aid money to control misappropriation. A third option is for aid agencies to promote competition among local leaders. The paper concludes by emphasizing that the solution to elite capture lies in people's empowerment, which is a lengthy process, because it requires building awareness and increasing access to information.

Ndulu, while recognizing the problem of elite capture in countries with authoritarian rule, is in favor of CBD to enhance transparency in the execution of development projects. In the case of donor-funded projects, donors have imposed aid conditionalities to ensure a greater voice for the people; however, even where recipient governments have accepted conditionalities, they have not always honored their commitments. Ndulu draws upon micro-level studies to show that participation by local people in project design and implementation has led to more effective targeting of programs for the poor.

The main challenge, therefore, is to redefine aid relationships so that recipient countries play the main role in designing, implementing, and monitoring programs in cooperation with the donors. However, local ownership should not be confined to national governments, but should include civil society. In addition, greater decentralization to subnational levels is needed, as this would give rise to empowerment through greater access to information, enhanced accountability, and increased local capacities and participation. The key to development is governance. In the last decade, African countries have embarked on reforms with greater involvement by the private sector and a more liberalized economic system. Nevertheless, progress has been slow, with poverty and low rates of economic growth persisting.

The main issue emerging from the two papers centers around the efficacy of CBD, as opposed to a top-down system of delivery of services for the poor. Several developing countries, including India, are currently pursuing inclusiveness, participation, and voice in the execution of development programs. The two papers focus on CBD through facilitators, nongovernmental organizations (NGOs), and user groups. An alternative system is elected bodies at the district and sub-district levels right down to the village. One example of this is the *panchayati raj* system in India.

India has a large NGO sector that is growing rapidly. Some of the NGOs have done excellent work. They have been funded by international agencies, the government, and private sources. For these organizations, resources are not a constraint. Their major problem is widening their operations without diluting the standards of service they have set for themselves. By contrast, many other organizations have mushroomed simply to take advantage of the funding opportunities that have become available in recent years both because of increases in foreign funding and because of government policy to promote development by the voluntary sector. The Council for Advancement of People's Action and Rural Technology (CAPART), a government agency that provides financial and technical support to NGOs for rural development in India, has been grappling with this problem of identifying the good NGOs for funding and rejecting the bogus NGOs. Many NGOs have been blacklisted by one Government Department or another or by CAPART; however, they simply reappear under a different name.

The international aid agencies would have to face this problem even if the multilateral reputation mechanism worked effectively. The problem donor agencies face in reaching communities through NGOs and local leaders without formal titles is more acute precisely because these organizations are not subject to the rules and regulations

that local governments and semigovernment structures have to adhere to. In the event that local organizations embezzle funds, the only option available to the aid agency may be to file criminal charges against the NGO or local leaders. Given the inadequacies of the judicial process, this is not a realistic option.

More recent years have seen the evolution of several community organizations in India, including self-help groups, water user groups, watershed associations, joint forest management committees, and village and district education committees. Again, the impact of these has been disparate. In some areas and in some sectors they have succeeded in mobilizing people so that they can be effective in demanding services that meet their perceived needs, and also in designing and implementing projects and programs. Consider the following examples.

People's participation and its impact on the delivery of health services has been amply demonstrated by the Rogi Kalyan Samiti initiative in Madhya Pradesh. In the wake of an epidemic of bubonic plague in 1994, the district administration of Indore, along with the people of the district, initiated efforts to prevent an outbreak of the disease in the district. Maharaja Yashwant Rao Hospital was completely evacuated and its patients shifted to another hospital, and the district's residents cooperated in a cleanup. This experiment in participation was later institutionalized for hospital management across the state. The Rogi Kalyan Samiti initiative also worked out a schedule of user charges to raise resources for continuous improvements in hospital infrastructure. This innovation has been widely commended. In 2000, the project won the first Annual Global Network Award in Tokyo.

The NGO Tarun Bharat Sangh (TBS) was established in 1975 to help manage natural resources. TBS concentrated on the harvesting and conservation of water, the most critical resource in Rajasthan, one of India's driest states. TBS recognized that water held the key to sustainable livelihoods in the area. Unless people were motivated, organized, and cooperated with each other in regenerating the traditional water harvesting structures, meeting their water needs would be difficult. TBS undertook walking tours to bring about awareness among the people and, to solicit their cooperation, held *panchayat* meetings for the same purpose. Their objective was to undertake a holistic development of the catchment areas. TBS raised finances by taking contributions from villagers for restoration of the water harvesting structures. For mediating social conflicts among villages, and even within a community, TBS set up water user committees in the catchment area of local rivers. This committee decides on the rules regarding the use of water, and villagers themselves enforce the rules. As a result of TBS's efforts, five rivers that used to dry up after the monsoon have become perennial. The head of TBS received the Magasaysay award two years ago.

In Andhra Pradesh, the self-help movement has led to the empowerment of women. Women are mobilized and organized into groups that focus on thrift and savings. As a result, they are able to raise bank loans and obtain government subsidies to start micro-enterprises. The formation of self-help groups has itself provided the women with a potent tool for their socioeconomic betterment. The self-help group concept is being replicated throughout the country with differing levels of success.

Even in a democratic country like India, where democratic decentralization has been effected through a constitutional mandate, elite capture takes place at different levels in the administrative hierarchy. At the local district and subdistrict levels, the *panchayati raj* institutions are riddled with corruption, and the capture of power and resources by a few elected representatives is widespread. Such representatives have tended to collude with the official machinery to reinforce the unequal access to power and resources that characterize village life. Yet this system has its advantages. Reservations for women and marginalized groups in the *panchayats* are certainly creating awareness about the potential of local self governance, and over time will empower them, first politically and then economically. I also believe that decentralized development will ensure that funds flow to villages and that monies will be spent there. This too would create employment and incomes, as well as community assets at the local level, which would impinge favorably on the lives of the poor. Over the last decade, several instances have come to light of motivated and dynamic heads of *panchayat* who have mobilized the people for the common good.

To make *panchayats* transparent and accountable, the *gram sabha*, a body that comprises all adult members (including women) of the community residing in a *panchayat*, has to be strengthened. *Gram sabhas* should have full powers to determine priorities in villages and approve budgets, and they should be involved in managing natural resources. They must perform the role of a watchdog, with the *panchayats* being held accountable to them. The empowerment of both *panchayati raj* institutions and the *gram sabhas* is at different levels of development in the various states. It is most effective in Kerala, which launched the People's Planning Campaign in 1996. In Kerala, 35 to 40 percent of the state's plan outlay is devolved to local bodies, and people's needs are identified at *gram sabhas* meetings. This results in the preparation of an annual plan for each *gram panchayat* that is approved by *gram sabhas*. Individual annual plans are then integrated into block and district plans. Kerala's high literacy levels have helped engender awareness among villagers and have facilitated the training of functionaries. Even in other states, a system of social auditing by the *gram sabhas* is being developed, accompanied by the enactment of freedom of information legislation.

Thus decentralization and participation are critical for the success of development projects and programs, whether through civil society organizations, community-based organizations and user groups, and/or elected local bodies. However, they are by no means sufficient. Appropriate macroeconomic policies, economic diversification, and greater investment in human capital and physical infrastructure are necessary to boost the development process, especially in Africa. Whether funded through domestic resources or bilateral development assistance from donors or multilateral agencies, expenditures should conform to local priorities and be subject to public scrutiny. I am glad that themes such as inclusiveness and people's voice are receiving the attention of economists and donor agencies, with their outcomes being assessed by the impact of these social variables on people's lives.

Comment on "Community-Based Development in the Context of Within-Group Heterogeneity" by Jean-Philippe Platteau and "Inclusiveness, Accountability, and Effectiveness of Development Assistance in Sub-Saharan Africa" by Benno J. Ndulu

VIJAYENDRA RAO

The papers by Jean-Philippe Platteau and Benno J. Ndulu present contrasting views of the efficacy of community-based development (CBD). For Ndulu, CBD is an important part of the solution to the central problems of poor accountability and transparency that have plagued African development. For Platteau, also talking mainly from a Sub-Saharan African perspective, CBD is eminently vulnerable to the problem of capture by entrenched and powerful elites, particularly in the context of weak states and the extreme disempowerment of the poor. As a discussant I therefore seem to be located in the middle of a serious debate, and one potentially useful role I could play is to review some of the thinking and evidence on this topic to place the debate in context. Therefore, to the exclusion of the many other interesting issues these papers raise, I will focus on shedding some light on the question of who is right on CBD. Is CBD, as Ndulu says, achieving substantial results on the ground, or, to paraphrase Platteau, is it leading to an expansion of opportunities for local elites to direct development funds toward their private benefit? For these comments I will draw heavily from a recent review of the literature (Mansuri and Rao forthcoming).

Ndulu cites various papers to claim that "when projects are conceived, designed, and executed with the participation of local populations, they tend to be more successful than those without such participation." However, every one of the papers Ndulu cites in support of this claim simply establishes a correlation between greater participation and successful projects. None satisfactorily demonstrates a causal link. Therefore what we know is that communities that have the wherewithal—that is, the social capital, or what one might more accurately call the capacity for collective action (Woolcock and Narayan 2000)—to engender more participation and inclusion are better able to manage their own projects. It is a long leap from here to push for community-based approaches as being the solution to the problem of how to

improve information and manage development projects at the local level. As Mansuri and Rao (forthcoming) demonstrate, there is an astonishing lack of reliable, representative impact evaluations of community-based projects, and the assumption that they are the next panacea for development seems more grounded in the cycles of development fashion trends than in results. The crucial question here is whether community-based approaches work better than other approaches in delivering services to the poor. While both theory and anecdotal evidence suggest that the approach could be promising, with one exception—a study looking at 60 water projects in northwest Pakistan that found that community-managed projects were generally better maintained than those managed by others (Khwaja 2001)—we know very little.

Another problem is that CBD is sometimes confused with empowerment, or processes that improve the agency of the poor. As Platteau, here and in previous work (Abraham and Platteau 2004), argues, in situations with preexisting economic and cultural inequities, which according to Platteau characterize most of Sub-Saharan Africa, CBD is likely to fail as a device for improving the welfare of the poor.

While I think this basic caution is not inappropriate, economic and social inequality is not always deeply entrenched and CBD might well be the optimal approach to pursue in these circumstances. Furthermore, even in situations of deep inequality CBD might work reasonably well. For instance, a case study of the Jamaican Social Investment Fund finds that distinguishing between elite domination that is malevolent and that which is benevolent is important (Rao and Ibanez forthcoming). In this case, while elites certainly dominated community-based processes of decisionmaking, these elites were often acting in the broader interests of the community. CBD requires relatively well-educated people to manage the complex process of applying for funds, managing budgets and accounts, filling in forms, and so on. This is something that necessarily requires a relatively high level of education, which typically means that leaders tend to be elites. However, even in the most unequal communities (such as those that are characterized by deep divisions by landholdings and caste), elites gain a great deal of public regard and status by doing good. Thus effectively managing a community-based project could lead to a nonmonetary private benefit for the elite. It is thus not clear that elite capture will always be malevolent, though this is not to say that instances of outright corruption and theft do not occur; however, this is not an inevitable outcome.

Note that at a conceptual level, much of the policy discussion on CBD is characterized by an element of naivety. Take, for instance, the concept of participation, which is assumed to be inherently good. As Mansuri and Rao (forthcoming) argue, what this concept implies in practice is not that straightforward, and participation as an idealized, tranformational concept has been challenged on a number of grounds. First, observers have noted that the exercise of voice and choice could be quite costly under certain conditions. At the most trivial level it may involve real or imputed financial losses resulting from the time commitments required for adequate participation.

Second, participation may lead to significant psychological, and even physical, duress for the most socially and economically disadvantaged, who are typically the prime potential beneficiaries of CBD projects, because genuine participation for such groups may require taking up positions that are contrary to the interests of more powerful groups. The premise of participatory approaches is that the potential benefits of participation generally outweigh such costs, but this is by no means obvious. The mainstreaming of participation has also made it an instrument for promoting pragmatic policy interests, such as cost-effective delivery, low maintenance costs, and so on, rather than a vehicle for the radical transformation of society. The main outcome in such contexts may simply be to shift some of the costs of service delivery to potential beneficiaries.

Third, the notion that exposure to participatory experiences will transform the attitudes and implementation styles of authoritarian bureaucracies (governments or donors) may be quite naive. The routinization of participatory planning exercises into the work of public sector implementation agencies puts new pressures on resources, while leaving implementers unclear about their potential gains from this new accountability.

Moreover, by means of careful ethnographic work Mosse (2001) and others show that the actual practice of participation may reflect an imposition of donors' point of view rather than a listening and learning process whereby donors learn from beneficiaries. This is often unintended but is a consequence of the elite status of donor representatives within a community. In an effort to please donors, beneficiaries may tell them exactly what they want to hear. Platteau argues that this is based on a strategy to obtain funds, but it could also be based on a less sinister desire to appear accommodating and to please the "important" folks from the city. Thus the exercise of voice has to be viewed as a social construction, as a reaction, and as a strategy. A development worker for the Department for International Development tells a story about going to a village in southern India to conduct a participatory rural appraisal. He asked the villagers why they thought they were poor, upon which they promptly started drawing a hexagonal representation of the nexus of poverty that was exactly reproduced from the department's poverty strategy. In other words, instead of hearing the voices of the poor we may be hearing the voices of the donors.

Another key concern with the CBD methodology is the manner in which it is often implemented. Mansuri and Rao (forthcoming) review a great deal of qualitative evidence that suggests that CBD projects often work with young, inexperienced facilitators. These "street-level bureaucrats" are critical actors in building participatory processes, and a large burden is placed on them. They are required, at least in theory, to mobilize communities, build communities' capacity for collective action, ensure adequate representation and participation, and, where necessary, break through elite domination. To do this effectively they must be culturally and politically sensitive, charismatic leaders, trainers, anthropologists, engineers, economists, and accountants. Despite their centrality to the CBD process, however, virtually no generalizable

evidence on their role in CBD initiatives is available. What evidence there is comes from case studies, which present a rather dismal picture.

Jackson's (1997) analysis of field workers' diaries in India indicates that field-level personnel tend to be driven by the incentives they face and that these are often not well aligned with the needs of the project. In particular, Jackson notes that field workers tend to gloss over local power relationships in a rush to show results. Vasan (2002) shows how facilitators in the forest management projects she looked at overlooked project goals when their personal incentives contradicted them. Botchway (2001), studying participatory projects in north Ghana, notes that project facilitators are often quite vulnerable to manipulation and control by locally powerful individuals and groups because they are typically young, inexperienced, and poorly paid. Michener (1998) echoes this and points out that the problem extends beyond field-level implementation staff. Senior members of the implementing agency are often former bureaucrats with little experience in community empowerment work. Their vision of development, guided by a lifetime of work with line ministries, is closer to what Tendler and Serrano (1999) call supply-driven/demand-driven development.

So where does this leave us? Is CBD to be abandoned as a strategy? Is its fate to become yet another discarded fad in the annals of development history? This conclusion seems extreme. There are several stories of highly successful initiatives that provide grounds for much optimism. An obvious example is the inspiring story of Sewa (Rose 1993), which, inspired by Gandhian notions of self-reliance, provides inexpensive credit and insurance to poor women workers. While Sewa has never been subjected to an impact evaluation, it is difficult to imagine that it has not been a successful effort given the number of women who claim to have benefited from it. The success of the Porto Allegre, Brazil, participatory budgeting movement, which, by all accounts, has strengthened the hand of the poor in directing more equitable budgetary allocation, is another iconic example (Calderon and Szmuckler 2004). Similarly, the Amul movement in India has created a cooperative that has been remarkably successful in creating new markets and higher incomes for dairy farmers across the country (Kurien 1997).[1]

The question is whether such movements can be rapidly replicated by the force of an external intervention led by a large bureaucracy. There are two problems here. One is that it is difficult to mimic the success of a highly motivated group of charismatic individuals who are able to sustain a long-term vision of structural transformation with dedication, patience, and creativity. Such individuals have been integrally involved in each success story. When such tasks are handed over to salaried professionals motivated by more mundane preoccupations, such as wages and promotions, the incentives change, usually for the worse.

What are the prerequisites for successfully conducting large-scale, community-based projects that are truly community driven and are not merely "boutique" interventions confined to a few lucky villages? Mansuri and Rao (forthcoming) argue that first, the process has to be gradual. Given the contextual complexities involved, initial designs based on best practices are bound to be imperfect; indeed,

the best practice may be the absence of a best practice. Rapidly scaling these initial designs up, particularly in countries that have had little experience with community-based projects, will likely result in failed projects and, ultimately, a backlash against CBD.

Second, the process requires a strong ethic of learning-by-doing. This requires effective and honest evaluations, with good treatment, control groups, baseline and follow-up data, and reliable monitoring systems to provide constant feedback. Ideally these should mix qualitative and quantitative methods to provide both reliable estimates of impact and an in-depth examination of context and process (Rao and Woolcock 2003). Lessons learned from monitoring and evaluation should be incorporated into the next phase of project design to correct mistakes, which are inevitable.

Third, careful and adequate attention should be paid to the training and development of a core cadre of facilitators, because they are the fulcrum of successful community-based interventions. This too cannot be done in haste and is part of the learning-by-doing process. Inexperienced facilitators should be given a chance to learn and grow under the supervision and leadership of more experienced individuals.

Fourth, the country should be committed to a cultural change in the institutional environment, which has to be become more participatory, responsive, transparent, and accountable. These are not just buzzwords. As Uphoff, Esman, and Krishna (1998, p. 202) argue: "If the expansion is occurring because the government or donor sources are promoting it, this is less persuasive than if there is a spontaneous joining of the program or if local governments take over responsibility." Without an enabling institutional environment that is committed to bottom-up development, communities will quickly face the problem of dashed expectations by encountering unresponsive bureaucrats.

Finally, and perhaps most important, all this requires an extremely long-term horizon. Both institutions such as the World Bank and countries that take on a CBD agenda need to realize that changing from top-down to bottom-up development in a manner that is effective and sensitive to local context and culture cannot be done hastily. CBD should also not be judged hastily: initial evaluations may well be poor. The key is to fix the problems observed in the evaluations and to work toward incremental improvements. All this requires a new vision for development that is long-term, well evaluated, honest, and open to errors and to learning from errors and less prone to the fashions of the moment. Absent these conditions, instead of turning development upside-down, CBD is likely to join the long list of discarded fads that litter the history of development. Therefore one valid issue that needs to be debated is how the current institutional culture within multilateral organizations needs to change so that enthusiastic adoption of the CBD strategy would be to the organizations' comparative advantage.

Therefore neither Ndulu nor Platteau is completely right. The truth lies somewhere in between—perhaps. We cannot know for sure because of a lack of reliable evidence.

Note

1. Uphoff, Esman, and Krishna (1998) document several similar examples.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

_____. 2004. "Participatory Development: When Culture Creeps In." In Vijayendra Rao and Michael Walton, eds., *Culture and Public Action*. Stanford, Calif.: Stanford University Press.

Botchway, Karl. 2001. "Paradox of Empowerment: Reflections on a Case Study from Northern Ghana." *World Development* 29(1): 135–53.

Calderon, Fernando, and Alicia Szmuckler. 2004. "Political Culture and Development." In Vijayendra Rao and Michael Walton, eds., *Culture and Public Action*. Stanford, Calif.: Stanford University Press.

Jackson, Cecile. 1997. "Sustainable Development at the Sharp End: Field-Worker Agency in a Participatory Project." *Development in Practice* 7(3): 237–47.

Khwaja, Asim Ijaz. 2001. "Can Good Projects Succeed in Bad Communities? Collective Action in the Himalayas." Harvard University, Cambridge, Mass. Processed.

Kurien, V. 1997. "The AMUL Dairy Cooperatives: Putting the Means of Development into the Hands of Small Producers in India." In Anirudh Krishna, Norman Uphoff, and Milton J. Esman, eds., *Reasons for Hope: Instructive Experiences in Rural Development*. West Hartford, Conn.: Kumarian Press.

Mansuri, Ghazala, and Vijayendra Rao. Forthcoming. "Community Based (and Driven) Development: A Critical Review." *World Bank Research Observer.*

Michener, Victoria. 1998. "The Participatory Approach: Contradiction and Co-option in Burkina Faso." *World Development* 26(12): 2105–18.

Mosse, David. 2001. "'People's Knowledge,' Participation, and Patronage: Operations and Representations in Rural Development." In Bill Cooke and Uma Kothari, eds., *Participation: The New Tyranny*. London and New York: Zed Books.

Rao, Vijayendra, and Ana Maria Ibanez. Forthcoming. "The Social Impact of Social Funds in Jamaica: A Mixed-Methods Analysis of Participation, Targeting, and Collective Action in Community Driven Development." *Journal of Development Studies.*

Rao, Vijayendra, and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In François Bourguignon and Luiz Pereira Da Silva, eds., *Evaluating the Poverty and Distributional Impact of Economic Policies: A General Introduction*. Washington, D.C.: World Bank.

Rose, Kalima. 1993. *Where Women Are Leaders: The Sewa Movement in India*. London: Zed Books.

Tendler, Judith, and Rodrigo Serrano. 1999. *The Rise of Social Funds: What Are They a Model of?* New York: United Nations Development Programme.

Uphoff, Norman, Milton J. Esman, and Anirudh Krishna. 1998. *Reasons for Success: Learning from Instructive Experiences in Rural Development*. West Hartford, Conn.: Kumarian Press.

Vasan, Sudha. 2002. "Ethnography of the Forest Guard: Contrasting Discourses, Conflicting Roles, and Policy Implementation." *Economic and Political Weekly* XXXVII(40): 4125–33.

Woolcock, Michael, and Deepa Narayan. 2000. "Social Capital: Implications for Development Theory, Research, and Policy." *World Bank Research Observer* 15(2): 225–49.

Scaling Up and Evaluation

Using Randomized Experiments and Structural Models for Scaling Up: Evidence from the PROGRESA Evaluation

ORAZIO P. ATTANASIO, COSTAS MEGHIR, AND MIGUEL SZÉKELY

The evaluation of welfare programs, and more generally of government or international organization interventions, is often posed as a one-off question, in that evaluators ask whether a specific intervention achieves a specific objective in a specific situation. Recently, however, the more general question of whether results from given studies can be used to predict the effects of other interventions, possibly in different contexts, has received a considerable amount of attention. The ability to extrapolate success stories and avoid failures in different situations would obviously be highly desirable. Unfortunately, a rigorous and successful extrapolation is extremely difficult. This paper discusses the issues involved in the evaluation of social interventions and attempts to scale them up. In particular, it discusses the relative merits of (a) nonparametric evaluation strategies that rely on possibly experimental exogenous variations to estimate the impact effects and (b) more structural approaches. The difference between the two approaches is particularly relevant in relation to extrapolation and scaling up. Two types of extrapolation are possible: (a) predicting the effects of a program that is different from the one that was evaluated, or (b) predicting the effects of exporting an existing program from a context where it was evaluated to a different one. This paper focuses on the latter. After discussing the conceptual and technical issues, we apply the ideas we discuss to the evaluation of PROGRESA, a large welfare program in Mexico for which a randomized evaluation sample is available and has been extensively studied.

Introduction

When resources are scarce, the most valuable information for policymakers is knowing the marginal effect of allocating a budgetary unit to a specific program or activity. Unfortunately, this kind of information is rarely available, especially in the case of social policy interventions. Therefore resources for welfare programs are not usually allocated based on evidence about their impact or on comparisons with other possible interventions.

In recent years, the evaluation of social programs has received considerable attention. Some programs have started to generate information that can be used to assess their impact, and this is certainly a necessary ingredient for improving policymaking (Morley and Coady 2003). However, the evaluation of welfare programs, and more generally of interventions by government or international organizations, is often posed as a one-off question, in that evaluators ask whether a specific intervention has achieved a specific objective in a specific situation. Many evaluations are silent about what the effects of an even slightly different program would be or what the effects of the same program applied to a different context would be.

Recently the more general question of whether results from given studies can be used to predict the effect of different interventions, possibly in different contexts, has received a considerable amount of attention. The ability to extrapolate success stories and avoid failures in different situations would obviously be highly desirable. Unfortunately, a rigorous and successful extrapolation is extremely difficult, though perhaps not unexpected given the problems that are often encountered in assessing the effects of social programs in nonexperimental settings.

This paper discusses the issues involved in evaluating social interventions and in attempts to scale them up. In particular, it discusses the relative merits of (a) nonparametric evaluation strategies that rely on possibly experimental exogenous variation to estimate the impact effects and (b) more structural approaches. The difference between the two approaches is particularly relevant in relation to extrapolation and scaling up. Two types of extrapolation and scaling up could be considered. First, one might want to predict the effects of a program that is different from the one that was evaluated or the effects of changing some aspects of the evaluated program. Second, one might want to predict the effects of exporting an existing program from a population where its effects were evaluated (the evaluation population) to a different population (the implementation population). In what follows we focus on the latter problem and discuss the former only in passing.

After considering the conceptual and technical issues involved with these types of exercises, we apply the ideas to the results of the evaluation of PROGRESA/Oportunidades,[1] a large welfare program in Mexico for which a randomized evaluation sample is available and has been extensively studied. In particular, we divide the seven Mexican states in which the evaluation was carried out into two groups and check the extent to which the results in one group can be extrapolated to the other. The advantage of such a strategy is that one can compare the extrapolation results with the actual ex post evaluation using either a simple comparison of means

or structural methods. The extrapolation here is based on a structural model. Rather than building a new model, we use the one recently proposed by Attanasio, Meghir, and Santiago (2001).

The case of PROGRESA/Oportunidades is especially relevant for our analysis. The program started in 1997 in rural areas and was expanded into urban areas during 2001 and 2002 under the same scheme, with the same benefit structure and levels, even though the characteristics of the urban poor and of the urban environment are very different from those in rural areas. Unfortunately, no evidence on the potential impact on the new areas was available so that decisionmakers could make a more informed decision about whether the program had to be modified in some way or whether complementary interventions were required. An impact evaluation of the urban component of PROGRESA/Oportunidades is currently underway, and the results will be available about two years after the expansion was implemented.

The rest of the paper is organized as follows. The next section discusses the conceptual and technical issues related to the scaling up of evaluations. This is followed by a description of the proposed structural model and the essence of the exercise performed on the Mexican data. The next section describes some details of the program and of the evaluation sample, followed by the results of the estimation and of the main simulation exercise. The final section concludes.

Issues in Scaling up

The basic discussion that follows considers a context where the impact of a policy varies by observable, and possibly by unobservable, characteristics. If the impact is constant, many of the issues discussed here become much simpler. When considering the possibility of applying the findings from one evaluation to a different area, we need to take a number of factors into account. These factors have to do with the way that the areas differ as well as with the way that the evaluation was conducted, namely, the way it was designed and the extent to which it has features that make it generalizable.

In terms of the differences across areas we need to consider the following factors: (a) the distribution of observable characteristics may be different—for example, the area where the program was evaluated may be wealthier or may contain more educated individuals than the area in which we wish to implement the policy; (b) the preferences and unobserved characteristics may differ; and (c) the institutions may be different in the two areas—for example, laws governing child labor may differ or existing laws may be implemented with greater vigor in one area than the other.

The design of the original evaluation will define what can be learned about other setups. Randomized experiments, natural experiments, and matching-based evaluations may all identify different parameters that are more or less relevant for other contexts. In what follows we discuss each of these issues. We argue that in practice, the best chance that a reliable scaling up can be obtained is if a reliable and well-identified structural model is available. Transferring results from one setting to

another without reference to any theoretical context is hard. However, we will also argue that, at least conceptually and with enough data, one could obtain many of the evaluation results from a structural model using a nonparametric approach and many randomized experiments. The main problem with this strategy is that such a wealth of data is typically impossible to obtain.

To inform the discussion and establish notation, consider the following simple problem. Suppose we are considering the impact of a given policy on an outcome variable Y. Such a policy could be a conditional cash transfer, such as the ones distributed by PROGRESA/Oportunidades, and the outcome variable could be the probability that a child enrolls in school. Suppose that an individual i has outcomes Y_i^1 under the policy and the same individual has outcomes Y_i^0 if he or she is not exposed to the effects of the policy. An evaluation will at best estimate some aspects of the distribution of the gains $Y_i^1 - Y_i^0$. What precise aspects of this distribution will be identified will depend on the structure of the evaluation. We also introduce the characteristics X_i and an assignment rule for the policy $D_i = 1$ or $D_i = 0$. Given these, examples of parameters that are the objects of evaluation are the average treatment effect (ATE), $E(Y_i^1 - Y_i^0)$, the impact of treatment on the treated $E(Y_i^1 - Y_i^0 | D_i = 1)$, and versions of these conditional on characteristics X (see Heckman, Lalonde, and Smith 1999).

Differences in the Distribution of Observable Characteristics

Suppose that a program has been evaluated in a particular area or country, for example, by a randomized experiment, and ATE has been estimated. Such randomization permits estimating the counterfactual conditional and unconditional distributions of the outcome variable Y. We can therefore directly compute $E(Y_i^1 - Y_i^0)$ or $E(Y_i^1 - Y_i^0 | X_i)$. Now suppose that we can characterize individuals by a vector of outcome-relevant characteristics X_i whether they are observable or unobservable. In this case ATE can be written as

$$ATE = \int E[Y_i^1 - Y_i^0 | X_i] \, dF(X_i),$$

where the expectation is taken over some random noise that is assumed to be identical across all areas. Central to the argument is the distribution of characteristics in the evaluation area, $F(X_i)$. If $F(X_i)$ differs from the distribution of characteristics in the area where we now want to implement the policy, and if the impacts vary with X_i, then the ATE parameter we have estimated from the evaluation has little to say about the impact of the program in the new area. Two main issues are pertinent here. The first is whether the distribution of characteristics in the evaluation area and the new area have common support (that is, whether all the types of people characterized by X that can be found in the evaluation area can also be found in the implementation area). If the implementation area includes individuals who do not exist in the evaluation area, the evaluation cannot say much about the impact of the policy on these people, except through some form of extrapolation based on parametric assumptions. In other words we can only

hope to predict the impact if we have some form of credible model capable of extrapolating in different circumstances. In any case, such results will nearly always be steeped in controversy.

A second, less serious, problem is one that arises when the distribution of characteristics is different over the common support. In this case, knowledge of the *ATE* in the evaluation area is insufficient to estimate the impact in the implementation area. Here we need to know $E[Y_i^1 - Y_i^0 | X_i]$. An implementation area *ATE* can then be estimated by averaging, using the distribution of characteristics in the implementation area (over the common support). While this is easy to say, these conditional *ATE*s are frequently not well estimated because *X* can be multidimensional and because the evaluation sample sizes may not be large enough. Again, some form of parametric assumptions may prove useful here.

Differences in the Distribution of Unobserved Characteristics

As noted earlier, if the distribution of characteristics varies and if this is relevant for the impact, the outcomes must be reweighed to match the distribution in the implementation area. However, this is not directly possible when some of the characteristics are unobservable. In this case, knowledge of $E[Y_i^1 - Y_i^0 | X_i]$, referred to as $ATE(X_i)$, for observable X_i in the evaluation areas is insufficient for knowledge of $ATE(X_i)$ in the implementation area because

$$ATE(X_i) = \int E[Y_i^1 - Y_i^0 | X_i, u_i] dF(u_i | X_i),$$

and hence the result depends on the distribution of $F(u_i | X)$, which could be area specific. Overcoming this problem is, of course, difficult. A parametric structural approach may allow us to identify $F(u_i | X_i)$ and $E[Y_i^1 - Y_i^0 | X_i, u_i]$ separately in the evaluation area; however, we may still have no way to identify the distribution of unobservables in the implementation area.

Differences in Institutions and Aggregate Conditions

Technically the problem of differences in institutions and aggregate conditions is similar to the one where the distribution of unobservables is not the same. We may be able to identify what we believe are key institutional differences and inform policy on the basis of judgment. However, when the evaluation and implementation areas are quite different, this is a key problem for scaling up. No obvious formal solution can be found unless (a) we can characterize the institutional differences by means of a small set of observable, aggregate variables and (b) we also possess evaluations over a large set of areas with sufficient variability in these characteristics to be able to identify similar environments to the one where we now wish to implement the policy. This is a similar problem to the support problem mentioned earlier.

In addition to institutional differences, we need to consider differences in aggregate macroeconomic conditions. Wages and labor market opportunities in general

may affect the outcomes of a policy. Moreover, differences might exist in infrastructure, services availability, and geography (which could complicate logistics in relation to the delivery of services). Consequently, implementation in a context with different macroeconomic and environmental conditions may give rise to different impacts. This concerns both the transfer of the policy from one area to another as well as from one time period to another.

Differences in the Precise Nature and Intensity of the Policy

Given obvious environmental differences, implementing an identical policy in a new area is unadvisable. However, in the absence of instruments to extrapolate the results of a given evaluation, replicating a successful program may be a tempting and cheap option, for example, the extension of PROGRESA/Oportunidades to urban areas mentioned earlier (and to many other countries around the world). But such a practice neglects some important issues. If the policy under consideration concerns, say, a school subsidy, different levels of the subsidy may be envisaged as well as different ways of means testing. Or policymakers might want to combine a certain intervention with other interventions that strengthen the school infrastructure in situations in which this is particularly weak. Indeed, the evaluation itself may provide hints about how the design could be improved to increase the impacts or to reduce costs. However, the evaluation will usually yield only a specific result relating to the particular rules and intensity envisaged in that case. To go beyond the confines of the particular program evaluated we need to combine the evaluation data with some kind of structural parametric model that will allow us, subject to assumptions, to glean information from other sources that may be indirectly informative about the impact of the policy.

If one does not want to rely on the assumptions necessary to estimate and implement a structural model to predict the likely effects of changes in a program, one needs exogenous variation in all the dimensions of interest. For instance, in the case of PROGRESA/Oportunidades, in addition to evaluating the effect of a particular grant on the enrollment in school of children of a certain age, one might be interested in evaluating the effect of changing the level of the grant for different groups of children. If the localities in the PROGRESA/Oportunidades sample had been randomly assigned to different versions of the program, one could have evaluated in a fully nonparametric way the effect of changes in the program. Similarly, one could estimate the effects of different changes in the program by means of an appropriate randomization scheme. The difficulty is that one is unlikely to have data that would allow the evaluation of any aspect of interest. The use of a carefully constructed structural model is one way to make parsimonious use of any exogenous variation available in the data to extrapolate in different dimensions. We come back to this issue later, but at this point we should stress that, at least conceptually, one could use appropriately constructed experimental data to estimate any elasticity that one would infer from a structural model; however, this strategy is limited by the availability of data.

The Nature of the Evaluation

The precise nature of the original evaluation and which parameter has been identified is a central issue. Frequently evaluations are designed to estimate the parameter treatment on the treated, that is, $E(Y_i^1 - Y_i^0 \mid D_i = 1)$, or the local average treatment effect (*LATE*), rather than the average treatment effect that we have been assuming to this point. The former depends on the precise assignment rule to the policy, while the latter depends on both the assignment rule and precisely which policy is being considered. Attempting to learn about implementation in the new environment from such parameters can be extremely hard, because we would have to model the way in which the assignment rules differ. Thus scaling up is more likely to be successful with an evaluation design that allows the estimation of average treatment effects, subject to the provisos mentioned earlier.

The Ethics and Political Economy of Evaluations

As already noted, a major obstacle to the exclusive use of nonparametric methods in the evaluation of policy interventions is the availability of experimental data in which the assignment of individuals (or localities) to programs (and possibly to different kinds of programs) is random. Randomized data sets collected explicitly for evaluation purposes are few and far between, and politicians and administrators typically strongly resist proposals to construct such data sets. Many reasons account for this reluctance. Obviously there are some important ethical issues. Excluding individuals from programs that researchers believe to be effective in some important dimensions is clearly problematic, and such difficulties tend to be amplified in the political process. Additional problems arise from the short time horizon that seems to be relevant for many politicians.

Experience has shown, however, that some dimensions can be used to overcome such resistance. Typically large programs take time to reach full coverage. The expansion phase can then be used so that instead of randomizing in terms of who gets the program and who does not, one randomizes based on the expansion timetable so that some individuals or communities are randomly assigned to the beginning of the program while others are placed at the end of the line. The PROGRESA/Oportunidades evaluation constitutes a good example of this strategy. The control communities were not excluded from the program forever, but consisted of those communities where, because of budgetary limitations, the program arrived late during the expansion program (that is, roughly two years after the program had first been implemented in the treatment communities included in the evaluation sample).[2] Of course, this strategy leads to a different set of problems: if individuals in the "control" sample know that they are going to get the program and react to this information, this could contaminate the evaluation. Anticipation effects among control individuals or communities are another argument in favor of the use of structural models: one can explicitly introduce the information about the future implementation of a program into the structural model. This is not an issue in completely static problems.

An alternative strategy is to work on pilots that could be used as a basis for designing the details of a specific program. Studies of pilot programs are particularly interesting for at least two reasons. First, from a political point of view, introducing exogenous or random variation in program implementation in a small set of areas or for a small set of individuals might be easier than doing so for the population at large. Second, experimenting with various versions of the actual program might be easier within a pilot and with the explicit purpose of fine-tuning the details of a specific program. The main limitation of pilot studies is likely to be the short time horizon over which they need to be performed.

A final interesting situation is where important budgetary limitations prevent application of the program to all applicants. In this case, random assignment of individuals to the program may be the fairest and most efficient response. This was the case, for instance, in a voucher program in Colombia that Angrist and others (2002) analyzed.

To make the discussion in this section concrete, in what follows we consider the evaluation of a specific welfare program: PROGRESA/Oportunidades in Mexico. As noted, PROGRESA/Oportunidades has been subject to extensive evaluation, and, given the availability of a large and high-quality database, has been studied extensively. The aim of the following sections is not to provide additional evaluation. This has been done using a variety of techniques in many different papers. Rather, in what follows we take PROGRESA/Oportunidades as a specific example to illustrate the various issues discussed (in particular, how, under what conditions, and with what limitations can one use a structural model to evaluate a welfare program and to scale up the results obtained to different situations).

A Structural Model for Evaluation and Scaling Up

The setting we consider is one where the program to be evaluated is a school subsidy program such as PROGRESA/Oportunidades. We assume that a randomized trial has taken place that has split up the relevant population into a treatment group and a control group. Thus one treatment with its entire set of rules is compared with no treatment. Therefore the randomized trial can provide information about the impact of the policy taken as a whole, on average, by subgroup depending on the age of the child or the school grade, and so on. As such, there is little we can learn for any but the most similar settings, that is, those that can be considered as stratified random samples from an identical population. To go further we need to use a model based on assumptions about behavior. This will be identified partly through the randomized experiment and partly through extra assumptions about the validity of certain cross-sectional assumptions. This model will also indirectly highlight the areas where extra-randomized evaluation could usefully inform policy.

We use a simple, dynamic school participation model recently developed by Attanasio, Meghir, and Santiago (2001). Here we provide only one version of the model without discussing or extensively justifying the many assumptions made

along the way. These discussions and alternative specifications of the model can be found in Attanasio, Meghir, and Santiago (2001).

Each child (or his or her parents) decides whether to attend school or to work taking into account the economic incentives involved in such choices. In our model, we assume that children have the option of going to school until they are 17, at which time all formal schooling ends. In the data, almost no children older than 17 are in school. We assume that children who go to school do not work and vice versa. We also assume that children necessarily choose one of these two options. If they decide to work they receive a wage that is specific to their village, education, and age. The model we consider is dynamic for two main reasons. First, because children cannot attend school past the age of 17, this means that going to school now provides the option of completing some grades in the future, that is, a 6-year-old child who wants to complete secondary education has to go to school (and pass the grade) every single year, starting with the current year. This source of dynamics becomes particularly important when we consider the impact of the PROGRESA-Oportunidades grants. Second, we allow for state dependence: the number of years of schooling affects the attractiveness of attending school during this period.

The Basic Framework

The structure of the model is as follows. In each period, going to school involves pecuniary and nonpecuniary costs, in addition to losing the opportunity of working for a wage. The current benefits come from the utility of attending school and possibly, as far as the parents are concerned, by the child care services that the school provides during the working day. The benefits are also assumed to be a function of past attendance. The costs of attending school include the costs of buying books and other school supplies, as well as clothing such as shoes. Transport costs are also involved if the village does not have a secondary school. For households that are entitled to participate in PROGRESA/Oportunidades and live in a treatment village, going to school involves receiving the grade- and gender-specific grant.

Because we are currently using a single cross-section, we use the notation t to signify the age of the child in the year of the survey. Variables with a subscript t may vary with age. We denote the utility of attending school for individual i in period t who has already attended ed_{it} years as

$$u_{it}^s = \mu_i + a'z_{it} + b\,ed_{it} + 1(p_{it} = 1)\beta^p x_{it}^p + 1(s_{it} = 1)\beta^s x_{it}^s + \varepsilon_{it},$$

where z_{it} relates to a number of taste shifter variables, including parental background and age. The variable $1(p_{it} = 1)$ denotes attendance at primary school, while the variable $1(s_{it} = 1)$ denotes attendance at secondary school. x_{it}^p and x_{it}^s represent factors affecting the costs of attending primary school and secondary school, respectively. These factors may interact with other characteristics, such as age or parental education. The term ε_{it} represents a logistic error term that is assumed independently and is identically distributed over time and individuals. Note that the presence

of ed_{it} introduces an important element of dynamics, which we discuss later. Finally, the term μ_i represents unobservables that we assume have a constant impact over time. We assume that μ_i is a discrete random variable whose points of support and probability distribution we estimate.

The utility of not attending school is denoted by

$$u_{it}^w = \vartheta_i w_{it},$$

where w_{it} represents potential earnings when out of school. The wage is a function (estimated from data) of age and educational attainment and village of residence. ϑ_i is a random variable representing heterogeneity in the sensitivity of child i's decision to attend school to earning opportunities. When we consider this additional form of heterogeneity, we assume that ϑ_i is a discrete random variable whose point of support and probability distribution we estimate along with those of μ_i. These are the unobserved characteristics that may or may not have the same distribution in the new implementation area.

The PROGRESA/Oportunidades grant can easily be added to this framework. Let $g(ed_{it}, z_i^p, s)$ denote the grant a child in grade ed_{it} receives if the child is a beneficiary ($z_i^p = 1$) and goes to school ($s = 1$). Then the utility of going to school will be

$$u_{it}^s = \mu_i + a'z_{it} + b\,ed_{it} + 1(p_{it} = 1)\beta^p x_{it}^p + 1(s_{it} = 1)\beta^s x_{it}^s + \theta\,g(ed_{it}, z_i^p, s) + \varepsilon_{it},$$

where the parameter θ reflects the effect that the grant has on the relative choice between school and work. The relative size of θ and the coefficient associated to the wage is of some interest. A model with completely selfish parents would predict a coefficient on the grant of the same size as the one on the wage. In the absence of the exogenous variation in the availability of the grant induced by the randomization, one would be forced to estimate the effect of the program through the coefficient on the wage. This is, for instance, the strategy Todd and Wolpin (2003) follow, and would be the only possible alternative if one wanted to estimate the effect of the program before its implementation. The availability of the randomization allows us to estimate a richer structural model, in that it permits differences between the effect of the grant and that of the wage.

After age 17, we assume that individuals work and earn wages that depend on their level of education. The number of choices open to the individual after school include working in the village, migrating to the closest town, or even migrating to another state. Because we do not have data that would allow us to model these choices (and schooling as a function of these), we model the terminal value function simply as a quadratic function of years of schooling, with the parameters to be estimated alongside the other parameters of the model.[3]

Because the problem is not separable over time, schooling choice involves comparing the costs of schooling now with its future and current benefits. The latter are intangible preferences for attending school, including the potential childcare benefits that parents may enjoy.

Our model has two sources of uncertainty. The first is an iid shock to schooling costs, modeled by the logistic random term ε_{it} Given the structure of the model, having a logistic error in the cost of going to school is equivalent to having two extreme value errors, one in the cost of going to school and one in the utility of work. Even though individuals know ε_{it} in the current period, they do not know its value in the future. Because future costs will affect future schooling choices, they indirectly affect current choices. Note that the term μ_i, while known (and constant) for the individual, is unobserved by the econometrician.

The second source of uncertainty originates because a child may not be successful in completing a grade. If that is the case, we assume that the level of education does not increase. We assume that the probability of failing to complete a grade is exogenous and does not depend on effort or on the willingness to continue schooling. However, we allow this probability to vary with the grade in question and with the age of the individual, and we assume that it is known to the individual.[4] We estimate the probability of failure for each grade as the ratio of individuals who are in the same grade as the year before at a particular age. As we know the completed grade for those not attending school, we include this in the calculation. This may be important, because failure may discourage school attendance. We denote by $I \in \{0,1\}$ the random increment to the grade that results from attending school at present. If successful, then $I = 1$, otherwise $I = 0$. We denote the probability of success at age t for grade ed as $p_t^s(ed_{it})$.

Thus the value of attending school for someone who has completed successful ed_i years in school, is of age t already, and has characteristics z_{it} is

$$V_i^s(ed_i, t \mid z_{it}) = u_{it}^s + \beta\{p_t^s(ed_i + 1)\, E\max\left[V_i^s(ed_i + 1, t + 1),\, V_i^w(ed_i + 1, t + 1)\right]$$
$$+\, (1 - p_t^s(ed_i + 1))\, E\max\left[V_i^s(ed_i, t + 1),\, V_i^w(ed_i, t + 1)\right]\},$$

where the expectation is taken over the possible outcomes of the random shock ε_{it}. The value of working is similarly written as

$$V_i^w(ed_i, t \mid z_{it}) = u_{it}^w + \beta E\max\left\{V_i^s(ed_i, t + 1), V_i^w(ed_i, t + 1)\right\}.$$

The difference between the first terms of the two equations reflects the current costs of attending, while the difference between the second two terms reflects the future benefits and costs of schooling. Finally, the parameter β represents the discount factor. In practice, because we do not model savings and borrowing explicitly, this will reflect liquidity constraints or other factors that lead households more or less to disregard the future.

Estimation

$$P(Attend_{it} = 1 \mid z_{it}, x_{it}^p, x_{it}^s, ed_{it}, wage) = F\{u_{it}^s - u_{it}^w\, \beta[E\max\{V_i^s(ed_i + I, t + 1),$$
$$V_i^w(ed_i + I, t + 1)\} - E\max\{V_i^s(ed_i, t + 1),$$
$$V_i^w(ed_i, t + 1)\}]\}$$

In terms of estimation, the problem in the absence of unobserved heterogeneity $(\mu_i \equiv \mu, \forall\, i)$ other than through the iid shock ε_{it}, is relatively simple. The likelihood function is based on the probability of attending school that takes the form

$$(1) \quad P(Attend_{it} = 1 \mid z_{it}, x_{it}^p, x_{it}^s, ed_{it}, wage)$$

$$= F\{u_{it}^s - u_{it}^w + \beta[E \max\{V_i^s\,(ed_i + I, t + 1), V_i^w(ed_i + I, t + 1)\}$$

$$- E \max\{V_i^s\,(ed_i, t + 1), V_i^w(ed_i, t + 1)\}]\}$$

where the expectation is taken over both ε and I where relevant.

The difference between the current values of going to school and working will reflect both the pecuniary trade-offs (the effect of the wage and the cost of going to school) and other relevant factors, such as the disutility of work, and possibly the utility of going to school. The most general version of our model allows these effects to be heterogeneous across individuals through the terms μ_i and ϑ_i. The difference in square brackets reflects the difference between the future value function implied by the current choice.

Assuming that the unobserved preference shock ε_{it} is logistic, when the discount factor (β) is zero, our model collapses to a simple logit model. With a positive discount factor instead, the model needs to be solved at each iteration to compute the future value functions V_{it+1}^s and V_{it+1}^w. In our case these computations are relatively simple, because the expected value of the value functions can be computed analytically given our distributional assumption. Given assumptions on the terminal value function for each final grade, the expressions in equation (1) can be computed by backward recursion.

In the presence of unobserved heterogeneity, we assume that the constant μ_i (and possibly ϑ_i) is a discrete random variable, distributed independently of all characteristics $z_{it}, x_{it}^p, x_{it}^s$, and the $wage_{it}$.[5] However, given the structure of our model and the fact that we use a single cross-section, we have an important initial conditions problem because we do not observe the entire history of schooling for the children in the sample. That is, we cannot assume that the random variable μ_i (and ϑ_i) is independent of past school decisions as reflected in the current level of schooling ed_{it}.

To solve this problem we specify a reduced form for educational attainment up to the current date. We assume that conditional on unobserved heterogeneity κ_i, the level of schooling achieved up to now follows a Poisson distribution with mean $\exp(h_i'\zeta + \kappa_i)$, where h_i includes variables reflecting past schooling costs, such as the availability of secondary schools in pre-experimental years. The probabilities of the stock of schooling and of attending school in this period are conditionally independent (given $z_{it}, x_{it}^p, x_{it}^s, h_i, wage_{it}$, and the unobservables $\mu_i, \vartheta_i, \kappa_i$). Hence we can write the probability of $ed_{it} = e$ and of child i attending school as

$$P(ed_{it} = e, Attend_{it} = 1 \mid z_{it}, x_{it}^p, x_{it}^s, h_i, wage_{it}, \mu_i, \vartheta_i, \kappa_i) =$$

$$P(Attend_{it} = 1 \mid z_{it}, x_{it}^p, x_{it}^s, wage_{it}, ed_{it}, \vartheta_i, \mu_i)$$

$$P(ed_{it} = e \mid z_{it}, x_{it}^p, x_{it}^s, h_i, wage, \kappa_i).$$

The endogeneity of the stock of schooling is captured by the potential dependence of ϑ_i, μ_i, and κ_i. Thus we assume that we can model this joint distribution as

$$F(\mu_i = m, \vartheta_i = s, \kappa_i = k) = p_{msk}$$

for $m \in M$, $s \in S$, and $k \in K$, where M, S, and K are the set of points of support for μ, ϑ, and κ. Hence for an individual with observable characteristics z_{it}, x_{it}^p, x_{it}^s, h_i, *wage*, the observed probability of attending and having reached a level of schooling e is

$$P(ed_{it} = e, Attend_{it} = 1 \,|\, z_{it}, x_{it}^p, x_{it}^s, h_i, wage_{it})$$

$$\sum_{m \in M} \sum_{s \in S} \sum_{k \in K} p_{msk} \{P(Attend_{it} = 1 \,|\, z_{it}, x_{it}^p, x_{it}^s, wage_{it}, ed_{it}, \vartheta_i = s, \mu_i = m)$$

$$P(ed_{it} = e \,|\, z_{it}, x_{it}^p, x_{it}^s, h_i, wage, \kappa_i = k)\}.$$

The number of points of support, as well as the values that m, s, and k can take and the probabilities at these points, can be estimated.

Using the Model for Addressing Scaling Up Issues

The foregoing model is one of many possibilities. We chose it because it incorporates some key issues in educational choice, namely, the trade-offs between costs and benefits that are likely to characterize behavior in a broad set of circumstances. It also incorporates the intertemporal trade-offs that are central to PROGRESA/ Oportunidades. The model provides a way to evaluate the impact of changing the program's parameters, such as the amount offered and the way the amounts vary by age. Because of the forward-looking nature of the model, it also allows us to distinguish the impact of the program compared with a no program state, as opposed to the randomized experiment, which provides the impact of having the program compared with expecting to receive it in 18 months time.

However, all these advantages do not come for free. They come because of a number of assumptions that we have made that allow us to identify a rich behavioral model, combining the randomized experiment with further cross-sectional variation. In particular, over and above the exogenous variation that is induced by the experimental design, we also use the structure of the grant across different age groups to identify how the impact changes with the amount of the grant, and also how individuals react to the promise of a future payment. Obviously a richer evaluation framework that would have generated variation in the amounts, and possibly in the age structure, would have led to a model that would be identified using fewer assumptions.

Given the model, we can immediately deal with the issue of the distribution of observable characteristics by applying the model to a random sample drawn from the implementation area. This will ensure that when we aggregate the impact, we have applied the right weights. Of course, this process is also possible without a structural model. In terms of dealing with issues relating to the support, this is just a matter of comparing the support of the Xs in the evaluation data with the support in the implementation data. This can be done by using the device of the propensity

score, which is defined here as the probability of being in the implementation data, given the *X*s. If there is a lack of common support, one then has to decide if extrapolation is to be used based on the parametric model.

The key difficulty, however, is the treatment of unobserved heterogeneity. In the estimation section we showed how one can estimate the distribution of unobserved heterogeneity in a parametric setting. The distribution of unobservables can have a large effect on the impact of the policy, but we know nothing about this distribution in the implementation area. In practice, the only choice is to assume that the distribution does not vary across these areas, which is, of course, unsatisfactory and can be a source of errors in the ex ante evaluation.

The model allows the examination of different structures of the program, such as variation in the age rules and in the amounts. It also allows us to take differing aggregate conditions into account, so long as these vary within the evaluation areas. This is not the case for PROGRESA/Oportunidades to any important degree, except perhaps as far as some variation in the wage is concerned. Finally, the model can technically predict both the average treatment effect and the treatment on the treated. Note that the randomized experiment permitted the identification of *ATE* without further assumptions that would be required to extract *ATE* from treatment on the treated.

The PROGRESA/Oportunidades Program and Its Evaluation Sample

This section starts by describing the main features of the PROGRESA-Oportunidades program. It then turns to a discussion of the evaluation sample and some of the results that have been obtained from its analysis. Finally, it splits the sample into two parts and describes some of the features of each of the two groups.

The PROGRESA/Oportunidades Program

In 1997, the Mexican government started a large program, known eventually as PROGRESA/Oportunidades, to reduce poverty in rural areas. The program was innovative in that introduced a number of incentives plus conditions with which participant households had to comply to keep receiving the program's benefits.

PROGRESA/Oportunidades focuses on health, nutrition, and education. The health component consists of a number of initiatives aimed at improving information about vaccination, nutrition, contraception, and hygiene and of a program of visits to health centers for children and women. Participation in the health component is a prerequisite for participating in the nutrition component, which in addition to a basic monetary subsidy received by all beneficiary households, gives some in-kind transfers to households with extremely young infants and pregnant women. The largest component of the program is the education one. Beneficiary households with school-age children receive grants conditional on school attendance. The size of the grant increases with the grade, and for secondary education is slightly higher for girls than for boys. In addition to the bimonthly payments, beneficiaries with school-age children receive a small annual grant for school supplies. Finally, all the transfers are provided to the mother in the household.

The program initially targeted the poorest communities in rural Mexico. Roughly speaking, the two criteria communities had to satisfy to qualify for the program were a certain degree of poverty (as measured by a so-called index of marginalization, basically the first principal component of a certain number of village-level variables routinely collected by the government) and access to certain basic structures (schools and health centers). The reason for the second criterion was the conditional nature of the program: without some basic structures within a certain distance, beneficiary households could not comply with the basic conditions for retaining their beneficiary status (participation in vaccination and check-up visits for the health and nutrition components and school attendance for the education component).

Once a locality qualified, individual households could or could not be eligible for the program, depending on a single indicator, once again the first principal component of a number of variables (such as income, house type, presence of running water, and so on). Eligibility was determined in two steps. First, a general census of the PROGRESA-Oportunidades localities measured the variables needed to compute the indicator, and each household was defined as poor or not poor (where poor was equivalent to eligibility). Subsequently, in March 1998, an additional survey was carried out and some households were added to the list of beneficiaries.

For logistic and budgetary reasons, the program was phased in slowly but is currently extremely large. In 1998 it was started in fewer than 10,000 localities, but by the end of 1999 it had been implemented in more than 50,000 localities and covered about 2.6 million households, or 40 percent of all rural families. The program now has a budget of about US$1 billion and is by far the largest welfare program in Mexico. It is the first program of its kind to survive a change of administration. The Fox administration decided not only to continue it, but also to expand it to poor urban areas. The program has received a considerable amount of attention and publicity, and similar programs are currently being implemented in Argentina, Colombia, Honduras, Nicaragua, and Turkey (for a detailed evaluation of various aspects of the program see the International Food Policy Research Institute web page www.ifpri.org/themes/progresa.htm).

The program represents substantial help for its beneficiaries. The nutritional component was Mex$100 (US$10) per month in the second half of 1998, which corresponded to 8 percent of the income of beneficiaries in the evaluation sample.

Table 1 reports the nutrition support and details of the education grant. As noted earlier, the grants are conditional on children's enrollment in school and attendance, and can be cumulated within a household up to a maximum of Mex$625 (US$62.5) per month per household, or 52 percent of the average beneficiary's income. The average grant per household in the sample we use was Mex$348 per month for households with children and 250 for all beneficiaries, or 21 percent of the beneficiaries' income. To keep the grant, children have to attend at least 85 percent of classes. If they fail a grade children are still entitled to the grant for the same grade, but if they fail the grade a second time, they lose eligibility.

TABLE 1.
PROGRESA/Oportunidades, Bimonthly Monetary Benefits, 1998 and 1999

(current Mex$)

Type of benefit	January–June 1998	July–December 1998	January–June 1999	July–December 1999
Nutrition support	190	200	230	250
Primary school				
Third grade	130	140	150	160
Fourth grade	150	160	180	190
Fifth grade	190	200	230	250
Sixth grade	260	270	300	330
First year of secondary school				
Boys	380	400	440	480
Girls	400	410	470	500
Second year of secondary school				
Boys	400	400	470	500
Girls	440	470	520	560
Third year of secondary school				
Boys	420	440	490	530
Girls	480	510	570	610
Maximum support	1,170	1,250	1,390	1,500

Note: US$1 is approximately equivalent to Mex$10.

Source: PROGRESA/Oportunidades.

The Evaluation Sample

Before starting the program, the agency running it decided to start collecting a large data set to evaluate its effectiveness. Among the beneficiary localities, 506 were chosen randomly and included in the evaluation sample. The 1997 survey was supplemented in March 1998 by a richer survey in these villages, which were located in 7 of Mexico's 31 states. All households in these villages were interviewed, for a total of roughly 25,000 households. Using the information from the 1997 and March 1998 surveys, each household can be classified as poor or non-poor, that is, each household can be identified as being entitled or not entitled to the program.

One of the most interesting aspects of the evaluation sample is that it contains a randomization component. The agency running PROGRESA/Oportunidades used the fact that, for logistical reasons, the program could not be started everywhere simultaneously to randomly allocate the villages in the evaluation sample to treatment and control groups. In particular, 320 randomly chosen villages of the evaluation sample were assigned to communities where the program started early, that is in May 1998. The remaining 186 villages were assigned to communities where the program started almost two years later, in December 1999.

An extensive survey of the evaluation sample was carried out. After the initial data collection between the end of 1997 and the beginning of 1998, an additional four instruments were collected in November 1998, March 1999, November 1999, and April 2000. Within each village in the evaluation sample, the survey covers all households and collects extensive information about consumption, income, transfers, and a variety of other issues. For each household member, including each child, there is information about age, gender, education, current labor supply, earnings, school enrollment, and health status. The household survey is supplemented by a locality questionnaire that provides information about prices of various commodities, average agricultural wages (for both males and females), institutions present in the village, and distance of the village from the closest primary and secondary school (in kilometers and minutes).

The evaluation sample has been studied extensively. In addition to several reports by the International Food Policy Research Institute, efficiently summarized by Skoufias (2001), several papers have looked at various outcomes, including Behrman and Todd (2000), Gertler (2000), and Schulz (2000). The Behrman and Todd paper is particularly important, because it looks at differences between treatment and control localities in preprogram variables. By and large, the randomization was successful in that, with a few exceptions, there are no apparent differences between treatment and control villages.[6]

Some of the most important results from the evaluation are the following. The nutritional component of the program has increased height and weight by 16 percent for children aged 12 to 36 months. The health services plus the nutritional benefits have resulted in a 12 percent decline in diseases among children and a 19 percent reduction in the number of days that adults are unable to attend work because of health reasons. The schooling incentives have generated an increase in primary school completion rates from 65 to 74 percent, a reduction of 24 percent in secondary school dropouts, and an estimated 38 percent increase in enrollment in secondary school.

Two Groups of States and Their Features

To illustrate the issues involved with scaling up, we divide the PROGRESA/ Oportunidades evaluation sample into two parts, conduct the evaluation in one and use the other to perform an ex ante evaluation of the program. We can then compare the results of these evaluations with the results obtained in the ex post evaluation. The seven states included in the evaluation sample were divided into two groups. The first group of states are the four poorest: Guerrero, Puebla, Veracruz, and Hidalgo. The second group consists of slightly better off states: Michoacan, Queretaro, and San Luis Potosi. Table 2 presents sample means and standard deviations for a number of variables that are likely determinants of the outcome of interest, as well as preprogram outcomes. Among the first group of variables we include child-level variables (completed years of schooling), household-level variables (income, mother's education, ethnicity), and community-level variables (agricultural wage, presence of asecondary school, and distance from the closest secondary school). All these

TABLE 2.
Differences between the Two Groups of States

Variable	Group 1[a] Mean	Group 1[a] Standard deviation	Group 2[b] Mean	Group 2[b] Standard deviation
Preprogram enrollment	0.748	0.419	0.856	0.370
Completed years of education	5.03	2.74	5.05	2.58
Mother's education				
Less than completion of primary education	0.377	0.485	0.295	0.456
Less than completion of secondary education	0.345	0.475	0.403	0.491
Secondary education or more	0.278	0.448	0.302	0.459
Household Income	637	948	814	1056
Agricultural male wage	25.5	6.6	37.5	11.6
Distance from secondary school (in minutes)	79.12	99.7	55.7	61.8
Percentage of population that is indigenous	0.349	0.477	0.038	0.192
Percentage of program beneficiaries	0.859	0.348	0.801	0.348

Boys older than 5 and younger than 18.

a. Puebla, Guerrero, Veracruz, and Hidalgo. The number of observations was 16,905.

b. Queretaro, Michoacan, and San Luis Potosi. The number of observations was 3,655.

Source: Authors' calculations.

variables, which are likely to affect the effectiveness of the program, were included among the determinants of schooling choices in the theoretical model.

The differences between the sample localities in the two groups of states are remarkable. As expected, the localities in the first group are considerably poorer than those in the second group. Every single indicator, from household income, to agricultural wage, to the percentage of children belonging to beneficiary families, points in that direction. Particularly remarkable is the percentage of children belonging to indigenous families, which is 35 percent in the first group and less than 4 percent in the second group. Preprogram enrollment is also quite different: in group 1, 75 percent of the boys aged 6 to 18 are in school, while in group 2 the figure is 86 percent.

Of course, with so many dimensions, summarizing the differences between the two groups of states in terms of these conditioning variables is difficult. For such a purpose, we estimate a simple probit model where the probability of a child being in the first or second groups is estimated as a function of the conditioning variables included in table 2 (excluding the outcome variables, such as preprogram schooling). Most of the variables inserted in the regression turned out to be statistically significant, indicating systematic differences in the distribution of these variables in the two groups of states.

For all observations in our sample we can now compute the propensity score (for belonging to the first group), which we use as a summary statistic for the difference in the distribution of the dependent variables between the two groups. Figure 1 plots the distribution of propensity scores in the two groups of states. Not only is the
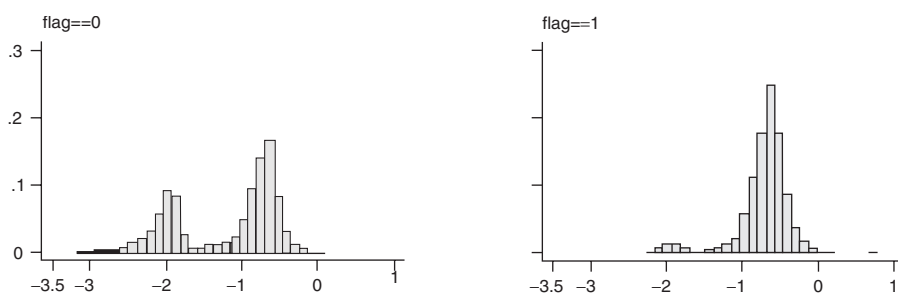
distribution quite different (as could be inferred from the fact that most variables were significant in the probit regression), but we also see that for a sizeable proportion of observations, the support of propensity scores in the two samples does not overlap: 6 percent of the observations in group 1 have a value of the propensity score higher than the highest value in group 2, and 5 percent of the observations in group 2 have a propensity score lower than the lowest level observed in group 1.

For scaling up, this result implies that even if unobserved heterogeneity does not constitute a problem, we can use the results of the evaluation in group 1 for an ex ante evaluation of the program in group 2 states only for 95 percent of the observations. Moreover, 28 percent of the observations in group 1 have a propensity score between −2 and −3, while only 0.3 percent of the observations in group 2 have a propensity score of less than −2. These issues will impose important limits to the scaling up exercise unless, of course, the independent variables we have been analyzing here turn out to be irrelevant for the effectiveness of the program.

The Effect of PROGRESA-Oportunidades in the Two Groups of States Estimated by Difference in Difference

Before presenting the results of the estimation of the structural model, we briefly discuss the effects of the program as estimated by applying a diff in diff estimator in the two groups of states. It is now well documented that most of the effect of PROGRESA-Oportunidades was on slightly older children (see Schulz 2000). For that reason, here and in what follows, we focus on boys age 10 to 18. However, we use a diff. in diff. (difference in difference) estimator that controls for a few determinants of school choice to improve efficiency.[7] In particular, we control for boys' age, mothers' education, distance to secondary schools, and towns' average cost of secondary schools.

The results indicate a remarkable difference in the estimated effect of the program in the two groups of states. In group 1, the effect is estimated at around 2.3 percent (with a standard error, which takes into account cluster effects at the municipality level, of 0.012). In group 2, the effect is estimated at 7.4 percent (with a standard error of 0.017). The difference between the two effects is statistically significant (p-value of 1.3 percent).

Two observations are in order. First, in the presence of anticipation effects in the control localities, the effect of the program is likely to be underestimated. This, following the discussion in Attanasio, Meghir and Santiago (2001), is our maintained

assumption. Second, to the best of our knowledge, this dramatic difference in the effect of PROGRESA-Oportunidades in different states had not been documented before. While the study of these differences is not the main goal of this paper, we stress the difference in the effectiveness of the program in raising enrollment rates between the poorest states and the others.

Estimation and Simulation of a Structural Model

In this section we first estimate the model described earlier in both groups of states. This set of results will allow us to quantify the differences in the distribution of unobserved heterogeneity in the two groups of states. Any difference will prevent a straightforward use of the evaluation in one of the groups of states to scale up the other group. However, in the second part of the section we ignore these differences,

TABLE 3.
Estimates of the Structural Model

Coefficients	Group 1[a] Parameter estimate	Standard error	Group 2[b] Parameter estimate	Standard error
Intercept p1	−26.12824	2.85159	−36.05473	3.90073
Intercept p2	−19.27025	2.31334	−24.36846	2.7138
Wage p1	−0.08644	0.06442	0.17190	0.08117
Wage p2	0.64341	0.12377	0.87642	0.13170
Grant	0.67240	0.11736	0.77147	0.20139
Distance from secondary school	0.09381	0.01325	0.19183	0.02608
Cost of secondary school	0.00793	0.00164	0.00463	0.00176
Age	3.24929	0.34110	3.91940	0.46625
Years of education	−2.55861	0.28772	−2.74321	0.42009
Father: primary school	−0.19454	0.14100	−0.45956	0.22372
Father: incomplete secondary school	−0.42453	0.16610	−0.77902	0.27938
Father: secondary school or more	−0.99106	0.37631	−1.39751	0.65107
Mother: primary school	−0.15970	0.14841	−0.42837	0.22448
Mother: incomplete secondary school	−0.38623	0.17568	−0.70545	0.26758
Mother: secondary school or more	−2.08134	0.53666	−1.08697	0.72461
Non-indigenous	−0.55368	0.14295	−0.95029	0.38256
Distribution of unobserved	0.6396	0.1855	0.8251	0.463	0.5238	0.8701
heterogeneity						
(intercepts: rows	0.1276	0.0473	0.1749	0.1173	0.0126	0.1299
slopes: columns)	0.7672	0.2328	1.0000	0.4637	0.5363	1.0000

Boys older than 9 and younger than 18.

Note: The specification also includes state dummies, dummies for beneficiaries, and dummies for treatment localities. The discount factor is 0.9.

a. Puebla, Guerrero, Veracruz, and Hidalgo. The number of observations was 16,905.

b. Queretaro, Michoacan, and San Luis Potosi. The number of observations was 3,655.

Source: Authors' calculations.

as well as the potential differences in the coefficients on the observable variables, and apply the model estimated in the first group of states to scale up the program in the second group of states. This ex ante evaluation is then compared with the one that can be obtained from the evaluation sample itself.

Estimation Results

Table 3 reports the estimation results for the two groups of states. The specification of the model is identical for the two groups of states. In both cases we performed a grid search on the discount factor and concluded that a value of around 0.9 maximizes the likelihood function. In both cases we assumed that the program is perceived to reach the control communities a year after its implementation in the treatment communities. This assumption is discussed in Attanasio, Meghir and Santiago (2001). Among the three scenarios tried in estimation (the program is never implemented in the control municipalities, the program is implemented after two years, and the program is implemented after one year), the third is the one that yields the largest likelihood function.

The first two rows of the table report the estimated points of support for the discrete random variable μ_i: the intercepts of the value of going to school. All the coefficients are expressed as costs of going to school, so that the probability of going to school decreases with the size of this coefficient. In both sets of states we can identify two groups of children, the first of which is much more likely to go to school. However, in group 2 both points of support are much lower, reflecting the overall larger enrollment of these children. At the bottom of the table we report the estimated probability distribution of the unobserved heterogeneity. Considering the marginal probabilities for the intercepts (reported in the third column), we see that in both groups of states the group of children more likely to go to school accounts for more than 80 percent of the sample. The differences in the level of these coefficients obviously constitute important problems for scaling up.

In rows three and four we report the estimated coefficients on the wage. Here the differences between the two states are also quite evident. Children in group 2 seem to be much more sensitive to wages: both coefficients are positive and significant for group 2 (indicating that an increase in children's wages decreases the probability of attending school), while for group 1, one of the two coefficients is not significantly different from zero. The marginal distribution is also different between the two groups. In group 1, the two points account for 76 and 23 percent of the sample (notice that 76 percent are therefore insensitive to the wage), while in group 2, the split is 46 and 54 percent.

Going down the rows, all coefficients have the expected sign and there are, once again, marked differences between the two groups. These differences constitute another potential problem for extrapolation and scaling up. Not only is the distribution of background variables different between the two groups of states, as documented in the previous section, but the effect of these variables on the outcome of interest seems to be different. Using the results of one evaluation to extrapolate to a different sample is therefore difficult.

Finally, note the difference (this time not too large) in the estimated coefficient of the grant. However, given the differences in all other coefficients and the nonlinearity of the model, relating the size of this particular coefficient to its marginal effect is hard. We now turn to simulations of the model to evaluate these types of effects.

Simulation Results

We now use the structural model to see how it performs in estimating the potential effects of the policy. The counterfactual simulations use the data from our chosen implementation area (group 1) with a number of different combinations of parameters. For all counterfactual simulations we use the distribution of unobserved heterogeneity as estimated in the evaluation area (group 2). Before presenting the counterfactual simulations, however, we report the effects of the program as estimated by the model estimated in the implementation area. This is the kind of exercise that would not be possible to do ex ante and that should be replaced by scaling up.

The effect we report in the first column of table 4 is the one predicted by the model as compared with the case where no policy is expected to be in operation. The effect, estimated at 0.08, is larger than the one predicted by the randomized experiment, because the anticipation effect we estimate for the control group is netted out.

In the next column, we carry out another simulation experiment that in practice is not feasible. We use the estimated parameters from the evaluation area, but use the state dummies estimated in the implementation areas (the latter would not normally be known in a genuine scaling up exercise). This shows how close one could get just by using the observable characteristics if one knew the contribution of the unobserved area effects. The impact we get is about a quarter less than the baseline one. The difference is attributable to unobserved area effects.

In the third, fourth, and fifth columns we present feasible predictions based on the coefficients from the evaluation area, the data from the implementation area, and the state dummy from each of the three states in the evaluation areas used in turn. This should also be accompanied (not done yet) by an analysis that would show which state is likely to be most similar to the implementation area. From the results we see that none of the effects are particularly close to the one estimated using implementation area data. In one case the effect is less than half.

Where does this leave us? At the moment we are still in the awkward position of having to accept that we are not well enough equipped for a scaling up exercise that

TABLE 4.
Effect in Group 1 States

Predicted model effect	Effect with own state dummies	Effect with Queretaro dummy	Effect with San Luis Potosi dummy	Effect with Michoacan dummy
0.080	0.061	0.039	0.058	0.110

Source: Authors' calculations.

is reliable. We know that a strong trade-off exists here between the scope of the original evaluation and the modeling assumptions one is prepared to make. Clearly our model may be far from perfect. It makes strong behavioral assumptions. Moreover, one can argue that it may not have a rich enough specification. However, it does seem to fit the data reasonably well. What we do show in this paper is that knowledge of observed individual characteristics is unlikely to be sufficient in practice to extrapolate the effects of a policy to a different context.

Conclusions

This paper discusses issues to do with scaling up, that is, using knowledge obtained from the evaluation of a policy in one area for predicting the effectiveness of this policy in another. We start from the premise, which is empirically supported, that the effects we are interested in vary substantially both in terms of observed characteristics and unobserved characteristics. Even though one can achieve something close to ideal with an elaborate experimental design, in the near future we are unlikely to have the resources, or even the will, to carry out the requisite complicated experimentation. We can go some of the way if we have extremely large samples for the kinds of experiments available now. This would at least allow us to correct carefully for the differences in the distribution of unobserved characteristics. But even there we are some way off from having enough data to produce reliable predictions, suitably reweighted. Thus the next best thing is to combine a structural model with the data we have. This allows us to fill in the gaps in a theoretically coherent way and offers a framework for redesigning policies. Inevitably this requires assumptions, but at least they are made in a coherent and transparent way. Nevertheless, in our first attempt we show that our ability to predict the actual effects is limited, particularly by the lack of knowledge about aggregate area effects. This points to a need to collect data at the area level, and perhaps to design evaluations in such a way that more variation is indiced at that level and not only at the individual-level data.

Subject to the foregoing caveats, the results provided in this paper have policy relevance. For instance, lower impacts in the relatively poorer areas suggest that to get the same effects, additional policy interventions might be necessary or the benefit structure might have to change. For instance, it might be necessary to improve school quality (to make it more adequate for indigenous children), to improve parents' education through literacy programs, or even to expand the supply of education services so that beneficiaries in poorer areas have to spend less time getting to school. An important item on the research agenda should be modeling differences in impact evaluation and linking them to observable variables.

Notes

1. The program was originally called PROGRESA from the Spanish acronym for Program for Health, Education, and Nutrition. When it was expanded after 2001, the program was renamed PROGRESA/Oportunidades.

2. Another distinguishing feature of the PROGRESA evaluation was that the randomization was done at the locality level rather than at the individual level. In that particular situation, this procedure was preferable for at least two reasons. First, by randomizing at the community level one can, in principle, estimate spillover and general equilibrium effects induced by the program. Second, from a political point of view, the random exclusion of a number of individuals in small communities where the program was operating might have been harder to sustain.

3. We have used some information about urban and rural returns to education at the state level along with some information about migration in each state to try to model such a relationship. Unfortunately, we have no information about migration patterns, and the data on the returns to education are extremely noisy. This situation has motivated our choice of estimating the returns to education that best fit our education choices.

4. Because we estimate this probability from the data, we could also allow for dependence on other characteristics.

5. In practice, dependence with the wage rate can be allowed for. However, the wage data are not rich enough to estimate a joint model of school participation and wages.

6. Of course, one would expect 5 percent of rejections at the 5 percent level. Unfortunately, one of the few preprogram variables that turned out to be statistically different between treatment and control villages was school enrollment. This difference, whose origin is not clear, has motivated the use of diff. in diff. estimators. In the structural model we proxy for it with a treatment dummy.

7. As the diff. in diff. estimator explicitly uses the randomization of the program between treatment and control localities, which is by construction uncorrelated with all independent variables, the only reason to "control" for such variables is to improve efficiency.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92: 1535–58.

Attanasio, O. P., C. Meghir, and A. Santiago. 2001. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate a Welfare Program." University College London. Processed.

Behrman, J., and P. Todd. 2000. "Randomness in the Experimental Samples of PROGRESA-Oportunidades." International Food Policy Research Institute, Washington, D.C. Processed.

Gertler, P. J. 2000 "Final Report: The Impact of PROGRESA-Oportunidades on Health." International Food Policy Research Institute, Washington, D.C. Processed.

Heckman, J., R. Lalonde, and J. Smith. 1999. "The Evaluation of Labor Market Policies." In O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*. Amsterdam: North Holland.

International Food Policy Research Institute. "PROGRESA—Programa de Educación, Salud y Alimentación." http://www.ifpri.org/themes/progresa.htm.

Morley, S., and D. Coady. 2003. *From Social Assistance to Social Development: A Review of Targeted Education Subsidies in Developing Countries.* Washington International Food Policy Research Institute and Institute for Global Development.

Schulz, T. P. 2000. "School Subsidies for the Poor: Evaluating a Mexican Strategy for Reducing Poverty." International Food Policy Research Institute, Washington, D.C. Processed.

Skoufias, E. 2001. "PROGRESA-Oportunidades and Its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI." International Food Policy Research Institute, Washington, D.C. Processed.

Todd, P., and K. Wolpin. 2003. "Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico." University of Pennsylvania, Philadelphia, Penn. Processed.

Scaling Up and Evaluation

ESTHER DUFLO

This paper discusses the role that impact evaluations should play in scaling up. Credible impact evaluations are needed to ensure that the most effective programs are scaled up at the national or international levels. Scaling up is possible only if a case can be made that programs that have been successful on a small scale would work in other contexts. Therefore the very objective of scaling up implies that learning from experience is possible.

Because programs that have been shown to be successful can be replicated in other countries while unsuccessful programs can be abandoned, impact evaluations are international public goods, thus the international agencies should have a key role in promoting and financing them. In doing this, they would achieve three important objectives: improve the rates of return on the programs they support, improve the rates of return on the programs other policymakers support by providing evidence on the basis of which programs can be selected, and build long-term support for international aid and development by making it possible to credibly signal what programs work and what programs do not work.

The paper argues that considerable scope exists for expanding the use of randomized evaluations. For a broad class of development programs, randomized evaluation can be used to overcome the problems often encountered when using current evaluation practices.

Introduction

Scaling up and evaluation are often presented as conflicting objectives, and for most international development agencies, "going to scale" has to be given priority. The United Nations Children's Fund (UNICEF), for example, lists as its first priority for HIV/AIDS education "moving away from small scale pilot projects" and "expanding effective and promising approaches to national scale."[1] The trade-off is explicit: by moving away from pilots and projects before their impact on behavior leading to HIV/AIDS has been convincingly established, one has to commit to expanding projects that are only promising—not enough projects have been proven to be "effective." The UNICEF web site on skill-based health education reports on 10 case studies of promising, school-based HIV/AIDS education programs, only one of which presents differences in outcomes between a treatment group and a comparison group. These approaches are the programs that UNICEF can recommend be implemented on a national scale.[2]

This paper argues that for international agencies there is no real trade-off between scaling up and evaluation. On the contrary, evaluation can give them an opportunity to leverage the impact of their programs well beyond their ability to finance them. The very idea of scaling up implies that the same programs can work in different environments and that learning from experience is possible. Therefore reliable program evaluations serve several purposes. First, a well-conducted evaluation can offer insights into a particular project. For example, all programs should be subject to process evaluations to ensure that funds are spent as intended and to receive feedback from stakeholders on how programs could be improved. However, while process evaluations are necessary, they are insufficient to determine program impact. A second purpose of rigorous evaluations of programs' impacts is that this information can be shared with others. The benefits of knowing which programs work and which do not extend far beyond any program or agency, and credible impact evaluations are global public goods in the sense that they can offer reliable guidance to international organizations, governments, donors, and nongovernmental organizations (NGOs) in their ongoing search for effective programs. Thus when evaluation is used to find out what works and what does not, the benefits extend far beyond the selection of projects within the organization. While a prospective impact evaluation may require postponing the national expansion of a program for some time, evaluation can be part of the backbone of a much larger expansion: that of the project on a much larger scale (if proven successful), and that of the ability to fund development projects. Providing these international public goods should be one of the important missions of international organizations.

In this paper I argue that for a broad class of development programs, randomized evaluations are a way to obtain credible and transparent estimates of program impact that overcome the problems often encountered when using other evaluation practices. Of course, not all programs can be evaluated with randomized evaluations; for example, issues such as central bank independence must rely on other methods of evaluation. Programs targeted to individuals or local communities, such as sanitation,

education, and health programs and local government reforms, are likely to be strong candidates for randomized evaluations. This paper does not recommend conducting all evaluations with randomized methods; rather, it starts from the premise that there is scope for substantially increasing their use, and that even a modest increase could have a tremendous impact.

This paper proceeds as follows: The next section presents the impact evaluation problem and the opportunities for evaluation and discusses examples of evaluations, drawn mostly from India. The following section discusses the potential of randomized evaluation as a basis for scaling up. The paper then turns to current practices and the role international agencies can play in promoting and financing rigorous evaluations, and then the paper concludes.

The Methodology of Randomized Evaluation

This section discusses the methodology of randomized evaluation: the problem it tries to solve, and the solution it provides. It presents various examples where the methodology is applied.

The Evaluation Problem

Any impact evaluation attempts to answer essentially counterfactual questions: How would individuals who did benefit from the program have fared in the absence of the program? How would those who did not benefit have fared if they had been exposed to the program? The difficulty with these questions is immediate: at a given point in time, an individual is observed either exposed to the program or not exposed. Comparing the same individual over time will not, in most cases, provide a reliable estimate of the impact the program had on him or her, because many other things may have changed at the same time that the program was introduced. We therefore cannot seek to obtain an estimate of the impact of the program on each individual. All we can hope for is to be able to obtain the average impact of the program on a group of individuals by comparing them with a similar group that was not exposed to the program.

Thus the critical objective of impact evaluation is to establish a credible comparison group, that is, a group of individuals who, in the absence of the program, would have had outcomes similar to those who were exposed to the program. This group gives us an idea of what would have happened to the program group if it had not been exposed, and thus allows us to obtain an estimate of the average impact on the group in question. Generally, in the real world, individuals who were subjected to the program and those who were not are very different, because programs are placed in specific areas (for example, poorer or richer areas), individuals are screened for participation in the program (for instance, on the basis of poverty or on the basis of their motivation), and the decision to participate is often voluntary. For all these reasons, those who were not exposed to a program are often not comparable to those who were. Any difference between them could be attributed to two factors: preexisting

differences (the so-called selection bias) and the impact of the program. Because we have no reliable way to estimate the size of the selection bias, we cannot decompose the overall difference into a treatment effect and a bias term.

To solve this problem, program evaluations typically need to be carefully planned in advance to determine which group is a likely comparison group. One situation where the selection bias disappears is when the treatment and the comparison groups are selected randomly from a potential population of beneficiaries (individuals, communities, schools, or classrooms can be selected into the program). In this case, on average, we can be assured that those who are exposed to the program are no different than those who are not, and that a statistically significant difference between them in the outcomes that the program was planning to affect after the program is in place can be confidently attributed to the program. I will now discuss several examples of randomized evaluations.

Prospective Randomized Evaluations

Random selection of treatment and comparison groups can happen in several circumstances: during a pilot project because the program's resources are limited, or because the program itself calls for random beneficiaries. The next two subsections discuss examples of these different scenarios. In addition, in some circumstances where a program was not randomly allocated, because of favorable circumstances a credible control group nevertheless exists.

Pilot Projects

Before a program is launched on a large scale, a pilot project, necessarily limited in scope, is often implemented. In most circumstances, the beneficiaries of the pilot can be randomly chosen, because many potential sites (or individuals) are equally good locations for the project. The pilot can then be used not only to find out if the program turns out to be feasible (which is what most pilots are currently used for) but also whether the program has the expected impacts. Job training and income maintenance programs are prominent examples of randomized evaluations. A growing number of such pilot projects are evaluated, often by means of collaboration between an NGO and academics (see, for example, Kremer 2003 for several references). To illustrate briefly how such studies can work in practice, consider an example from India analyzed in Banerjee and others (2001). This study evaluated a program where an Indian NGO, Seva Mandir, decided to hire second teachers for the nonformal education centers it runs in villages. Nonformal schools seek to provide basic numeracy and literacy skills to children who do not attend formal school and, in the medium term, to help mainstream these children into the regular school system. These centers are plagued by high teacher and child absenteeism. A second teacher, often a woman, was randomly assigned to 21 out of 42 schools. The hope was to increase the number of days the school was open, to increase children's participation, and to improve performance by providing more individualized attention to the children. By providing a female teacher, the NGO also hoped to make school more attractive for girls. Teacher and child attendance were regularly monitored in program and comparison

schools for the entire duration of the project. The impact of the program on learning was measured by testing children at the end of the school year. The program reduced the number of days schools were closed: one-teacher schools were closed 39 percent of the time, whereas two-teacher schools were closed 24 percent of the time. Girls' attendance increased by 50 percent. However, test scores did not differ.

Carefully evaluated pilot projects form a sound basis for the decision to scale the project up. In the example just discussed, the NGO did not implement the two-teacher program on a full scale on the grounds that its benefits did not outweigh its costs. The NGO used the savings to expand other programs.

By contrast, positive results can help build consensus for a project that has the potential to be extended far beyond the scale that was initially envisioned. The PRO-GRESA program (subsequently renamed as PROGRESA-Oportunidades) in Mexico is the most striking example of this phenomenon.[3] PROGRESA offers grants, distributed to women, conditional on children's school attendance and preventative health measures (nutrition supplementation, health care visits, and participation in health education programs). In 1998, when the program was launched, Mexican government officials made a conscious decision to take advantage of the fact that budgetary constraints made it impossible to reach the 50,000 potential beneficiary communities of PROGRESA all at once, and instead started with a pilot program in 506 communities. Half of those were randomly selected to receive the program, and baseline and subsequent data were collected in the remaining communities (Gertler and Boyce 2001). Part of the rationale for starting with this pilot program was to increase the probability that the program would be continued in case of a change in the party in power. The proponents of the program understood that to be scaled up successfully, the program would require continuous political support.

The task of evaluating the program was given to academic researchers through the International Food Policy Research Institute. The data were made accessible to many different people, and a number of papers have been written on the program's impact (most of them are accessible on the institute's web site).[4] The evaluations showed that PROGRESA was effective in improving health and education. Comparing PROGRESA beneficiaries and nonbeneficiaries, Gertler and Boyce (2001) show that children had about a 23 percent reduction in the incidence of illness, a 1 to4 percent increase in height, and an 18 percent reduction in anemia. Adults experienced a reduction of 19 percent in the number of days of work lost because of illness. Shultz (forthcoming) finds an average of a 3.4 percent increase in enrollment for all students in grades 1 through 8. The increase was largest among girls who had completed grade 6: 14.8 percent.

In part because the program had been shown to be successful, it was indeed maintained when the Mexican government changed hands: by 2000 it was reaching 2.6 million families (10 percent of the families in Mexico) and had a budget of US$800 million or 0.2 percent of Mexico's gross domestic product (Gertler and Boyce 2001). It was subsequently expanded to urban communities, and, with support from the World Bank, several neighboring Latin American countries are implementing similar programs. Mexican officials transformed a budgetary constraint into an opportunity and made evaluation the cornerstone of subsequent scaling up. They were

rewarded by the expansion of the program and by the tremendous visibility that it acquired.

Replication and Evaluation of Existing Projects

A criticism often heard against the evaluation of pilot projects is that pilot projects may be different from regular projects. The fact that pilot projects are evaluated can create problems with the interpretation of the results. If the project is unsuccessful, it may be because it faced implementation problems during the first phase of the program. If it is successful, it may be because more resources were allocated to it than would have been under a more realistic scenario, because the context was favorable, or because the participants in the experiment had a sense of being part of something and changed their behavior. Moreover, all programs are implemented under particular circumstances, and the conclusions may be hard to generalize.

A first answer to some of these concerns is to replicate successful (as well as potentially successful) experiments in different contexts. This has two advantages. First, in the process of transplanting a program, circumstances will require changes, and the program will show its robustness if its effectiveness survives these changes. Second, obtaining several estimates in different contexts will provide some guidance about whether the impacts of the program differ significantly for different groups. Replication of the initial evaluation study in the new context does not imply delaying full-scale implementation of the program if the latter is justified on the basis of existing knowledge. More often than not, the introduction of the program can only proceed in stages, and the evaluation only requires that beneficiaries be phased into the program in random order.

Two studies on school-based health interventions provide a good illustration of these two benefits. The first study (Miguel and Kremer 2004) evaluates a program of twice-yearly, school-based mass treatment with inexpensive deworming drugs in Kenya, where the prevalence of intestinal worms among children is high. Seventy-five schools were phased into the program in random order. Health and school participation improved not only at program schools, but also at nearby schools because of reduced disease transmission. Absenteeism in treatment schools was 25 percent (or 7 percentage points) lower than in comparison schools. Including this spillover effect, the program increased schooling by 0.15 years per person treated. Combined with estimates about the rates of return to schooling, the estimates suggest extremely high rates of return of the deworming intervention. The authors estimate that deworming increases the net present value of wages by more than US$30 per treated child at a cost of only US$0.49.

One of the authors then decided to examine whether these results generalized among preschoolers in urban India (Bobonis, Miguel, and Sharma 2002). The baseline revealed that, even though worm infection was present, the levels of infection were substantially lower than in Kenya (in India, "only" 27 percent of children suffer from some form of worm infection, compared with 92 percent in Kenya). However, 70 percent of children had moderate to severe anemia. The program was therefore modified to include iron supplementation. It was administered through a network of preschools

in urban India. After one year of treatment, the authors found a nearly 50 percent reduction in moderate to severe anemia, large weight gains, and a 7 percent reduction in absenteeism among four- to six-year-olds (but not for younger children). The results of the previous evaluation were thus by and large vindicated.[5]

A second answer is to evaluate the impact of programs that have already shown their potential to be implemented on a large scale. In this case, concerns about the ability to expand the program are moot, at least at the level at which it was implemented. It may also make evaluating the program in several sites at the same time easier, thereby alleviating some of the concerns about external validity. A natural occasion for such evaluation is when the program is ready to expand, and the expansion can be phased in in random order.

The evaluation of a remedial education program by Banerjee and others (2003) is an example of this approach. The program has been run by Pratham, an Indian NGO, which implemented it in 1994. Pratham now reaches more than 161,000 children in 20 cities. The remedial education program hires a young woman from the children's community to provide remedial education in government schools to children who have reached grades 2, 3, or 4 without having mastered the basic grade 1 competencies. Children who are identified as lagging behind are pulled out of their regular classroom for two hours a day to receive this instruction. Pratham wanted to evaluate the impact of this program, one of their flagship interventions, at the same time as they were looking to expand. The expansion into a new city, Vadodara, provided an opportunity to conduct a randomized evaluation. In the first year (1999–2000), the program was expanded to 49 randomly selected schools out of the 123 Vadodara government schools. In 2000–01, the program was expanded to all the schools, but half the schools got a remedial teacher for grade 3 and half got one for grade 4. Grade 3 students in schools that got the program in grade 4 served as the comparison group for grade 3 students in schools who got the program in grade 3. At the same time, a similar intervention was conducted in a district of Mumbai, where half the schools got the remedial teachers in grade 2 and half got them in grade 3. The program was continued for one more year, with the schools switching groups. Thus the program was conducted in several grades, in two cities, and with no school feeling that it had been deprived of resources relative to the others, because all schools benefited from the program.

After two years, the program increased the average test score by 0.39 standard deviations, which represents an increase of 3.2 points out of a possible 100 (the mean in the control group was 32.4 points), and had an even stronger impact on the test scores of those children who had low scores initially (an increase of 3.7 points, or 0.6 standard deviation, on a basis of 10.8 points). The impact of the program is rising over time, but it is similar across cities and genders. Hiring remedial education teachers from the community appears to be 10 times more cost-effective than hiring new teachers. One can be relatively confident in recommending the scaling up of this program, at least in India, on the basis of these estimates, because the program was continued for a period of time, it was evaluated in two quite different contexts, and it has shown its ability to be rolled out on a large scale.

Program-Induced Randomization

In some instances, fairness or transparency considerations make randomization the best way to choose the recipients of a program. Such programs are natural candidates for evaluation, because the evaluation exercise does not require any modification of the program's design.

When some schools are oversubscribed, allocation to particular schools is often done by lottery. In some school systems in the United States, students have the option of applying to so-called magnet schools or schools with special programs, and admission is often granted by lottery. Cullen, Jacob, and Levitt (2002) use this feature to evaluate the impact of school choice in the Chicago school system by comparing lottery winners and losers. Because each school runs its own lottery, their paper is, in effect, taking advantage of 1,000 different lotteries. They find that lottery winners are less likely to attend their neighborhood schools than lottery losers, but more likely to remain in the Chicago school system. However, their subsequent performance is actually worse than that of lottery losers. This is in sharp contrast to expectations and what a "naive" comparison would have found. When one simply compares the results of all the children who attended the school of their choice to the results of all those who did not, one finds that the results of children who attended a school of their choice are indeed better than the results of those who did not. The results from the randomized evaluation show, however, that, if anything, the causal effect of attending a school of one's choice is negative. The "naive" difference, which is positive, simply reflects the fact that the children who decided to change schools were highly motivated.

Voucher programs constitute another example of programs that often feature a lottery. The sponsor of the program allocates only a limited budget to the program, the program is oversubscribed, and a lottery is used to pick the beneficiaries. Angrist and others (2002) evaluate a Colombian program in which vouchers for private schools were allocated by lottery because of the program's limited budget. Vouchers were renewable conditional on satisfactory academic performance. The authors compare lottery winners and losers. Lottery winners were 15 to 20 percent more likely to attend private school; 10 percent more likely to complete grade 8; and scored 0.2 standard deviations higher on standardized tests, equivalent to a full grade level. Winners were substantially more likely to graduate from high school and scored higher on high school completion and college entrance examinations. The benefits of this program to participants clearly exceeded the costs, which were similar to the costs of providing a public school place.

When nationwide policies include some randomization aspect, this provides a unique opportunity to evaluate a policy that has already been scaled up in several locations. The knowledge gained from this experience can be used to inform policy decisions to expand the policy in the countries, to continue with the program, or to expand the policy in other countries. However, because the randomization is part of the program design rather than a deliberate attempt to make evaluating it possible, the data necessary for the evaluation are not always available. International agencies

can play two key roles in this respect. First, they can organize and finance limited data collection efforts. Second, they can encourage governments and statistical offices to link up existing data sources that can be used to evaluate the experiments. Set-asides for women and minorities in the decentralized government in India (the *panchayat* system) are an interesting example. In 1993, the 73rd amendment to the Constitution of India required the states to set up a three-tiered *panchayat* system (village, block, and district levels), directly elected by the people, to administer local public goods. Elections must take place every five years, and *panchayat* councils have the latitude to decide how to allocate local infrastructure expenditures. The amendment also required that one-third of all positions (council members and council chairs) be reserved for women, and that a share equal to the representation of disadvantaged minorities (scheduled castes and scheduled tribes) be reserved for these minorities. To avoid any possible manipulation, the law stipulated that the reserved positions be randomly allocated.

Chattopadhyay and Duflo (forthcoming) evaluate the impact of reserving seats for women in West Bengal. They collected data on 465 villages across 165 councils in 1 district, and find that women tend to allocate more resources to drinking water and roads and less to education. This corresponds to the priorities men and women expressed through their complaints to *panchayats*. Then they collected the same data in a poor district of Rajasthan, Udaipur. They find that in Rajasthan, women invest more in drinking water and less on roads, and that this once again corresponds to the complaints expressed by men and women. These results were obtained in two very different districts with different histories: West Bengal had had a *panchayat* since 1978, while Rajasthan had none until 1995, plus Rajasthan has particularly low female literacy among Indian states. Thus the results suggest that the gender of policy-makers matters in both more and less developed political systems. Furthermore, it provides indirect, but powerful, evidence that local elected officials do have power even in relatively young systems. Chattopadhyay and Duflo (forthcoming) also evaluate the impact of reservations for scheduled castes and find that a larger share of goods is assigned to scheduled caste hamlets when the head of a *panchayat* is from a scheduled caste.

In principle, the data to evaluate the impact of this experiment on a much larger scale are available: village-level census data are available for 1991 and will become available for 2001. The National Sample Survey Organization conducts large-scale consumption and labor surveys every five years, with detailed data on outcomes. However, administrative barriers make these data difficult to use for the purpose of evaluating this program, because the census does not contain any information about which *panchayat* a village belongs to. In addition, the information about *panchayat* reservations and composition is not centralized, even at the state level, and is available only at the district level. Likewise, the National Sample Survey contains no information about *panchayats*. This is an example where, at a relatively small cost, information could be made available that would be useful for evaluating an extremely large program. It requires coordinating various people and agencies, a task that international organizations should be well placed to accomplish.

Other Methods to Control for Selection Biases

Natural or organized randomized experiments are not the only methodology that can be used to obtain credible impact evaluation of program effects. To compensate for the lack of randomized evaluations, researchers have developed alternative techniques to control for selection bias as best as possible. Labor economists in particular have made tremendous progress. (For excellent technical and nontechnical surveys of the various techniques, their value, and their limitations, see, for example, Angrist and Krueger 1999, 2001; Card 1999; and Meyer 1995.) Here I briefly mention some of the techniques that are most popular with researchers.

One strategy is to try to find a control group that is as comparable as possible with the treatment group, at least along observable dimensions. This can be done by collecting as many covariates as possible, and adjusting the computed differences through a regression or by matching the program and the comparison group, that is, by forming a comparison group that is as similar as possible to the program group. One way to proceed is to predict the probability that a given individual is in the comparison or the treatment group on the basis of all the available observable characteristics and to form a comparison group by picking people who have the same probability of being treated as those who actually got treated. Rosenbaum (1995) refers to this as propensity score matching. The challenge with this method, as with regression controls, is that it hinges on having identified all the potentially relevant differences between treatment and control groups. In cases where treatment is assigned on the basis of a variable that is not observed by the researcher, such as demand for the service, this technique will lead to misleading inferences.

When a good argument can be made that the outcome would not have had differential trends in regions that received the program if the program had not been put in place, it is possible to compare the growth in the variables of interest between program and nonprogram regions (this is often called the difference-in-differences technique). Whether the argument is good and the identification assumptions are justified is, however, often hard to judge. This identification assumption cannot be tested, and to even ascertain its plausibility, one needs to have long time series of data from before the program was implemented to be able to compare trends over long enough periods. One also needs to make sure that no other program was implemented at the same time, which is often not the case. Finally, when drawing inferences, one needs to take into account that regions are often affected by time-persistent shocks, which may look like a program effect (Bertrand, Duflo, and Mullainathan 2004).

Duflo (2001) takes advantage of a rapid school expansion program that took place in Indonesia in the 1970s to estimate the impact of building schools on schooling and subsequent wages. Identification is made possible because the allocation rule for schools is known (more schools were built in places with low initial enrollment rates), and because the cohorts benefiting from the program are easily identified (children 12 or older when the program started did not benefit from the program). The faster growth of education across cohorts in regions that got more schools suggests that access to schools contributed to increased education. The trends were similar before the program and shifted clearly for the first cohort that was exposed to the program,

which reinforces confidence in the identification assumption. This identification strategy is not often valid, however. Frequently when policy changes are used to identify the effect of a particular policy, the policy change is itself endogenous to the outcomes it tried to affect, which makes identification impossible (see Besley and Case 2000).

Program rules often generate discontinuities that can be used to identify the effects of the program by comparing those who were just above the threshold to qualify for a program to those who were just below the threshold. For example, if scholarships are allocated on the basis of a certain number of points, one can compare those just above to those just below the threshold. Angrist and Lavy (1999) use this technique, known as regression discontinuity design (Campbell 1969) to evaluate the impact of class size in Israel. In Israel, a second teacher is allocated whenever the size of a class would be larger than 40 children. This generates discontinuities in class size when the enrollment in a grade goes from 40 to 41 (class size changes from 40 to 20 and 21), 80 to 81, and so on. Angrist and Lavy compare test score performance in schools just above and just below the threshold and find that those just above the threshold have significantly higher test scores than those just below, which can confidently be attributed to the class size, because it is difficult to imagine that schools on both sides of the threshold have any other systematic differences. Discontinuities in program rules, when enforced, are thus a source of identification. However, they often are not enforced, especially in developing countries. For example, researchers tried to use the discontinuity in Grameen Bank, the flagship micro-credit organization in Bangladesh that lends only to people who own less than one acre of land (Pitt and Khandker 1998), as a source of identification. However, in practice Grameen Bank lends to many people who own more than one acre of land, and there is no discontinuity in the probability for borrowing at the threshold (Morduch 1998). In developing countries rules are probably frequently not enforced strictly enough to generate discontinuities that can be used for identification purposes.

Alternatives to randomized evaluation exist, and they are useful; however, identification issues need to be tackled with extreme care and they are never self-evident. They generate intense debate in academic circles whenever such a study is conducted. Identification is less transparent, and more subject to divergence of opinion, than in the case of randomized experiments. The difference between good and bad evaluations of this type is thus more difficult to communicate. The study and the results are also less easy to convey to policymakers in an effective way with all the caveats that need to accompany them. This suggests that, while a mix of randomized and non-randomized evaluation is necessary, international organizations should commit themselves to running some randomized evaluations.

Scaling Up and Randomized Evaluations

The previous section showed that when programs' beneficiaries are individuals or communities, rather than an entire country, for example, randomized evaluations are

often a possible way to obtain reliable estimates of the programs' effects. This section discusses how the results of these evaluations can be used to scale up development programs.

Obtaining Reliable Estimates of Program Impact

When the evaluation is not planned ex ante, to evaluate the impact of a program researchers must resort to before and after comparisons (when a baseline was conducted), or to comparisons between beneficiaries and communities that, for some reason, were not exposed to the program. When the reasons why some people were exposed to the program and some were not are unknown, or worse, when they are known to be likely to introduce selection bias, those comparisons are likely to be biased. The data collection is often as expansive as for a randomized evaluation, but the inferences are biased. As argued earlier, controlling for observable differences between treatment and control groups through a regression analysis or propensity score matching will correct for the bias only if beneficiaries and nonbeneficiaries are known with certainty to be comparable conditional on these characteristics. This is unlikely to be true unless the program was randomly allocated conditional on these characteristics. In particular, a project officer trying to optimally allocate a program typically has more information than a researcher, and will (and should) make use of it when allocating resources.

These concerns have serious practical implications. Studies comparing experimental and nonexperimental estimates with the same data show that the results from randomized evaluation can be quite different from those drawn from non-randomized evaluation. In a celebrated analysis of job training programs, LaLonde (1986) finds that many of the econometric procedures and comparison groups used in program evaluations did not yield accurate or precise estimates and that such econometric estimates often differed significantly from experimental results. Glewwe and others (forthcoming) compare retrospective and prospective analyses of the effect of flip charts in schools on test scores. Retrospective estimates using straightforward ordinary least squares regressions suggested that flip charts raised test scores by up to 20 percent of a standard deviation, robust to the inclusion of control variables, while difference-in-differences estimates suggested a smaller effect of about 5 percent of a standard deviation, an effect that is still significant, though sometimes only at the 10 percent level. In contrast, prospective estimates based on randomized evaluations provided no evidence that flip charts increased test scores. These results suggest that using retrospective data to compare test scores seriously overestimates the charts' effectiveness. A difference-in-differences approach reduced, but did not eliminate, the problem and, moreover, whether such a difference-in-differences approach has general applicability is not clear. These examples suggest that ordinary least squares estimates are biased upward rather than downward. This is plausible, because in a poor country with a substantial local role in education, inputs are likely to be correlated with favorable, unobserved community characteristics. If the direction of omitted

variable biases were similar in other retrospective analyses of education inputs in developing countries, the effects of inputs may be even more modest than retrospective studies suggest.

Some of the results are more encouraging. For example, Buddelmeyer and Skoufias (2003) use randomized evaluation results as a benchmark to examine the performance of regression discontinuity design for evaluating the impact of the PROGRESA program on child health and school attendance. The researchers found the performance of regression discontinuity design in this case to be remarkably good: impact estimates with this quasi-experimental method agreed with experimental evidence in 10 out of 12 cases, and the 2 exceptions both occurred in the first year of the program. Such research can provide invaluable guidance about the validity and potential biases of quasi-experimental estimators.

Another important source of bias in program effects are publication biases. Positive results tend to receive a large amount of publicity. Agencies that implement programs seek publicity for their successful projects, and academics, as well as academic journals, are much more interested in and able to publish positive results than modest or insignificant results. However, many programs fail, and publication bias may be substantial if only positive and significant results are published.

The problem of publication bias may be much larger with retrospective evaluations. Ex post the researchers or evaluators define their own comparison group, and thus may be able to pick a variety of plausible comparison groups. In particular, researchers obtaining negative results with retrospective techniques are likely to try different approaches or not to publish.

Available evidence suggests that the publication bias problem is severe (DeLong and Lang 1992). In the case of natural experiments and instrumental variable estimates, publication bias may actually more than compensate for the reduction in bias caused by the use of an instrument, because they tend to have larger standard errors, and researchers looking for significant results will select only large estimates. For example, Ashenfelter, Harmon, and Oosterbeek (1999) provide strong evidence of publication bias of instrumental variables estimates of the returns to education: on average, the estimates with larger standard errors also tend to be larger. This accounts for most of the oft-cited results that claim that instrumental estimates of the returns to education are higher than ordinary least squares estimates. In contrast, randomized evaluations commit in advance to a particular comparison group. Once the work to conduct a prospective randomized evaluation has been done, researchers just need to ensure that the results are documented and published even if the results suggest quite modest effects, or even no effects at all (such as some of the studies discussed in this paper). As I discuss later, putting institutions in place to ensure that negative results are systematically disseminated is important (such a system is already in place for the results of medical trials).

Several sources of bias are specific to randomized evaluation, but they are well known and can often be corrected for. The first possibility is that the initial randomization is not respected; for example, a local authority figure insists that the school in his village be included in the group scheduled to receive the program, or parents

manage to reallocate their children from a class or a school without the program to a class or school with the program. Or conversely, individuals allocated to the treatment group may not receive the treatment, for example, because they decide not to take the program up. Even though the intended allocation of the program was random, the actual allocation is not. In particular, the program will appear to be more effective than it is in reality if individuals allocated to the program ex post also receive more of other types of resources, which is plausible. This concern is real, and evaluations certainly need to deal with it; however, it can be dealt with relatively easily. Even though the initial assignment does not guarantee in this case that someone is actually either in the program or in the comparison group, in most cases it is at least more likely that someone is in the program group if he or she was initially allocated to it. The researcher can thus compare outcomes in the initially assigned group (this difference is often called the intention to treat estimate) and scale up the difference by dividing it by the difference in the probability of receiving the treatment in those two groups (Imbens and Angrist 1994). Krueger's (1999) reanalysis of the Tennessee student/teacher achievement ratio class size experiment used this method to deal with the fact that some parents had managed to reallocate their children from regular classes to small classes.[6] Such methods will provide an estimate of the average effect of the treatment on those who were induced to take the treatment by the randomization, for instance, on children who would have been in a large class had they not been placed in the treatment groups. This may be different from the average effect in the population, because people who anticipated benefiting more from the program may be more likely to take advantage of it. It may, however, be a group that policymakers especially care about, because they are likely to be the ones who are more likely to take advantage of the policy if it is implemented on a large scale.

A second possible source of bias is differential attrition in the treatment and comparison groups: those who benefit from the program may be less likely to move or otherwise drop out of the sample than those who do not. For example, the two-teacher program Banerjee and others (2001) analyze increased school attendance and reduced dropout rates. This means that when a test was administered in the schools, more children were present in the program schools than in the comparison schools. If children who are prevented by the program from dropping out of school are the weakest in the class, the comparison between the test scores of the children in treatment and control schools may be biased downward. Statistical techniques can be used to deal with this problem, but the most effective way is to try to limit attrition as much as possible. For example, in the evaluation of the remedial education program in India (Banerjee and others 2003), an attempt was made to track down all children and administer the test to them, even if they had dropped out of school. Only children who had left for their home villages were not tested. As a result, the attrition rate remained relatively high, but was the same in the treatment and comparison schools and does not invalidate test score comparisons.

A third possible source of bias is when the comparison group is itself indirectly affected by the treatment. For example, Miguel and Kremer's (2004) study of the Kenyan deworming program showed that children in treatment schools and in

schools near the treatment schools were less likely to have worms, even if they were not themselves given the medicine. The reason is that worms easily spread from one person to another. In previous evaluations, treatment had been randomized within schools. Its impact was thus underestimated, because even comparison children benefited from the treatment. The solution in this case was to choose the school rather than the pupils within a school as the unit of randomization.

Randomizing across units—for example, across schools or communities rather than individuals within a unit—is also often the only practical way to proceed. For example, offering a program to some villagers and not others may be impossible, but the fact that randomization takes places at the group rather than the individual level needs to be explicitly taken into account when calculating the confidence interval of the estimates of the impact of the program. Imagine, for example, that only two large schools take part in a study, and that one school is chosen at random to receive new textbooks. The differences in test scores between children in the two schools may reflect many other characteristics of the treatment and comparison schools (for example, the quality of the principal). Even if the sample of children is large, the sample of schools is actually small. The grouped nature of the data can easily be taken into account, but it is important to take it into account when planning design and sample size.

In summary, while randomized evaluations are not a bullet-proof strategy, the potential for biases is well known, and those biases can often be corrected. This stands in sharp contrast with biases of most other types of studies, where the bias caused by the nonrandom treatment assignment cannot be either signed or estimated.

Generalizing the Results of Evaluation

Randomized evaluation can therefore provide reliable estimates of treatment effects for the program and the population under study. To draw on these estimates to assess the prospects for scaling up the program, however, one has to make the case that these estimates tell us something about the effect of the program after it is scaled up. There are different reasons why the results of a well-executed experiment may not be generalizable.

First, the experiment itself may have affected the treatment or the comparison samples, for example, the provision of inputs might temporarily increase morale among beneficiaries, and this could improve performance (known as the Hawthorne effect). While this would bias randomized evaluations, it would also bias fixed-effect or difference-in-differences estimates. As mentioned earlier, either the treatment or the comparison group may also be temporarily affected by being part of an experiment (known as the John Henry effect). These effects are less likely to be present when the evaluations are conducted on a large scale and over a long enough time span, and some experimental designs can minimize the risk of such effects. For example, in Pratham's remedial education program analyzed by Banerjee and others (2003), all the schools received the program, but not all the grades. Trying to assess whether these effects are present is, however, important. In his reanalysis of

the project student/teacher achievement ratio data, Krueger (1999) exploits variation in class size within the control group occasioned by children's departure during the year to obtain a second estimate of the class size effect, which is, by definition, not contaminated by John Henry or Hawthorne effects, because all the teachers in this sample belong to the control group. He finds no difference in the estimates obtained by these two methods.

Second, treatment effects may be affected by the scale of the program. For example, the Colombian voucher program Angrist and others (2002) analyze was implemented on a pilot basis with a small sample, but the rest of the school system remained unchanged, in particular, the number of students affected was too small to have an impact on the composition of the public and private schools. If this program were to be implemented on a large scale, it could affect the functioning of the school system, and could therefore have a different impact (Hsieh and Urquiola 2002). More generally, partial equilibrium treatment effects may be different from general equilibrium treatment effects (Heckman, Lochner, and Taber 1998). Addressing these problems requires randomized evaluation to be performed at the level of the economy. This may be possible for programs such as voucher programs, where the general equilibrium effects will likely take place at the community level, and where communities can be randomly affected or not affected by the program, but I am not aware of an evaluation of this type.

Third, and perhaps most important, no project will be replicated exactly: circumstances vary and any idea will have to be adapted to local circumstances. In other words, internal validity is not sufficient. The evaluation also needs to have some external validity, that is, the results can be generalized beyond the population directly under study. Some argue that evaluation can never generalize. In its most extreme form (see, for example, Cronbach 1982; Cronbach and others 1980; see also the review of the education literature in Cook 2001), this argument contends that every school, for example, is specific and complex, and that nothing definitive can be learned about schools in general. This discourse has made its way into some international organizations,[7] but note that it is contradictory to the objective of going to scale. What is the point of rolling out a program on a large scale if one thinks that, for example, each school needs a different program? The very objective of scaling up has to be founded on the postulate that even if the impact of a program varies across individuals, thinking of average treatment effects makes sense. This is exactly the postulate that underlies the external validity of randomized evaluations.

A theory of why a specific program is likely to be effective is necessary to provide some guidance about what elements in the program and in its context were keys to its success. Theory will help disentangle the distinct components of a program and discriminate between variants that are likely to be important and variants that are not (Banerjee 2002). For example, an economic analysis of the PROGRESA program suggests that it may have been useful because of its impact on income, because of its effect on women's bargaining power, or because of its effect on incentives. Aspects of the program most likely to be relevant to the program's success are the size of the transfer, its recipient, and the conditionality attached to it. In contrast, the color of

the food supplement distributed to the families, for example, is unlikely to be important. Replication of the programs may then vary these different aspects to determine which of them is the most important. This also suggests that priority should be given to evaluating programs that are justified by some well-founded theoretical reasoning, because the conclusions from the evaluation are then more likely to generalize.

Theory provides some guidance about what programs are likely to work and, in turn, the evaluation of these programs forms a test of the theory's prediction. Because prospective evaluations need to be planned ahead of time, designing pilot programs in such a way that they help answer a specific question or test a specific theory is also often possible. For example, Duflo, Kremer, and Robinson (2003) report on a series of randomized evaluations conducted in Kenya in collaboration with International Christian Support (ICS) Africa, a Netherlands-based NGO active in the area. They were motivated by the general question: why do so few farmers in this region of Kenya use fertilizer (only about 10 percent), even though its use seems to be profitable and it is widely used in other developing countries, as well as in other regions of Kenya? They first conducted a series of trials on the farms owned by randomly selected farmers and confirmed that, in small quantities, fertilizer is extremely profitable: the rates of return were often in excess of 100 percent. They then conducted a series of programs to answer a number of other questions: Do farmers learn when they try fertilizer out for themselves? Do they need information about returns or about how to use them? Does the experiment need to take place on their farm, or can it take place on a neighbor's farm? Do they learn from their friends? To answer these questions, the researchers first randomly selected farmers to participate in the field trials and followed their adoption of fertilizer subsequently, as well as that of a comparison group. Second, they also followed adoption by the friends and neighbors of the comparison farmers. Finally, they invited randomly selected friends of farmers participating in the trials to the important stages in the development of the experiment and monitored their subsequent adoption.

These questions are extremely important to our understanding of technology adoption and diffusion, and the ability to generate exogenous variation through randomized program evaluation greatly helped in this understanding. Moreover, the answers also helped International Christian Support Funds develop a school-based agricultural extension program that has a chance to be effective and cost-effective. A pilot version of this program is currently being evaluated.

Thus theory and existing evidence can be used to design informative replication experiments and to sharpen predictions from these experiments. Rejection of these predictions should then be taken seriously and will inform the development of the theory. Replication is one area where international organizations, which are present in most countries, can play a key role if they take the time to implement randomized evaluations of programs that can be replicated. An example of such an opportunity that was seized is the replication of PROGRESA in other Latin American countries. Encouraged by the success of PROGRESA in Mexico, the World Bank encouraged (and financed) Mexico's neighbors to adopt similar programs. Some of these programs have included a randomized evaluation and are currently being evaluated.

Note also that the exogenous variation created by the randomization can be used to help identify a structural model. Attanasio, Meghir, and Santiago (2001) and Behrman, Sengupta and Todd (2002) are two examples of such an exercise using the PROGRESA data to predict the possible effects of varying the schedule of transfers. These studies rest on assumptions that one is free to believe or not, but at least they are freed of some assumption by the presence of this exogenous variation. The more general point is that randomized evaluations do not preclude the use of theory or assumptions. Indeed, they generate data and variation that can help identify some aspects of these theories.

Assessing the Feasibility of Randomized Evaluation

As noted in the introduction, randomized evaluations are not adapted for all types of programs. They are adapted to programs that are targeted to individuals or communities and where the objectives are well defined. For example, the efficacy of foreign aid disbursed as general budget support cannot be evaluated in this way. It may be desirable, for efficiency or political reasons, to disburse some fraction of aid in this form, although it would be extremely costly to distribute all foreign aid in the form of general budget support, precisely because it leaves no place for rigorous evaluation of projects. However, in many cases randomized evaluations are feasible.

The main cost of evaluation is the cost of data collection, and it is no more expensive than the cost of collecting any other data. Indeed, by imposing some discipline on which data to collect (the outcomes of interest are defined ex ante and do not evolve, as the program fails to affect them) may reduce the cost of data collection relative to a situation where what is being measured is not clear. Several potential interventions can also be evaluated in, say, the same groups of schools, as long as the comparison and treatment groups for each intervention are "criss-crossed." This has the added advantage of making is possible to directly compare the efficacy of different treatments. For example, in Vadodara, Pratham implemented a computer-assisted learning program in the same schools where the remedial education program evaluated by Banerjee and others (2003) was implemented. The program was implemented only in grade 4. Half the schools that had the remedial education program in grade 4 got the computer-assisted learning program, and half the schools that did not have the remedial education program got the computer-assisted learning program. The preliminary results suggest that the effect on mathematics is comparable to the effect of the remedial education program, but the cost is much smaller. Even keeping the budget of process evaluation constant, a reallocation of part of the money that is currently spent on unconvincing evaluation would probably go a long way toward financing the same number of randomized evaluations. Even if randomized evaluations turn out to be more expensive, the cost is likely to be trivial in comparison with the amount of money saved by avoiding the expansion of ineffective programs. This suggests that randomized evaluation should be financed by international organizations.

Political economy concerns sometimes make not implementing a program in the entire population difficult, especially when its success has already been demonstrated;

for example, the urban version of PROGRESA will not start with a randomized evaluation, because of the strong opposition to delaying some people's access to it. This objection can be tackled at several levels. First, opposition to randomization is less likely to falter in an environment where it has strong support, especially if a rule prescribes that an evaluation is necessary before full-scale implementation.

Second, if, as argued earlier, evaluations are not financed by loans but by grants, this may make it easier to convince partners of their usefulness, especially if they permit countries to expand programs. An example of such explicit partnership is a study on the effectiveness of HIV/AIDS education currently being conducted in Kenya (Duflo and others 2003). With support from UNICEF, the government of Kenya has put together a teacher training program for HIV/AIDS education. Because of a lack of funds, the program's coverage had remained minimal. The Partnership for Child Development, with grants from the World Bank, is funding a randomized evaluation of the teacher training program. ICS Africa is organizing training sessions with facilitators from the Kenyan government. The evaluation allowed training to be expanded to 540 teachers in 160 schools, which would not have been possible otherwise. The randomization was not grounds for the Kenyan authorities to reject the program. On the contrary, at a conference organized to launch the program, Kenyan officials explicitly appreciated the opportunity the evaluation gave them to be at the forefront of efforts to advance knowledge in this area.

The example of PROGRESA showed that government officials recognized the value of randomized evaluation and were actually prepared to pay for it. The favorable response to PROGRESA and the World Bank's subsequent endorsement of the findings will certainly influence how other governments think about experiments. Several examples of this kind could do a lot to move the culture.

Third, governments are far from being the only possible outlets through which international organizations could organize and finance randomized evaluation. Many of the evaluations discussed so far were set up in collaboration between NGOs and academics. NGOs have limited resources and therefore cannot hope to reach all the people they target. Randomized allocation is often perceived as a fair way to allocate sparse resources. In addition, members of NGOs are often extremely entrepreneurial, and as a result NGOs are willing to evolve in response to new information. NGOs tend to welcome information about the effectiveness of their programs, even if they find out that they are ineffective. For these reasons, many NGOs are willing to participate in randomized evaluations of their programs. For example, the collaboration between the Indian NGO Pratham and Massachusetts Institute of Technology researchers, which led to the evaluations of the remedial education and the computer-assisted learning program (Banerjee and others 2003) was initiated by Pratham, which was looking for partners to evaluate their program. Pratham understood the value of randomization and was able to convey it to the schoolteachers involved in the project. International organizations could finance randomized evaluations organized in collaboration with researchers (from their organizations or from academia) and genuine NGOs.

Timing Evaluation and Implementation

Prospective evaluations do take time: convincing studies often go on for two or three years. Obtaining information about a program's long-term impact, which can be extremely important and can differ from the short-run impact, takes even longer. For example, Glewwe, Illias, and Kremer (2003) suggest that a teacher incentive program caused a short-run increase in test scores but no long-run impact, which they attribute to practices of "teaching to the test." When the program targets children but seeks to affect adult outcomes, which is the case for most education or health interventions, the delay between the program and the outcomes may become long. In these cases, waiting for the answer before deciding whether or not to implement the program is not possible.

While this is a real concern, this should not prevent evaluation of the effect of the program on the first cohort to be exposed to the program. While policy decisions will have to be taken in the meantime, knowing the answer at some point is surely better than never knowing it, which would be the case without evaluation. Moreover, obtaining short-term results, which may be used to get an indication of whether or not the program is likely to be effective, is often possible and may guide policy in the short run. For example, in the case of the evaluation of the HIV/AIDS teacher training program, an assessment was performed a few weeks after the program was started and while it was still ongoing. Students in the schools where the teachers were first trained were interviewed about whether the curriculum in their school covered HIV/AIDS and were administered a knowledge, attitude, and practice test. The preliminary results suggested that the program was indeed effective in increasing the chance that HIV/AIDS would be mentioned in class and in improving students' knowledge about HIV/AIDS and HIV prevention. These results could be communicated immediately to the policymakers.

The first results of an evaluation can also be combined with other results or with theory to provide an estimate of what the final impact of the program is expected to be. Obviously, one has to be cautious about such exercises and carefully outline what comes out of the evaluation results and what is the result of assumptions. One should set up programs so that long-run outcomes can be tracked that can then vindicate or invalidate predictions. For example, Miguel and Kremer (2004) combined their estimate of the impact of the deworming program on school participation with estimates of returns to education in Kenya to provide an estimate of the long-term impact on adult productivity, which they used to construct their cost-benefit estimates. They are also continuing to track the children exposed to deworming drugs to directly estimate the drugs' long-run effect.

Finally, delaying some expenditures may actually be worthwhile, given that we know so little about what works and what does not, especially if this can give us an opportunity to learn more. It is disconcerting that we do not know more about what works and what does not work in education, for example, after spending so many years funding education projects. On this scale, the two or three years needed for an evaluation, or even the many more needed to obtain information about the long-run

outcomes, seem a short period of time. It may delay some expenditures, but it will accelerate the process of learning how to make these expenditures usefully. The U.S. Food and Drug Administration (FDA) requires randomized evaluation of the effects of a drug before it can be distributed. Occasionally, the delay the FDA imposes on the approval of new drugs has created resentment, most recently among associations representing AIDS victims; however, randomized trials have played a key role in shaping modern medicine and have accelerated the development of effective drugs.

The Role that International Agencies Can Play

This section discusses current evaluation practices and the role that international agencies can play in improving these practices.

Current Practice

The foregoing examples show that obtaining convincing evidence about the impact of a program is possible by organizing pilot projects, taking advantage of the expansion of existing projects, or taking advantage of project design. While not all programs can be evaluated using these methods, only a tiny fraction of those that could potentially be evaluated actually are. Most international organizations require that a fraction of the budget be spent on evaluation. Some countries also make evaluation compulsory; for example, the Constitution of Mexico requires evaluation of all social programs. However, in practice, this share of the budget is not always spent efficiently; for example, evaluations may be subcontracted to untrained consultancy outfits that are given little guidance about what they should achieve. Worse, they are sometimes entrusted to organizations that have an interest in the outcome, in which case the evaluators have a stake in the results they are trying to establish.

When an evaluation is actually conducted, it is generally limited to a process evaluation, that is, the accounts are audited; the flows of resources are followed; and the actual delivery of the inputs is confirmed, for example, whether textbooks reached the school. In addition, qualitative surveys are used to determine whether beneficiaries actually used the inputs (did the teachers use the textbooks?) and whether there is prima facie evidence that the program beneficiaries were satisfied by the program (were the children happy?). Process evaluation is clearly essential and should be part of any program evaluation: if no textbooks were actually distributed, finding that the program had no impact would hardly be surprising. However, just observing the beneficiaries' reactions to a program can lead to misleading conclusions about its effectiveness. Some programs may, from all observations, seem like resounding successes, even if they did not achieve their objectives. The emphasis on process evaluation implies that, more often than not, when impact evaluations take place they are an afterthought and are not planned for at the time the program starts.

India's District Primary Education Program (DPEP), the largest World Bank–sponsored education program, is an example of a large program that offered the

potential for interesting evaluations, but whose potential on this count was jeopardized by the lack of planning. The DPEP was supposed to be a showcase example of the ability to go to scale with education reform (Pandey 2000). Case (2001) provides an illuminating discussion of the program and the features that make its evaluation impossible. The DPEP is a comprehensive program that seeks to improve the performance of public education. It involves teacher training, inputs, and classrooms. Districts are generally given a high level of discretion in how to spend the additional resources. Despite the apparent commitment to a careful evaluation of the program, several features make a convincing impact evaluation of the DPEP impossible. First, the districts were selected according to two criteria: low level of achievement, as measured by low female literacy rates, but high potential for improvement. In particular, the first districts chosen to receive the program were selected "on the basis of their ability to show success in a reasonable time frame" (Pandey 2000, p. 14). The combination of these two elements in the selection process indicates that any comparison between the level of achievement of DPEP districts and non-DPEP districts would probably be biased downward, while any comparison of improved achievement between DPEP and non-DPEP districts (difference-in-differences) would probably be biased upward. This has not prevented the DPEP from putting enormous emphasis on monitoring and evaluation: the project collected large amounts of data and commissioned numerous reports. However, the data collection process was conducted only in DPEP districts. These data can only be used to do before and after comparisons, which clearly do not make any sort of sense in an economy undergoing rapid growth and transformation. If researchers ever found a credible identification strategy, they would have to use census or national sample survey data.

The Political Economy of Program Evaluation

I have argued that the problems of omitted variable bias that randomized evaluations are designed to address are real and that randomized evaluations are feasible. They are no more costly than other types of surveys and are far cheaper than pursuing ineffective policies. So why are they so rare? Cook (2001) attributes their rarity in education to the postmodern culture in American schools of education, which is hostile to the traditional conception of causation that underlies statistical implementation. Pritchett (2002) argues that program advocates systematically mislead swing voters into believing exaggerated estimates of program impacts. Advocates block randomized evaluations, because they would reveal programs' true impacts to voters. Kremer (2003) proposes a complementary explanation, whereby policymakers are not systematically fooled but have difficulty gauging the quality of evidence, knowing that advocates can suppress unfavorable evaluation results. Program advocates select the highest estimates to present to policymakers, while any opponents select the most negative estimates. Knowing this, policymakers rationally discount these estimates. For example, if advocates present a study showing a 100 percent rate of return, policymakers might assume that the true return is 10 percent. In this environment, if randomized evaluations are more precise (because the estimates are on average

unbiased), there is little incentive to conduct randomized evaluations because they are unlikely to be high enough or low enough that advocates will present them to policymakers.

Under such circumstances, international organizations can play a key role by encouraging randomized evaluations and funding them. Moreover, if policymakers and donors can more readily identify a credible evaluation when examples are already available, which seems plausible, this role can actually start a virtuous circle by encouraging other donors to recognize and trust credible evaluation, and thus advocate the generation of such evaluations as opposed to others. In this way, international organizations can contribute to a climate favorable to credible evaluation and overcome the reluctance noted earlier. The process of quality evaluation itself would then be scaled up above and beyond what the organizations themselves could promote and finance.

What International Agencies Can Do

The foregoing discussion suggests a number of actions that international organizations could undertake to strengthen the role of evaluations.

Defining Priorities for Evaluation

Demanding that all projects be subject to impact evaluation is almost certainly counterproductive. Clearly all projects need to be monitored to ensure that they actually happened, and thus to make sure that the international organization is functioning properly, which is the main responsibility of the organization's evaluation department. Some programs simply cannot be evaluated using the methods discussed in this paper, for example, monetary policy cannot be randomly allocated. Even among projects that could potentially be evaluated, not all need an impact evaluation. Indeed, the value of a poorly identified impact evaluation is low, and its cost, in terms of credibility, is high, especially if, as argued later, international organizations should take a leading role in promoting quality evaluation.

A first objective is thus to cut down on the number of wasteful evaluations. Any proposed impact evaluation should be reviewed by a committee before any money is spent on data collection to avoid a potentially large waste of money. The committee's responsibility would be to assess the ability to deliver reliable, causal estimates of the project. A second objective would be to conduct credible evaluations in key areas. In consultation with a body of researchers and practitioners, each organization should determine key areas for which it will promote impact evaluations. Organizations could also set up randomized evaluations in other areas when the opportunity occurs.

Setting up Autonomous Impact Evaluation Units

Given the scarcity of randomized evaluations, there may be some scope for setting up a specialized unit to encourage, conduct, and finance randomized impact evaluations and to disseminate the results. Such a unit would also encourage data collection and

the study of true natural experiments with program-induced randomization. As noted earlier, randomized evaluations are not the only way to conduct good impact evaluations: when randomization is infeasible, other techniques are available. However, such evaluations are conducted much more routinely, while randomized evaluations are much too rare given their value and the opportunities for conducting them. They also have common features and would benefit from a specialized unit with specific expertise. Because impact evaluation generates international public goods, the unit could finance and conduct rigorous evaluations in the key areas the organization identifies.

Setting up an autonomous unit would have several advantages. First, it would ensure that conducting evaluation is a core responsibility of a team of people. Second, this unit would be free of the fire-walling requirements that are necessary to make the evaluation divisions of international organizations independent, but make prospective evaluations difficult. For example, the director of the World Bank's Operations Evaluation Department reports directly to the board, and the department's teams are prevented from establishing close connections with the implementation team. This makes a prospective randomized evaluation essentially impossible. Third, randomized evaluation and nonrandomized evaluation should be clearly separated to avoid the "scaling down" effect caused by the political economy of evaluation.

Banerjee and He (2003) argue that the World Bank's decisions and reports have little impact on market decisions or on subsequent debates, that is, that the World Bank does not seem to have the role of a leader and promoter of new ideas that it could have. This may be in part because everybody recognizes that the World Bank, perhaps legitimately, operates under a set of complicated constraints, and that what justifies its decisions it not always clear. Credibility would require the Bank to be able to separate the results generated from randomized evaluation from the data reported by the rest of the organization. The results of studies produced or endorsed by the unit could be published separately from other World Bank documents.

Working with Partners

An evaluation unit would have a tremendous impact in terms of working with partners, in particular, NGOs and academics. For projects submitted from outside the unit, a committee within the unit, perhaps with the assistance of external reviewers, could receive proposals from within the international organizations or from outsiders and choose projects to support. The unit could also encourage the replication of important evaluations by sending out calls for specific proposals. Many NGOs would certainly be willing to take advantage of the opportunity to obtain funding. NGOs are flexible and entrepreneurial and can easily justify working with only some people, because they do not serve the entire population. The project could then be conducted in partnership with people from the unit or other researchers, especially academics, to ensure that the team has the required competencies. The unit could provide both financial and technical support for this project with dedicated staff and

researchers. Over time, based on the experience acquired, the unit could also serve as a more general resource center by developing and disseminating training modules, tools, and guidelines for randomized evaluation. It could also sponsor training sessions for practitioners.

Certifying and Disseminating Evaluation Results

Another role the unit could serve, after establishing a reputation for quality, is acting as a certifying body, clearinghouse, and dissemination agency. To be useful, evaluation results need to be accessible to practitioners within and outside the development agencies. A role of the unit could be to conduct systematic searches for all impact evaluations, assess their reliability, and publish the results in the form of policy briefs and in a readily accessible and searchable database. The database should include all the information needed to interpret the results (estimates, sample size, region and time, type of project, cost, cost-benefit analysis, caveats, and so on), as well as some rating of the validity of the evaluation and references to other related studies. The database could include both randomized and nonrandomized impact evaluations and clearly label the different types of evaluation. Evaluations would need to satisfy minimum reporting requirements to be included in the database, and all projects supported by the unit would have to be included in the database, whatever their results. This would help alleviate the publication bias problem, whereby evaluations that show no results are not disseminated. While academic journals may be uninterested in publishing the results of programs that failed, from the point of view of policymakers, this knowledge is as useful as knowing about projects that succeeded. Ideally, over time, the database would become a basic reference for organizations and governments, in particular, as they seek funding for their projects. This database could then jump-start a virtuous circle, with donors demanding credible evaluations before funding or continuing projects, more evaluations being done, and the general quality of evaluation work rising.

Conclusion: Using Evaluation to Build Long-Term Consensus for Development

Rigorous and systemic evaluations have the potential to leverage the impact of international organizations well beyond their ability to finance programs. Credible impact evaluations are international public goods: the benefits of knowing that a program works or does not work extend well beyond the organization or the country implementing the program. Programs that have been shown to be successful can be adapted for use in other countries and can be scaled up within countries, while unsuccessful programs can be abandoned. By promoting, encouraging, and financing rigorous evaluations of the programs they support, as well as of programs others support, the international organizations could provide guidance to the international organizations themselves, as well as to other donors, governments, and NGOs in the ongoing search for successful programs, and thereby improve the effectiveness of

development aid. Moreover, by credibly establishing which programs work and which do not, the international agencies could counteract skepticism about the effectiveness of spending on aid and build long-term support for development. This is the opportunity to achieve real scaling up.

Notes

1. See http://www.unicef.org/programme/lifeskills/priorities/index.html.
2. The World Bank is not immune to recommending programs whose effectiveness has not been established. A publication by Narayanan (2000) lists a series of programs recommended by the World Bank, of which few have been evaluated (Banerjee and He 2003).
3. PROGRESA is so called from the Spanish acronym for Program for Health, Education, and Nutrition.
4. See http://www.ifpri.org/themes/progresa.htm.
5. To make this point precisely, one would need a full cost-benefit analysis of both programs to see whether the same improvement in human capital were achieved with the same expenditure. The paper on India does not yet include a cost-benefit analysis.
6. Galasso, Ravallion, and Salvia (2002) use the same technique to control for endogenous take-up of a subsidy voucher and training program in Argentina, and Banerjee and others (2003) use it to control for the fact that only two-thirds of the schools allocated to the treatment group actually received the remedial education teachers.
7. A representative from a large organization once objected to the idea that randomized evaluations could be taught and "were not nuclear physics." His answer was that "studying human beings is much more complicated than nuclear physics." This exactly makes the point that, unlike for physics, there are no general laws of human behavior, and therefore nothing general can be said.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Angrist, Joshua, and Alan Krueger. 1999. "Empirical Strategies in Labor Economics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics,* vol. 3A. Amsterdam: North Holland.

_____. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15(4): 69–85.

Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): 533–75.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92(5): 1535–58.

Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek. 1999. "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* 6(4): 453–70.

Attanasio, Orazio, Costas Meghir, and Ana Santiago. 2001. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA." Inter-American Development Bank, Washington, D.C. Processed.

Banerjee, Abhijit. 2002. "The Uses of Economic Theory: Against a Purely Positive Interpretation of Theoretical Results." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Banerjee, Abhijit, and Ruimin He. 2003. "The World Bank of the Future." *American Economic Review Papers and Proceedings* 93(2): 39–44.

Banerjee, Abhijit, Suraj Jacob, and Michael Kremer with Jenny Lanjouw and Peter Lanjouw. 2001. "Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials." Harvard University and Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2003. "Remedying Education: Evidence from Two Randomized Experiments." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Behrman, Jere, Piyali Sengupta, and Petra Todd. 2002. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Mexico." University of Pennsylvania, Philadelphia. Processed.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Difference in Differences Estimates?" *Quarterly Journal of Economics* 119(1): 249–75.

Besley, Timothy, and Anne Case. 2000. "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *Economic Journal* 110(467): F672–F694.

Bobonis, Gustavo, Edward Miguel, and Charu Sharma. 2002. "Iron Supplementation and Early Childhood Development: A Randomized Evaluation in India." University of California, Berkeley. Processed.

Buddelmeyer, Hielke, and Emmanuel Skoufias. 2003. *An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA*. Discussion Paper no. 827. Bonn: Institute for Study of Labor.

Campbell, Donald T. 1969. "Reforms as Experiments." *American Psychologist* 244: 7–29.

Card, David. 1999. "The Causal Effect of Education on Earnings." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics,* vol. 3A. Amsterdam: North Holland.

Case, Anne. 2001. "The Primacy of Education." Princeton University, Princeton, N.J. Processed.

Chattopadhyay, Raghabendra, and Esther Duflo. Forthcoming. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*.

Cook, Thomas D. 2001. "Reappraising the Arguments Against Randomized Experiments in Education: An Analysis of the Culture of Evaluation in American Schools of Education." Northwestern University, Chicago. Processed.

Cronbach, L. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.

Cronbach, L., S. Ambron, S. Dornbusch, R. Hess, R. Hornik, C. Phillips, D. Walker, and S. Weiner. 1980. *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass.

Cullen, Julie Berry, Brian Jacob, and Steven Levitt. 2002. "Does School Choice Attract Students to Urban Public Schools? Evidence from over 1,000 Randomized Lotteries." University of Michigan, Ann Arbor. Processed.

DeLong, J. Bradford, and Kevin Lang. 1992. "Are All Economic Hypotheses False?" *Journal of Political Economy* 100(6): 1257–72.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91(4): 795–813.

Duflo, Esther, and Michael Kremer. Forthcoming. "Use of Randomization in the Evaluation of Development Effectiveness." In *Proceedings of the Conference on Evaluating Development Effectiveness*. Washington, D.C.: World Bank, Operations Evaluation Department.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2003. "Understanding Technology Adoption—Fertilizer in Western Kenya: Preliminary Results from Field Experiments." Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Duflo, Esther, Pascaline Dupas, Michael Kremer, and Samuel Sinei. 2003. "Evaluating HIV/AIDS Prevention Education in Primary Schools: Preliminary Results from a Randomized Controlled Trial in Western Kenya." Harvard University and Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Galasso, Emanuela, Martin Ravallion, and Agustin Salvia. 2002. "Assisting the Transition from Workfare to Work: A Randomized Experiment." World Bank, Development Research Group, Washington, D.C. Processed.

Gertler, Paul J., and Simone Boyce. 2001. "An Experiment in Incentive-Based Welfare: The Impact of PROGRESA on Health in Mexico." University of California, Berkeley. Processed.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2003. "Teacher Incentives." Harvard University, Cambridge, Mass. Processed.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. Forthcoming. "Retrospective Versus Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics.*

Heckman, James, Lance Lochner, and Christopher Taber. 1998. "General Equilibrium Treatment Effects: A Study of Tuition Policy." Working Paper no. 6426. National Bureau of Economic Research, Cambridge, Mass.

Hsieh, Chang-Tai, and Miguel Urquiola. 2002. "When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program." Princeton University, Princeton, N.J. Processed.

Imbens, Guido, and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2): 467–75.

Kremer, Michael. 2003. Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." *American Economic Review Papers and Proceedings* 93(2): 102–15.

Krueger, Alan. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2): 497–532.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training with Experimental Data." *American Economic Review* 76(4): 604–20.

Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics* 13(2): 151–61.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1): 159–218.

Morduch, Jonathan. 1998. "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh." Princeton University, Princeton, N.J. Processed.

Narayanan, Deepa, ed. 2000. *Empowerment and Poverty Reduction: A Sourcebook*. Washington, D.C.: World Bank.

Pandey, Raghaw Sharan. 2000. *Going to Scale with Education Reform: India's District Primary Education Program, 1995–99*. Education Reform and Management Publication Series, vol. I, no. 4. Washington, D.C.: World Bank.

Pitt, Mark, and Shahidur Khandker. 1998. "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy* 106(5): 958–96.

Pritchett, Lant. 2002. "It Pays to Be Ignorant: A Simple Political Economy of Rigorous Program Evaluation." *Journal of Policy Reform* 5(4): 251–69.

Rosenbaum, Paul R. 1995. "Observational Studies." In *Series in Statistics*. New York: Springer.

Shultz, T. Paul. Forthcoming. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." *Journal of Development Economics*.

Comment on "Using Randomized Experiments and Structural Models for Scaling Up: Evidence from the PROGRESA Evaluation" by Orazio P. Attanasio, Costas Meghir, and Miguel Székely and "Scaling Up and Evaluation" by Esther Duflo

FRANÇOIS BOURGUIGNON

Evaluating policies, programs, and projects as rigorously as possible is crucial for development strategies to be effective, but certain steps must be taken for such an evaluation to be feasible. Too often policies are initiated based on economic arguments that rely on a loose view of economic behavior and are evaluated *a posteriori* based on a few aggregate statistics with ambiguous significance. Of course, in some situations no other approach is possible: decisions have to be made quickly, data necessary to optimize the policy design are unavailable ex ante, and data needed to identify the actual effects of the policy ex post were not prepared with enough anticipation. In such circumstances, the intuition and capacity of policymakers and their economic advisers for explicit or implicit modeling are ultimately responsible for good or bad policy. However, in some other situations policy decisions may rely on better information and their results may be better identified ex post at a limited extra cost. The papers in this session discuss methods for this type of evaluation and the possibility of using such evaluation for "scaling up" policies or, in other words, modifying their design, expanding them either in terms of coverage or expenditure per beneficiary, or even exporting them to other countries or regions.

These two papers appear to be quite different because of the techniques they use. In effect, they lie at opposite ends of a range of techniques, the best use of which presumably depends on the policies being evaluated and the data that may be collected. Both papers are clear and well done. Discussing them in depth would run the risk of being excessively technical. Adopting a more general point of view and reflecting on the actual use to be made of the techniques illustrated by these two papers seemed to be more helpful.

I start by stating precisely what is required from evaluation tools, especially in connection with scaling up. I then establish a typology of existing tools based on two

fundamental criteria: structural versus reduced form modeling and experimental versus standard survey data. Finally, the discussion focuses on the adequacy of these techniques with respect to the policies being evaluated or the circumstances in which they have to be evaluated. By analogy with the two papers being discussed, I consider only policies with a micro household focus under the general assumption that they do not indirectly modify the economic environment of agents through general equilibrium effects.

A simple general way of characterizing these policies is as follows. The objective of the policy may be taken to be the modification of a set of outcomes, y, in the sphere of household activity. This set may include real income, food consumption, children's school attendance and actual learning, health status of household members, and so on. The channel used by the policy is a set of elements, z, in households' economic environment, such as taxes, cash transfers, and accessibility and quality of public services. Practically, the policy consists of modifying the channel variable(s), z, in order to modify the vector of outcomes, y, in a predetermined direction. Formally, one could say that z is transformed into $z + t$. By analogy with experimentation, t is the treatment imposed on households, and the policy may be characterized by the way in which this treatment depends on agents' characteristics, x.[1] Let $T(x)$ be that function. The policy is thus characterized by the nature of the channel variables, z, the policy design $T(\ )$, and the outcomes of interest, y. It is not difficult to check that most policies alluded to in the two papers can indeed be represented in this general way, even though the additivity of t and z sometimes appears as a notational artifact.

With these simple notations, "evaluating" a policy in a given population, P, of agents is equivalent to estimating the expected change in the outcomes, y, for given values of the channel variables, z, and other characteristics, x, and, of course, given the policy function, $T(\ )$. Policymakers are essentially interested in the expected change in outcome conditional on all the preceding factors. Evaluation therefore consists of estimating

$E\{\Delta y | z, x; T(\ ); P\}.$

The reason for the conditioning on the population P being considered is simply that agents in different populations or in different environments might not react in the same way to the policy $T(\ )$. In this same framework, scaling up may be defined simply as the effect on the preceding quantity of changing the design $T(\ )$ of the policy or the population P.

With this description of policies and scaling up, a rough typology of evaluation techniques may now be established. It is based on two related criteria. The first is whether the evaluation is to be carried out ex ante or whether it is done ex post when the policy $T(\ )$ is implemented, at least on an experimental basis in a few communities. The second criterion is whether the evaluation relies on a structural or reduced form representation of agents' behavior.

Ex ante evaluation necessarily relies on a structural model that relates individual characteristics and environment to outcomes. Given a sample of agents, i, with

information about their economic environment, z_i, individual characteristics, x_i, and outcomes, y_i, a simple specification of that model is given by:

$$y_i = a_P.x_i + b_P.z_i + c_P.x_i.z_i + u_i, \tag{1}$$

where u_i is a residual term and the behavioral parameters a, b, and c all depend on the population P being considered. Usual econometric techniques permit obtaining estimates (denoted by ^) of behavioral parameters. The ex ante evaluation is then given by

$$E(\Delta y_i \mid z_i, x_i; T(\ ); P) = (\hat{b}_P + \hat{c}_P z_i)T(x_i). \tag{2}$$

This evaluation is more or less precise depending on the quality of the estimators $\hat{b}$ and $\hat{c}$ or on the quality of the model and of the data.[2]

Ex post, a sample of observations may be collected with individuals receiving different treatments $t_i = T(x_i)$, and some of them receiving no treatment at all, whether by policy design or because data are experimental and they include a control group that has not been exposed to the policy. Indeed, it is in this latter case that referring to t_i as a "treatment" makes sense. A direct evaluation method then consists of comparing outcomes among individuals with different treatments, t_i. Generalizing the preceding approach, this is equivalent to estimating the following model:

$$y_i = a_P x_i + b_P z_i + c_P x_i z_i + A_P t_i + B_P t_i.x_i + C_P t_i.z_i + v_i. \tag{3}$$

The evaluation of the policy is then given by

$$E(\Delta y_i \mid z_i, x_i; T(\ ); P) = (\hat{A}_P + \hat{B}_P x_i + \hat{C}_P z_i)T(x_i). \tag{4}$$

This approach is reduced form in the sense that no constraint relates the coefficients A, B, and C that measure the effect of the policy with the behavioral coefficients, a, b, and c.

The main difficulty is ensuring that the coefficient estimates in the preceding expression are unbiased. With ordinary least squares, this requires that the treatment, t_i, be independent of unobserved outcome determinants in the residual term, v_i. The enormous advantage of the randomized experimentation framework Esther Duflo discusses in her paper is precisely to guarantee the absence of bias. In other cases, the bias must be corrected, which generally relies on extraneous information that is not always available.

The two preceding approaches may be combined into an ex post structural model. The paper by Orazio P. Attanasio, Costas Meghir, and Miguel Székely gives an example of such a model. The structural model may now be written as:

$$y_i = a_P x_i + b_P(z_i + t_i) + c_P x_i(z_i + t_i) + u_i, \tag{5}$$

and the policy can be evaluated according to equation (2) as with the ex ante model. The main aim of this approach with respect to the ex ante model is to make the estimates of behavioral parameters more precise, essentially by adding agents exposed to the program to the sample, whether this is on an experimental basis or with full policy implementation. This added information also permits a simple test of the structural

model. If it is unsatisfactory, then testing the reduced form specification (3) against equation (5) would lead to rejection of the structural model specification.[3]

Now that a range of evaluation approaches has been specified, the question is which one should be chosen depending on the type of policy and the nature of the data available or to be collected. From this point of view, the ex ante/ex post distinction emphasized earlier must not lead to confusion. It does not mean that only a structural approach is possible before a policy is actually launched. Experimentation with pilot programs and control groups may also be a possibility. With scaling up objectives in mind, is it therefore preferable to run these controlled experiments, or is it better to rely on the predictions of a structural model estimated on the basis of some initial survey, or should both approaches be systematically combined?

In answering this question, note that the preceding choice is not always possible. Some policies, programs, and projects simply cannot be evaluated on the basis of a structural model, either because some of the key variables for describing the behavior of agents are intrinsically unobservable, or because no obvious and simple-enough economic model may be invoked to relate y to x and z. The remedial education project Duflo cites falls in this category: no structural model of the educational process is available that would permit predicting the increase in test scores of the children exposed to this program. There are many other examples where only some kind of experimentation or pilot project would permit a more or less precise evaluation. Conversely, in some cases randomized experimentation is impossible for ethical or political reasons. Equity reasons may explain why a major program can be implemented in some communities and not in others. At best, some delay may be introduced across localities, as occurred with PROGRESA,[4] but then modifying the design of the program as it is progressively expanded becomes difficult. In that sense, scaling up is impossible.[5]

Cases where the alternative arises are when some structural estimate of the effect of the policy can be obtained and when experimentation or pilot projects are also possible. The trade-off then seems to be essentially between policy design flexibility and estimation precision. On the one hand, thinking of an experimentation framework that would permit testing a large range of policy designs $T(\ )$ is difficult. In principle, one could conduct a number of different projects in different communities, and this is not an exceptional experimental design, but sampling must take place not only across policy designs, but also across communities so as to have enough joint variability of policy and household characteristics. Practically, this limits the number of alternative policies to be compared with each other. This is not a constraint with the structural approach. On the other hand, the estimates obtained with the structural approach are likely to depend on the quality of the estimation work, that is, the quality of the fit of the structural model to the data. This problem of precision is less acute in the experimentation approach, which relies on sample means.

Thus the structural approach is preferable when it can rely on a model that relates outcomes to policy channels and agents' characteristics in a precise way. Scaling up to other policy designs is then relatively easy. However, if no precise

model is available for a variety of reasons, the experimentation ex post approach is preferable, even though scaling up to other designs will generally prove difficult. Thus randomized trials are more appropriate for policies with a limited number of alternative formats. Intermediate cases in between these situations are likely to be better covered by ex post structural models that combine the two approaches.

Two fundamental difficulties of scaling up remain to be discussed. The first refers to the population being covered. In the preceding equations, estimated parameters are indexed by the population that the observation sample was taken from. Indeed, Attanasio, Meghir, and Székely show quite clearly that estimating a structural model of PROGRESA for some Mexican states and simulating the resulting model for others leads to significant gaps between simulated and actual outcomes. Of course, this precaution is superfluous when structural modeling relies on nationally representative surveys, which is often likely to be the case.[6] Things are not as simple for approaches based on experimentation. As experimentation is likely to be conducted on limited geographical units, the problem of extrapolating to other localities may be more serious.

The second difficulty is common to the two approaches. It has to do with the potential general equilibrium effect of scaling up. The example of school vouchers Duflo cites provides a good example. That vouchers distributed randomly among a small sample of households contributed to better school achievement by beneficiaries is undoubtedly an interesting and important result. But suppose that the program is scaled up to the national level. It then seems unrealistic to keep assuming that nothing is modified on the supply side of the education market. Private schools will charge higher fees, or if they are not allowed to do so, they will increase their pupil-to-teacher ratios, thereby reducing their quality superiority and threatening the entire program. Some structural modeling is necessary to take these indirect effects into account, but it is of a different nature than the type discussed earlier.

Where does all this leave us? Evaluating policies, programs, and projects ex ante, as well as ex post, in terms of both expected and overall outcomes is an absolute necessity for making development policies more effective. When referring to household-oriented micro policies, a range of evaluation techniques is available, with randomized experimentation and structural modeling at both ends and ex post structural modeling at the middle. The main message is that there is no reason to prefer one approach over the other a priori. Both are equally relevant, but in different situations and for different types of policies. Randomized experimentation and related techniques have received a great deal of attention lately. They are certainly appealing, and the hope is that the use of controlled trials will spread rapidly among policymakers and program managers in developing countries. Yet these techniques are far from covering the entire range of needs. Users must not forget that structural modeling is a better evaluation tool in several cases, especially when scaling up is concerned. In some instances, it really is the only evaluation technique available, whereas in others it can usefully complement experimental techniques.

Notes

1. The function $T(\ )$ often depends on outcomes, y, too. As the paper by Attanasio, Meghir, and Székely shows, this is especially true for conditional cash transfer programs such as PROGRESA in Mexico. For simplicity, what follows ignores this possibility.

2. The intent of the interaction term is to make the effect of the policy dependent on individual characteristics, as can be seen in equation (2).

3. Todd and Wolpin (2003) conduct such a test for the PROGRESA program. They indeed show that a structural model estimated ex ante, or on the control groups in the first years of PROGRESA implementation, predicts the effects of the program as estimated through randomized experiments rather well. Like Attanasio, Meghir, and Székely, Todd and Wolpin's model is somewhat elaborate. An interesting issue is whether the same out-of-sample fit would have been obtained with a simpler specification.

4. PROGRESA is so called from the Spanish acronym for Program for Health, Education, and Nutrition.

5. From this scaling up point of view the pure randomized experiment aspect of PROGRESA was more interesting for the international development community than for Mexican policymakers. Yet this aspect permitted an overall ex post evaluation of the program, as initially designed, which would have been much more difficult otherwise.

6. Attanasio, Meghir, and Székely's virtual scaling up experiment is somewhat artificial, because they ignore the information on differences across states that could have been derived on a pure ex ante basis.

References

The word "processed" describes informally reproduced works that may not be commonly available in libraries.

Todd, P., and K. Wolpin. 2003. "Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico." University of Pennsylvania, Department of Economics, Philadelphia. Processed.

Comment on "Using Randomized Experiments and Structural Models for Scaling Up: Evidence from the PROGRESA Evaluation" by Orazio P. Attanasio, Costas Meghir, and Miguel Székely and "Scaling Up and Evaluation" by Esther Duflo

T. PAUL SCHULTZ

An argument Esther Duflo advances is that if program evaluation were done better, the more reliable evaluations would be accorded greater weight in the allocation of resources for development, public and nongovernmental organization resources would be spent in a more cost-effective manner, and the world would be more likely to increase its commitment of funds to bilateral and multilateral development organizations. These are reasonable expectations given our limited understanding of how public priorities are set, and they justify inclusion of the topic in this conference on accelerating development.

Both of the papers argue the advantages of one methodological approach for producing better evaluation studies. Duflo contends that randomized social experiments have the greatest promise for producing reliable conclusions on the effectiveness of specific program interventions relative to the main alternative form of evaluation study, which I will call for short matching studies. Orazio P. Attanasio, Costas Meghir, and Miguel Székely endorse and illustrate a less frequently followed evaluation methodology that has attracted a growing circle of economists. It involves hypothesizing a structural microeconomic model accounting for the program's objective or outcome into which the policy instruments of a public program can be introduced. Within such an estimated representation of technological possibilities and behavioral choice, the effectiveness of specific observed policies can be simulated, and the researcher can also forecast the effectiveness of policies that lie outside the range of observed experience and extrapolate the effects of policies to populations that have not been empirically evaluated.

Even though the authors of these two papers may share long-run policy goals, they endorse very different methods for producing better evaluation results. Duflo extends her argument to consider the characteristics of political and economic institutions

T. Paul Schultz is professor of economics at Yale University, New Haven, Connecticut.

and actors that may deter investing the justified amount in the production of randomized evaluation studies, which the international development community should view as a high priority public good.

I would like to start by drawing attention to some of the difficulties in modeling structurally the school enrollment decision cited by Attanasio, Meghir, and Székely based on the Mexican PROGRESA randomized experiment, which transfers cash to poor rural mothers if their children are enrolled in grades 3 through 9.[1] The authors divide the PROGRESA evaluation sample into two parts; estimate their structural model from the 80 percent of the data drawn from four poorer Mexican states; and then extrapolate their simulated results to the remaining 20 percent of the PROGRESA data from three richer states that happen to be closer to the dynamic, urban areas of central Mexico surrounding Mexico City. In the PROGRESA evaluation area, which encompasses about 500 poor rural villages, roughly two-thirds of the villages were offered the PROGRESA package of program transfers starting in mid-1998, while the remaining villages were surveyed but not brought into the program until December 1999. All households in all 500 villages were surveyed at approximately six-month intervals following the benchmark census survey of October 1997.

Several issues need clarification in relation to the authors' modeling strategy because they are not sufficiently described in their paper for me to infer how the variables listed in table 3 are defined or derived or how the model is estimated. To begin with, any economic structural model of the enrollment decision will depend on the private family costs and benefits of schooling. The costs are structurally represented by the student's foregone wage opportunities or what the child could produce additionally for the family or earn in the labor market if he or she were not going to school for one more year. The primary private benefit from enrollment is the extra wage an adult worker may expect to receive if he or she completes one more year of schooling, plus the program transfer for the child being enrolled. In the area where PROGRESA was implemented, better educated young people are more likely to leave their villages and migrate to urban areas, where wages are higher and the percentage wage premium for an educated worker is likely to be larger than in the rural areas where they went to school.

Unfortunately for the econometrician, few school-age children work in paid employment in the PROGRESA area. According to the baseline census of October 1997, only 2 to 5 percent of the children age 8 to 12 are reported working for a wage, and from age 13 to 16 the proportion rises from 10 to 44 percent for boys and from 4 to 15 percent for girls (Schultz forthcoming, table A-4). Thus there are few child observations on which to base child wages or the critical opportunity cost of school enrollment, and undoubtedly those children who do work for pay are not representative of those who do not, complicating any interpretation of the distribution of such wages by sex, age, and completed schooling.

Another possible source of wage data is the community (village) questionnaire, which is completed by an "informed person" in each community. Even though about 95 percent of the community questionnaires report a typical day wage for a male

casual worker in agriculture, only 55 percent report such a wage for adult females, and fewer than 10 percent report a wage for children. The PROGRESA data files I have consulted do not shed much systematic light on child wages. How the authors dealt with this limitation in the data is unclear, though I might note that similar problems arise in most surveys in low-income countries where parents are unable or unwilling to report a wage or assess the monetary value of their child's contribution to the family's resources.

Adults in these communities also rarely report adult wages, because workers are predominantly self-employed farmers, small entrepreneurs, or unpaid family workers. The proportion of individuals working for wages will tend to vary with education, and therefore the observed wage differentials by education are unlikely to measure differences in the average marginal productivity between these educational groups. The paper should also explain how the authors have closed this critical data gap, although it may have been set forth in an unpublished paper.

Specifying a structural model for school enrollment in a setting where little representative information is available on child wages, adult wages by education, migration probabilities, or urban wage opportunities by education is an ambitious research program. As a policymaker I might be inclined to prefer a more transparent reduced form estimate of how access to program benefits (that is, the effect of the intention to treat) is related to the probability of school enrollment, rather than rely on a structural representation for which several critical variables are unobserved. If the reduced form methodology is adopted to evaluate the effect of the program, then the alternatives Duflo describes may be considered: either a random experiment to proxy how access to the program treatment is associated with enrollment; or a method of matching individuals or groups on observed characteristics according to a propensity score; or a multivariate model of enrollment and a sample selection rule determining the probability of program treatment, as developed by Heckman (1979). The structural model also assumes that a child either works or goes to school, whereas the PROGRESA data reveal that this dichotomous specification may be inadequate. Many children work and attend school and some do neither. My own analysis implies that the availability of PROGRESA transfers for those who are poor and eligible has a larger effect on increasing school enrollment than on decreasing work by the school-age child, but the hours allocated to both activities, which are needed to convert these program effects into comparable units, are not available from the database (Schultz forthcoming).

Finally, a puzzling feature of the structural model relates to the note to table 3, which indicates that dummy variables are included for a state-level fixed effect on enrollment. If the paper's goal is to test whether the model fitted to one set of states is able to extrapolate "good" estimates for another set of states, how this initial fit can include controls for the original states that cannot be extrapolated to the other states is unclear. Table 4 reports how well the simulations do in approximating the randomized program effects for the extrapolated states, but to perform this calculation, the econometrician is assumed to know the correct value of the state dummies or relies on an arbitrary choice among the initial state fixed effects.

The structural modeling method tends to extend the research beyond the reach of classical statistical methods for hypothesis testing and comparisons between nested models. Applied econometricians attempt to assess standard statistical problems with relatively transparent specification tests, such as those that arise because of (a) omitted variable bias, (b) simultaneous equation bias, (c) errors in measuring explanatory variables, (d) errors in specifying functional forms, (e) non-normality in the distribution of endogenous explanatory variables, and (f) qualitative dependent variables. How well do structural models evaluate the robustness of these methods to the choice among plausible specifications and calibrated parameters?

The authors of both papers are encouraged to elaborate on what types of programs and what forms of program outcomes and behavioral changes are most plausibly modeled by either randomized experiments or structural micro-econometric models. Would Attanasio, Meghir, and Székely agree that time allocation between school and work in urban areas and urban wage returns to schooling could be more confidently modeled by their structural methods than the rural schooling decision they seek to explain in this paper? Duflo notes that her evaluation methods are more promising where a program has a clearly delineated treatment (and control) population at the level of the individual or the geographic area. She therefore concludes that monetary policy, for example, does not lend itself to her form of evaluation study, but presumably that education, preventive health, and possibly curative health do, and under special conditions local infrastructure and services may be evaluated, such as water, sanitation, electrical services, and irrigation. Would she agree that a longer gestation lag between program inputs and policy outcomes makes the evaluation process more difficult, as sample attrition increases and prospective follow-up surveys become more costly? Even in such cases as lifetime programs for human capital investment in early childhood health and nutrition, HIV/AIDS prevention, or schooling and mobility, random experimental results might be used to assess the early consequences of a program on, for example, a child's weight-to-height or years of completed schooling. In these examples, the short-term results from random experiments could be combined with a well-established structural model of lifecycle effects of child health or schooling on adult production and consumption opportunities to approximate internal rates of return to such human capital program inputs and private investments over a cohort's expected life.

Also if the decision to participate in the program is related to the effect of the program treatment on the participant, or if the allocation of program resources among program participants is determined by the participant and service providers, the reduced form random experimental methodology may be more reasonable than trying to specify and estimate a comprehensive dynamic structural model. For example, if teachers or doctors allocate their time and intervention therapies among students and patients differentially based on their professional knowledge of the student and patient, respectively, a comprehensive structural model is expected to account for the teacher's and doctor's treatment decisions and measure the public and private costs of each alternative treatment allocation.

Duflo argues in conclusion that international agencies, such as the World Bank, should internalize the international spillover benefits from reliable, high-quality, readily generalizable evaluation studies. Such studies, which are likely to be based on a randomized design, have the potential to be an important global public good in economic development and should not be charged against the loans or grants supporting a country-specific project, except to the extent that the evaluation findings can feed back into subsequent fine-tuning of that country's project design.

However, in encouraging the World Bank to create a separate division responsible for evaluation studies, Duflo does not stress strongly enough the obvious issue of conflict of interest when program managers are responsible for evaluating their own programs. The separation of responsibilities for evaluation studies could encourage a concentration of professional expertise in evaluation methods and more readily reward personnel for competent implementation of such studies and diffusion of their findings throughout the policy community, but personnel who are engaged in designing and negotiating loans and grants for projects should not be expected to evaluate their own projects without bias. In addition, evaluation studies are likely to be completed several years after a program has been funded, and the individual in the Bank who was responsible for its design and funding is likely to have been reassigned to another part of the Bank in the interim and may have no interest in, or authority to, monitor its final evaluation. Hence many sound arguments support Duflo's proposed separation of responsibilities between operations and evaluation studies in a development organization such as the World Bank.

The last recommendation in Duflo's paper may also need to be reemphasized. Evaluation builds on reliable preprogram benchmark survey or census data and postprogram survey data. These data could often be obtained at low marginal cost from established national statistical programs in a country, if the statistical programs were designed to include local area identifiers to facilitate a variety of potential program evaluations. International development agencies should therefore invest resources to build institutional capacity in survey design and data processing expertise that would help countries regularly collect national household surveys with local area identifiers, perhaps using universal global positioning satellite technology. These data resources and survey capacities would then be useful for program evaluation, and yet be professionally designed in a form that coding would protect the confidentiality of the responding individual, family, or firm. The public good attribute of such coordinated data collection might be a lesson the World Bank could share with all its member states, including the United States, where the Bureau of the Census restricts evaluation studies by releasing few local area identifiers in its public use files.

These two papers are of interest in distinctly different ways. Attanasio, Meghir, and Székely's paper is an innovative extension of microeconometric methodology that may be expected to provide an exploratory policy tool in some contexts. Duflo's paper presents a forceful case for organizations such as the World Bank to dedicate more resources to performing randomized evaluation studies of its major programs in education, preventive health, family planning, welfare, and additional public service sectors.

Note

1. PROGRESA is so called from the Spanish acronym for Program for Health, Education, and Nutrition.

References

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153–61.

Schultz, T. P. Forthcoming. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." *Journal of Development Economics*.

The Annual World Bank Conference on Development Economics (ABCDE) brings together the world's leading scholars and development practitioners for a lively debate on state-of-the-art thinking in development policy and the implications for the global economy.

The 15th conference, held in Bangalore, India, on May 21–23, 2003, was the first such conference to be held in a developing country. The major theme of the conference, accelerating development, was divided into four topics: fostering entrepreneurship, innovation, and growth; challenges of development in lagging regions; participation, inclusion, and results; and scaling up and evaluation.

IN THIS VOLUME

Introduction by François Bourguignon and Boris Pleskovic; opening address by S. M. Krishna; keynote addresses by Azim Hasham Premji, Nicholas Stern, and Rakesh Mohan; papers by T. N. Srinivasan, Robin Burgess and Anthony J. Venables, Partha Dasgupta, Justin Yifu Lin and Mingxing Liu, Jean-Philippe Platteau, Benno J. Ndulu, Orazio P. Attanasio, Costas Meghir, and Miguel Székely, and Esther Duflo; and comments by Gur Ofer, Sidney Winter, Govinda Rao, Karen R. Polenske, Rohini Nayyar, Vijayendra Rao, François Bourguignon, and Paul Schultz.

OXFORD
UNIVERSITY PRESS

THE WORLD BANK